Temporal selectivity declines in the aging human auditory cortex

Julia Erb[1], Lea-Maria Schmitt[1], Jonas Obleser[1]

[1]Institute of Psychology, University of Lübeck, Maria-Goeppert-Str. 9A, 23562 Lübeck, Germany

*Correspondence: Julia Erb, Maria-Goeppert-Str. 9a, 23562 Lübeck, Germany
Phone: +49 451 3101 3631, e-mail: julia.erb@uni-luebeck.de

Abstract

Current models successfully describe the auditory cortical response to natural sounds with a set of spectro-temporal features. However, these models have hardly been linked to the ill-understood neurobiological changes that occur in the aging auditory cortex. Modelling the hemodynamic response to a rich natural sound mixture in N=64 listeners of varying age, we here show that in the older listeners' auditory cortex, the key feature of temporal rate is represented with a markedly broader tuning. This loss of temporal selectivity is most prominent in primary auditory cortex and planum temporale, with no such changes in adjacent auditory or visual brain areas. Amongst older listeners, we observe a direct relationship between chronological age and temporal-rate tuning, unconfounded by auditory acuity or model goodness of fit. In line with senescent neural dedifferentiation more generally, our results highlight decreased selectivity to temporal information as a hallmark of the aging auditory cortex.

Keywords: fMRI, spectro-temporal modulations, healthy aging, auditory system, presbycusis, hearing loss, temporal rate coding

Impact Statement

The aged human auditory cortex shows preserved tonotopy, but temporal modulations are represented with a markedly broader tuning, highlighting decreased temporal selectivity as a hallmark of the aging auditory cortex.

Introduction

Age-related hearing loss is a frequent cause for a decline in speech comprehension, particularly in complex acoustic environments. As the temporal envelope of speech is crucial for speech comprehension (Shannon et al., 1995), the ability to accurately encode temporal features in the auditory system is pivotal for successful perception of speech and conspecific vocalizations (Peelle and Wingfield, 2016). Age-related deficits have been observed in different temporal psychoacoustic tasks (e.g. gap detection; Snell et al., 2002). Those psychoacoustic studies suggest that the precision of processing of temporal cues declines with age (e.g. Gordon-Salant et al., 2006).

While age-related hearing loss has been attributed to a variety of peripheral declines in auditory coding (e.g. Gates and Mills, 2005), fewer studies have looked at more central levels of auditory processing. Yet, declines in temporal processing are hypothesized to be not exclusively peripheral in origin, but more central than at least the mid-brain (Recanzone, 2018). In the aged macaque auditory cortex, neurons become more broadly tuned to temporal modulations and temporal fidelity of cortical responses decreases (Ng and Recanzone, 2018). Topographically, a neural dedifferentiation is suggested to take place in the older brain: In young macaques, the neuronal response to the inter-stimulus interval in primary field A1 and caudolateral belt area (CL) differs, in that A1 neurons have shorter response latencies. In aged animals, however, no difference was observed between A1 and CL such that aged A1 neurons had an equivalent response latency to CL neurons. Further, in aged neurons, a shift in neural coding strategy from less temporal coding towards more rate coding of temporal modulations was observed (Overton and Recanzone, 2016; for a review of both coding strategies see Joris et al., 2004). For those neurons that still adhered a temporal coding strategy, a decrease in temporal fidelity was observed, although the absolute number of neurons responsive to temporal modulations was unaffected by age.

At higher processing levels using fMRI, only a number of relatively unspecific changes have been reported in the older human listener (for review, see Peelle and Wingfield, 2016). For example, the frontal and cingulo-opercular cortical response is elevated during a challenging speech comprehension or gap detection task (Erb and Obleser, 2013, Vaden et al., 2015, Vaden et al., 2019). Those have been interpreted as compensatory mechanisms in response to loss of sensory acuity. Aging has been hypothesized to lead to a neural dedifferentiation in sensory cortices (e.g., Park et al., 2004). A general observation is that with age, decoding accuracy of stimulus conditions from auditory cortical fMRI responses declines (Lalwani et al., 2019) (but see discussion for contrasting findings in EEG speech tracking, e.g., Presacco et al., 2016). This finding in turn has been linked to an imbalance of excitation and inhibition in older adults, more specifically a reduction of GABA levels in auditory cortex. Accumulating evidence supports the theory of an age-related loss of inhibition in sensory cortices (Caspary et al., 2008).

Recent advances in computational modelling of fMRI data has led to substantial progress in our understanding of how natural sounds become represented in the auditory cortex. A series of studies have shown that cortical processing of sounds is optimized to represent the spectro-temporal modulations which are typically present in conspecific vocalizations such as speech or animal calls (Santoro et al., 2014, Hullett et al., 2016, Santoro et al., 2017, Erb et al., 2018). However, those models have not looked at the pressing case of the aged human cortex. It remains unknown whether — and if so, which — general response properties are altered in aged auditory cortex.

The present study combines fMRI and computational modeling to compare auditory cortical responses to natural sounds in young and older humans. We focus on the age-group comparison at the level of representation of fundamental acoustic features (spectro-temporal modulations). Results show that while the large-scale topographic organization of acoustic features is preserved in the older

auditory cortex, age groups differ in tuning to temporal modulations: Human aging is accompanied by a broadening of temporal rate tuning in auditory cortex.

Materials and Methods

Participants

We invited a total of *n = 75* participants for scanning from which we had to exclude one participant due to excessive movement, three participants due to incidental neurological findings, five participants due to inability of understand speech in noise, one participant due to broken headphones, one participant due to age (43 years, i.e., could not be assigned to young or older group). The remaining *n* = 64 participants were right-handed, young (*n* = 33; aged 18–32, 18 female) and older (*n* = 31; aged 51–78, 15 female) native German speakers. Simulations (custom matlab code) showed that a two-sample permutation test at a conventional type I error rate of 5 % with a sample size of *N* = 27 per group can detect medium to large effects of Cohen's *d* = 0.75 (as can be expected for the neural fMRI measures under consideration, e.g. Alavash et al., 2019) with a satisfactory power of 80 %. The power of our procedure reduces accordingly if the true effect in the population is smaller, but remains over 60 % also for true effects closer to Cohen's *d* = 0.5.

Younger participants had self-reported normal hearing. Older participants' hearing ranged from normal hearing to mild hearing loss: they had undergone pure-tone audiometry during the last two years (for audiograms see Fig. S1A) and were excluded from the study if pure-tone-average (PTA) of the better ear exceeded 30 dB. All participants gave informed consent and were financially compensated or received course credit. All procedures were approved by the local ethics committee of the University of Lübeck (ethical approval AZ 16-107).

93 **Stimuli and task**

94 Participants listened to 64 minutes of a freely narrated audiobook (Hertha Müller, "*Die Nacht ist aus*
95 *Tinte gemacht*") presented against a competing stream of resynthesized natural sounds ("sound
96 textures"; McDermott and Simoncelli, 2011) at 0 dB SNR. Textures were synthesized from the spectro-
97 temporal modulation content of a large set of real-life sounds (n = 192), including speech and vocal
98 samples, music pieces, animal cries, scenes from nature and tool sounds that had been used in
99 previous studies (Moerel et al., 2013, Santoro et al., 2017). Texture synthesis parameters were as
100 follows: Frequency range = 0.02 – 10 kHz, number of frequency bands = 30, sampling rate = 20 kHz,
101 temporal modulation range = 0.5 – 200 Hz, sampling rate = 400 Hz; maximum number of iterations =
102 60. Texture exemplars of 5 s length were concatenated to form the background stream. The order of
103 exemplars was pseudo-randomized across participants (in four different sound orders). In total, each
104 exemplar was repeated four times.

105 Participants were asked to listen to the story and answer three four-choice questions on its
106 semantic content after each run (24 questions in total); chance level was thus proportion correct =
107 0.25. Four-answer choice questions were displayed on a screen and participants responded via a
108 button box in their right hand. Both the young and older group performed on average above chance
109 (Fig. S1b), but young participants performed significantly better than older participants (mean
110 proportion correct difference =0.1, p = 0.009 (permutation test), Cohen's d = 0.65).

111 **Acquisition of MRI data**

112 We acquired functional and structural MRI at 3T (Siemens Magnetom Skyra) with a 64-channel RF
113 head array coil.

114 Functional T2*-weighted data were acquired using an echo planar imaging sequence. We collected
115 eight runs of eight minutes each (8 runs à 519 volumes) using continuous scanning. The acquisition

116 parameters were as follows: repetition time (TR) = acquisition time (TA) = 947 ms, echo time (TE) = 28
117 ms, acceleration factor = 4, flip angle = 60°, field of view (FOV) = 200 x 200 mm, 52 slices; voxel size =
118 2.5 mm isotropic (whole-brain coverage). Field maps for intensity inhomogeneity correction were
119 acquired after every second run (TR = 610 ms, TE1 = 4.92, TE2 = 7.38, flip angle = 60°, FOV = 200 x 200
120 mm, 62 slices, voxel size = 2.5 mm isotropic).

121 Anatomical T1-weighted images were acquired using an MPRAGE sequence (TR = 2400 ms, time to
122 inversion [TI] = 1000 ms, TE = 3.16 ms, flip angle = 8°, FOV = 256 x 256 mm, number of slices = 176,
123 voxel size = 1 mm isotropic, GRAPPA acceleration factor = 2). T2-weighted images were collected at
124 the end of the session (TR = 3200 ms, TE = 449 ms, FOV = 256 x 256 mm, number of slices = 176, voxel
125 size = 1 mm isotropic, GRAPPA acceleration factor = 2).

126 **Preprocessing of MR data**

127 Results included in this manuscript are based on a preprocessing pipeline of *fMRIPprep* 1.2.4 which
128 is based on *Nipype* 1.1.6 (Esteban et al., 2019). The following two paragraphs on preprocessing are
129 based on an automatically generated output of *fMRIPprep*.

130 **Anatomical data preprocessing**

131 The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU)
132 using N4BiasFieldCorrection (ANTs 2.2.0), and used as T1w-reference throughout the workflow. The
133 T1w-reference was then skull-stripped using antsBrainExtraction.sh (ANTs 2.2.0), using OASIS as target
134 template. Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1), and the brain mask
135 estimated previously was refined with a custom variation of the method to reconcile ANTs-derived
136 and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle. Spatial
137 normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c was performed
138 through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of

139 both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter
140 (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9).

141 **Functional data preprocessing**

142 For each of the eight BOLD runs, the following preprocessing was performed. First, a reference
143 volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. A
144 deformation field to correct for susceptibility distortions was estimated based on a field map that was
145 co-registered to the BOLD reference, using a custom workflow of *fMRIPrep*. Based on the estimated
146 susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-
147 registration with the anatomical reference. The BOLD reference was then co-registered to the T1w
148 reference using bbregister (FreeSurfer) which implements boundary-based registration. Co-
149 registration was configured with nine degrees of freedom to account for distortions remaining in the
150 BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation
151 matrices, and six corresponding rotation and translation parameters) are estimated before any
152 spatiotemporal filtering using mcflirt (FSL 5.0.9). BOLD runs were slice-time corrected
153 using 3dTshift from AFNI 20160207. The BOLD time-series (including slice-timing correction when
154 applied) were resampled onto their original, native space by applying a single, composite transform to
155 correct for head-motion and susceptibility distortions. All analyses (see below) were performed in
156 native space. Subsequently, results (best feature maps) were resampled to surfaces on the *fsaverage5*
157 template using mri_vol2surf (FreeSurfer).

158 **Computational modelling**

159 We applied a computational modelling approach to fMRI data as described in Erb et al. (2018).
160 Two modelling procedures were applied (Fig. 1). In a first *univariate encoding* approach, we calculated
161 a modulation transfer function (MTF) for each individual voxel. The MTF characterizes how faithfully

the modulation content of the stimulus gets transferred to the voxel. By assigning the feature value with the maximal response to each voxel, we obtained maps of a voxel's best features across the auditory cortex.

In a second *multivariate decoding* approach, data from voxels were jointly modelled within a model-based decoding framework. The combined analysis of signals from multiple voxels increases the sensitivity for stimulus information that is represented in patterns of activity, rather than in individual voxels. Further, the accuracy with which those features can be reconstructed provides an explicit measure of the amount of information about sound features available in cortex.

Extraction of frequency-specific modulation content

The modulation content of the stimuli was obtained by filtering the sounds within a biologically plausible model of auditory processing (Chi et al., 2005). This auditory model consists of an early stage that models the transformations that acoustic signals undergo from the cochlea to the midbrain; and a cortical stage that accounts for the processing of the sounds at the level of the auditory cortex. We derived the spectrogram and its modulation content using the "NSL Tools" package (available at http://www.isr.umd.edu/Labs/NSL/Software.htm) and customized Matlab code (The MathWorks Inc., Matlab 2014b/2018a).

Spectrograms for all sounds were obtained using a bank of 128 overlapping bandpass filters with equal width ($Q_{10dB} = 3$), spaced along a logarithmic frequency axis over a range of $f = 180–7040$ Hz. The output of the filter bank was band-pass filtered (hair cell stage). A midbrain stage modelled the enhancement of frequency selectivity as a first-order derivative with respect to the frequency axis, followed by a half-wave rectification and a short-term temporal integration (time constant $\tau = 8$ ms).

Next, the auditory spectrogram was further analyzed by the cortical stage, where the modulation content of the auditory spectrogram was computed through a bank of 2-dimensional filters selective

for a combination of spectral and temporal modulations. The filter bank performs a complex wavelet decomposition of the auditory spectrogram. The magnitude of such decomposition yields a phase-invariant measure of modulation content. The modulation selective filters have joint selectivity for spectral and temporal modulations, and are directional, i.e. they respond either to upward or downward frequency sweeps.

Filters were tuned to spectral modulation frequencies of Ω = [0.3, 0.4, 0.8, 1.3, 2.3, 4] cyc/oct, temporal modulation frequencies of ω = [1, 2, 4, 8, 16, 32] Hz, and centre frequencies of f = [232, 367, 580, 918, 1452, 2297, 3633, 5746] Hz. Our rationale for this choice of values was to use a decomposition roughly covering the temporal and spectral modulations present in the acoustic energy of natural sounds we used (for spectro-temporal modulation content of the sounds see supplemental Fig. S2). To avoid overfitting, for the decoding analyses in regions of interest (ROIs), we reduced the number of features such that filters were tuned to spectral modulation frequencies of Ω = [0.25, 0.5, 1, 2 , 4] cyc/oct, temporal modulation frequencies of ω = [1, 2.4, 5.7, 13.5, 32] Hz, and centre frequencies of f = [277, 576, 1201, 2502, 5213] Hz.

The filter bank output was computed at each frequency along the tonotopic axis and then averaged over time. This resulted for the encoding (ROI decoding) analysis in a representation with 6 (5) spectral modulation frequencies × 6 (5) temporal modulation frequencies × 8 (5) tonotopic frequencies = 288 (125) parameters to learn. The time-averaged output of the filter bank was averaged across the upward and downward filter directions (Santoro et al., 2014). Those processing steps were applied to all stimuli, resulting in an [N x F] feature matrix S of modulation energy, where N is the number of sounds, and F is the number of features in the modulation representation.

Prior to acoustic feature extraction, the continuous sounds (~8 min per run) were cut into snippets of the length of the TR (947 ms) resulting in a total of n = 4152 sounds which were subdivided

into training (n = 3114) and test sounds (n = 1038) for the cross-validation procedure (see below). This resulted in an equivalent the temporal resolution of the feature matrix and the fMRI data.

Note that for the encoding analysis, the number of parameters to estimate is thus smaller than the number of observations in the training set (n = 3114 training sounds). In the decoding analysis, the number of fitted features is limited by the number of voxels instead (Equation 5). Each feature was convolved with the standard double gamma model for the hemodynamic response function peaking at 4 s.

Univariate encoding analysis: Model estimation

Based on the training data only, the fMRI activity $Y_i\ [N_{train} \times 1]$ at voxel i was modeled as a linear transformation of the feature matrix $S_{train}\ [N_{train} \times F]$ plus a noise term $n\ [N_{train} \times 1]$:

$$Y_i = S_{train} C_i + n \qquad (1)$$

where N_{train} is the number of sounds in the training set, and C_i is an [F×1] vector of model parameters, whose elements c_{ij} quantify the contribution of feature j to the overall response of voxel i. Columns of matrices S_{train} and Y_i were converted to standardized z-scores. Therefore, Equation (1) does not include a constant term. The solution to Equation (1) was computed using kernel ridge regression (Hoerl and Kennard, 1970). The regularization parameter λ was selected independently for each voxel via generalized cross-validation (Golub et al., 1979). The search grid included 25 values between 10^{-6} and 10^{6} logarithmically spaced with a grid grain of $10^{0.5}$.

To obtain more stable estimates of the voxels' feature profiles, this computation was performed in a fourfold cross-validation procedure using different subsets of the eight runs. For each iteration, two out of the eight runs were selected and left out for testing, resulting in a subset of 3114 training

229 sounds on which the estimation was performed and 1038 test sounds. In this way, we obtained four
230 estimates of each voxel's feature profile which were averaged across iterations.

231 **Model evaluation**

232 To evaluate the model's prediction accuracy we performed a sound identification analysis (Kay et
233 al., 2008). To this end, we used the fMRI activity patterns predicted by the model to identify which of
234 the test sounds had been heard. Given the trained model $\tilde{C}$ $[F \times V]$, and the feature matrix S_{test} $[N_{test} \times F]$
235 for the test set, the predicted fMRI activity $\hat{Y}_{\text{test}}$ $[N_{test} \times V]$ for the test sounds was obtained as:

236 $$\hat{Y}_{\text{test}} = S_{test}\ \tilde{C} \qquad (2)$$

237 Then, we computed for each stimulus s_k the correlation r_s between its predicted fMRI activity
238 $\hat{Y}_{\text{test}}(s)$ $[1 \times V]$ and all measured fMRI responses $Y_{test}(s)$ $[1 \times V]$. The rank of the correlation between
239 predicted and observed activity for stimulus s_k was used as a measure of the model's ability to
240 correctly match $Y_{\text{test}}(s)$ with its prediction $\hat{Y}_{\text{test}}(s)$. The rank was then normalized between 0 and 1 as
241 follows to obtain the sound identification score m for stimulus s (Santoro et al., 2014):

242 $$m_s = 1 - \frac{rank(r_s) - 1}{N_{test} - 1} \qquad (3)$$

243 Note that $m = 1$ indicates a correct match; $m = 0$ indicates that the predicted activity pattern for
244 stimulus s_i was least similar to the measured one among all stimuli. Normalized ranks (sound
245 identification scores) were computed for all stimuli in the test set, and the overall model's accuracy
246 was obtained as the mean of the sound identification scores across stimuli. Accuracies were converted
247 to *z*-scores via Fisher's transformation to reduce deviations from normality.

Topographical best-feature maps

The response profiles for temporal modulation, spectral modulation and frequency were computed as marginal sums of the estimated stimulus-activity mapping function *C* of the frequency-specific modulation model by summing across irrelevant dimensions. For example, to obtain the temporal modulation transfer function (tMTF), we summed across the spectral modulation and frequency dimension:

$$tMTF(\omega) = \sum_{\Omega} \sum_{f} C(\omega, \Omega, \mathrm{f}) \qquad (4)$$

To calculate profiles for the spectral modulation transfer function (sMTF) and frequency transfer function (fTF), we correspondingly summed across irrelevant dimensions. The voxels' characteristic values (CTM, CSM, CF) were defined as the maximum of the tMTF, sMTF and fTF, respectively. Cortical maps were generated by color-coding the voxels' preferred values and projecting them onto an inflated representation of the cortex. To obtain group maps, individual maps were transformed to FreeSurfer's *fsaverage5* space and averaged.

Multivariate decoding analysis: Model estimation

In the multivariate decoding analysis, we evaluated the fidelity with which regions of interest (ROIs) in auditory cortex encode acoustic features by estimating decoders. In addition to a region including the whole auditory cortex, we selected five ROIs based on anatomical criteria (using the FreeSurfer labels): Heschl's gyrus/sulcus (HG/HS), planum temporale (PT), planum polare (PP), superior temporal gyrus (STG) and sulcus (STS). For a control ROI analysis, we selected visual cortex (calcarine sulcus). For each individual in each ROI, a linear decoder was trained for every feature of the modulation space based on the training data only (Santoro et al., 2017). Each stimulus feature S_j [N_{train} ×

269 *1*] was modelled as a linear transformation of the multi-voxel response pattern Y_{train} [$N_{train} \times V$] plus a
270 bias term b_j and a noise term n [$N_{train} \times 1$] as follows:

271 $$S_j = Y_{train} C_j + b_j 1 + n \qquad (5)$$

272 where N_{train} is the number of sounds in the training set, V is the number of voxels, 1 is a [$N_{train} \times 1$]
273 vector of ones, and C_j is a [$V \times 1$] vector of model parameters, whose elements c_{ji} quantify the
274 contribution of voxel *i* to the encoding of feature *j*. All parameters in Equation (5) were estimated
275 within the same ridge-regression cross-validation scheme as Equation (1). Here, the regularization
276 parameter λ was determined independently for each feature by generalized cross validation using the
277 identical search grid as in the univariate encoding model (see above).

278 **Estimation of multi-voxel modulation transfer functions (MTFs)**

279 Decoders were estimated on the train runs (see above) and tested on the test runs. Given the
280 trained model $\tilde{C}$ [$V \times F$], and the patterns of fMRI activity for the test sounds Y_{test} [$N_{test} \times V$], the predicted
281 feature matrix activity $\hat{S}_{\text{test}}$ [$N_{test} \times F$] for the test sounds was calculated as:

282 $$\hat{S}_{\text{test}} = Y_{test}\, \tilde{C} \qquad (6)$$

283 Predictions for all features from the test sets were concatenated and decoders were assessed
284 individually by computing the Pearson's correlation coefficient (*r*) between a predicted and a given
285 actual stimulus feature. For decoding from the whole auditory cortex (ROIs), this resulted in 288 (125)
286 correlation coefficients, which represented the MTF. To obtain marginal profiles of the MTFs, we
287 averaged across the momentarily irrelevant dimensions, respectively.
288 For statistical testing, all correlation values were transformed into z-scores by applying Fisher's *z*-
289 transformation and pooled together. Statistical significance of the reconstruction accuracy was
290 assessed with permutation testing. For each stimulus feature we computed the null-distribution of

accuracies at the single-subject level by randomly permuting (1000 times) the stimulus labels of the test sounds and computing the correlation coefficient with the predicted features for each permutation. In order to preserve the temporal correlations among sounds, the predictors were phase scrambled. The empirical chance level of correlation r_{chance} was defined as the mean of the null distribution. The *p*-value was computed as the proportion of permutations that yielded a correlation equal to or more extreme than the empirical one. The reconstruction accuracies were z-scored relative to the empirical null distribution.

Post-hoc statistical analysis of marginal MTFs

Group marginal profiles of MTFs were obtained as the median of all individual marginal MTFs. To assess the statistical significance of the observations on the MTF's marginal profiles, we performed the following post-hoc analyses.

Selectivity Index. Using the marginal profiles of the MTFs, we quantified tuning selectivity by the selectivity index

$$SI = \frac{\max_{j\in[k]}(r_j) - \frac{1}{k-1}\sum_{j=1, j\neq j_{max}}^{k} r_j}{\max_{j\in[k]}(r_j) + \frac{1}{k-1}\sum_{j=1, j\neq j_{max}}^{k} r_j} \qquad (7)$$

Where r_j is the reconstruction accuracy at feature *j*, *k* is the number of features, j_{max} is the index of the maximal reconstruction accuracy. *SI* was calculated for each acoustic dimension separately (i.e., rate, scale or frequency) and compared between age groups using an exact permutation test by calculating the mean group difference and permuting the age group labels 10000 times.

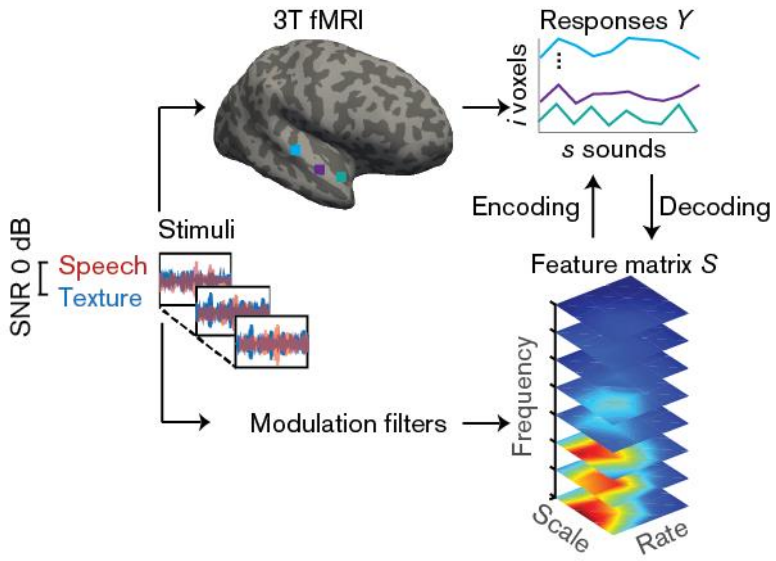


Figure 1. Experimental design and analysis pipeline. Participants listened to a story embedded in sound textures at an SNR of 0 dB while we acquired 3T-fMRI data. The sounds were decomposed into their modulation spectrum (Chi et al., 2005). In a first *univariate encoding* approach, we calculated a modulation transfer function (MTF) for each individual voxel. In a second *multivariate decoding* approach, linear decoders were trained on the response patterns in auditory cortex for each feature of the modulation representation. Predictions were then tested on a left-out testing data set.

317 **Results**

318 fMRI responses to sounds were modelled using both model-based encoding (Santoro et al., 2014)
319 and decoding (Santoro et al., 2017). In brief, sounds were decomposed into acoustic features, that is,
320 the frequency-dependent spectro-temporal modulation content (Chi et al., 2005). Analyses were
321 restricted to voxels in an anatomic mask of the auditory cortex. In a fourfold cross-validation
322 procedure, we used ridge regression to derive single- and multi-voxel MTFs.

323 **Encoding results**

324 We compared three models of hemodynamic responses to sound. The models respectively
325 describe the fMRI responses to the (1) speech stream (foreground), (2) textures (background), or (3) the
326 mixture of both streams.

327 **Sound identification accuracies.** In a first encoding analysis, we estimated a modulation transfer
328 function (MTF) for each voxel based on a subset of fMRI data (training) for each model. We then
329 assessed the ability of these models to accurately predict the fMRI responses to sounds of a new,
330 independent data set (testing data). We quantified prediction accuracy by means of a sound
331 identification analysis (see Materials and Methods).

332 Sound identification accuracy was higher than chance level (0.5) for all models and age groups (Fig.
333 2). Besides being the plausible model (as it reflects the actual acoustic input to the participant), the
334 sound mixture model had the highest sound identification accuracies (Fig. 2a) and thus described
335 most accurately the fMRI responses to sounds. All subsequent encoding and decoding analyses were
336 based on this mixture model. Young participants had significantly higher identification accuracies for
337 mixture than old participants. There was no significant correlation between chronological age and
338 sound identification accuracy for mixture within the young or older group (Fig. 2b).

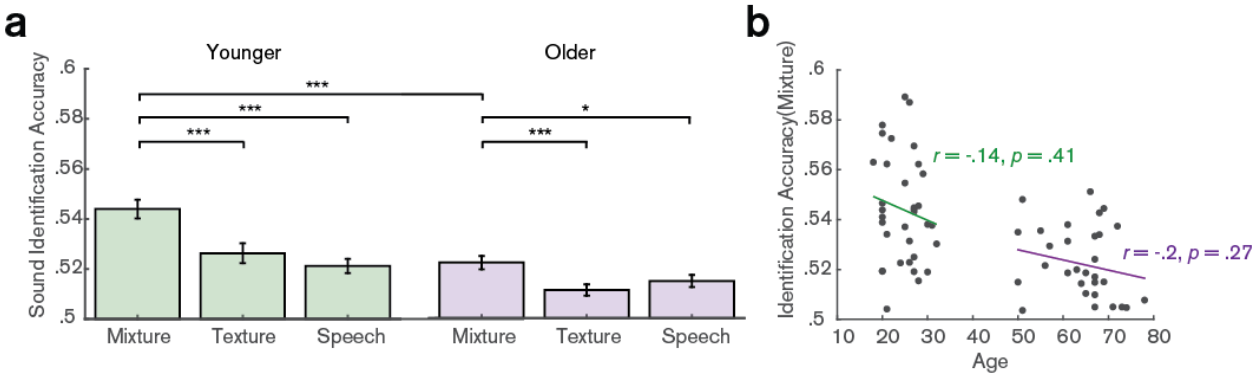


Figure 2. Sound identification accuracies for the different encoding models. (a) Accuracies for sound mixture, texture (background stream) and speech (foreground stream), for young (left, green) and older participants (right, violet). Bars indicate the mean (±SEM) of all participants per age group. Accuracies are normalized between 0 and 1; 0 denotes that the predicted activity pattern for a given stimulus was least similar to the measured one among all test stimuli; 1 denotes correct identification; chance level is 0.5. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, exact permutation test. (b) Scatter plot of identification accuracies for sound mixture as a function of age.

Topographical best-feature maps. Best features maps were obtained from univariate encoding models by marginalizing the single-voxel MTFs for the dimension of interest. In particular, tonotopic maps were obtained by averaging over the rate and scale dimension and assigning the frequency with the maximal response to a given voxel. Tonotopic maps showed the typical mirror-symmetric frequency gradients along Heschl's gyrus in both young and aged auditory cortex, irrespective of age (Fig. 3). We did not find significant age-group differences using two-sample voxel-wise *t* -tests (corrected for multiple comparisons). In fact, tonotopic group maps were correlated, indicating that topography for best frequency was preserved in older adults (Fig. 3a).

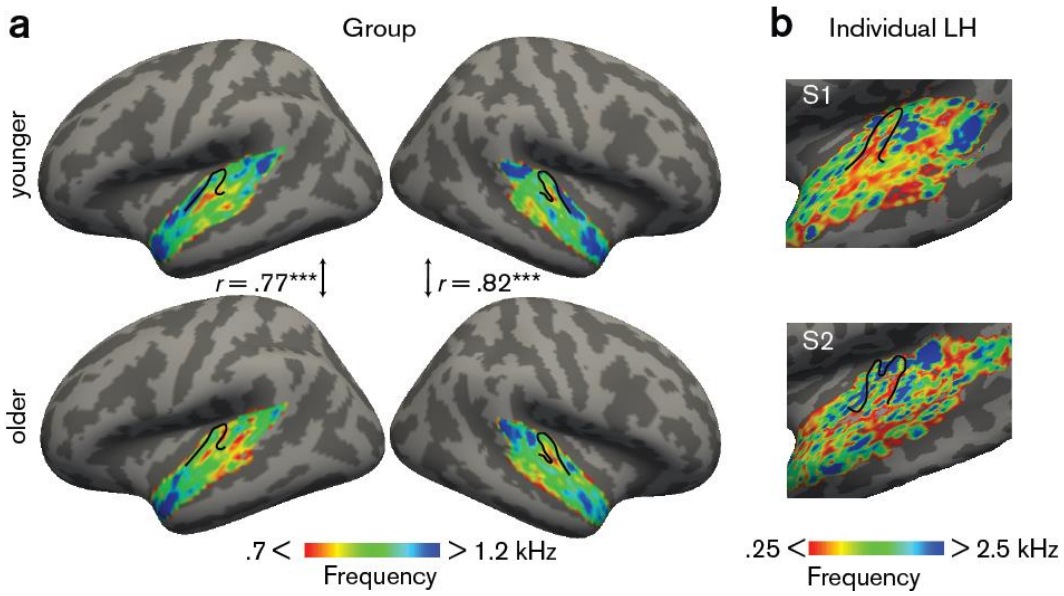


Figure 3. Tonotopic maps were derived by marginalizing modulation transfer functions (MTFs) for frequency and assigning the feature with the maximal response to a given voxel. (a) Group maps were obtained as median across subjects and are shown on FreeSurfer's *fsaverage5* template. Tonotopic maps for young group (upper panel) and older group (bottom panel) are correlated in both the right and left hemisphere. (b) Exemplary individual tonotopic maps for left hemisphere of a young (upper panel) and old participant (bottom panel) are displayed on individual surfaces. Black outlines indicate Heschl's gyrus. *** $p < 0.005$.

Topographical best-feature maps for rate and scale were more complex. Similarly, we did not find significant voxel-wise age-group differences using two-sample *t*-tests (corrected for multiple comparisons). On the contrary, both for temporal and spectral modulations, group maps were correlated between age groups (Fig. 4). For best temporal modulation maps, we observed a medial-to-lateral gradient for increasing temporal rate (Fig. 4a). For best spectral modulation maps, the locus of highest spectral resolution coincided with the low frequency region on Heschl's gyrus (Fig. 3a, Fig. 4b).

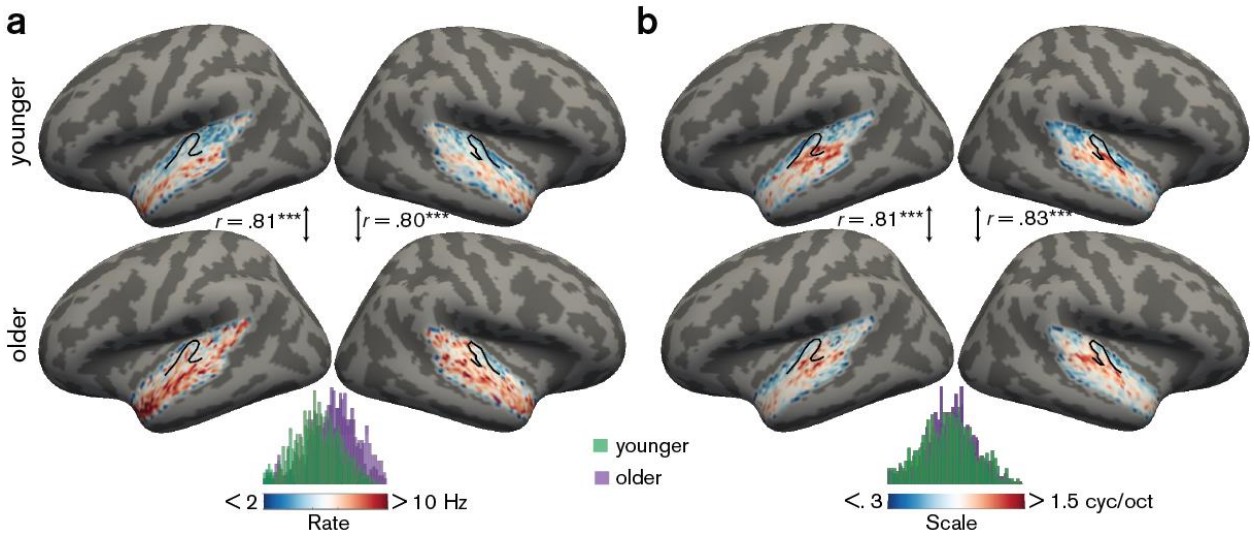


Figure 4. Group best-feature maps for temporal rate (a) and spectral scale (b). Group maps were obtained as median across subjects and shown on FreeSurfer's *fsaverage5* template. Black outlines indicate Heschl's gyrus. Topographic organization in young (upper panel) and older group (bottom panel) is significantly correlated (Pearson's correlation coefficient) for best temporal rate (a) and best spectral scale (b). Histograms of best feature values are shown for the young group (green) and older group (violet) above the color bars. *** $p < 0.005$.

Decoding results

Feature reconstruction from auditory cortex. Cortical sensitivity to acoustic features was quantified using model-based decoding. Decoders were trained on a subset of the data (training) for each acoustic feature separately (Fig.1). Decoding accuracies were obtained as Pearson's *r* between predicted and actual acoustic features in the testing data. Results were compared across age groups using exact permutation tests. Decoding yielded highest accuracies at frequencies of 230–580 Hz (median r = 0.52), scales of 0.25 cyc/oct (median r = 0.5), and rates of 4–8 Hz (median r = 0.42), irrespective of age (Fig. 5 a,b).

Selectivity index. A key observation were age differences in selectivity of tuning to temporal modulations (as quantified by the selectivity index SI, see materials and methods). Selectivity of temporal rate coding was higher in young than in older participants (mean SI difference = 0.01, p =

385 0.004 [permutation test], Cohen's d = 0.68, Fig. 5c, left). A lower selectivity index is consistent with a
386 broadening of temporal tuning in the aged auditory cortex. Notably, in the older group, the rate
387 selectivity index was significantly correlated with chronological age (i.e., Fig. 5d, left), but was
388 uncorrelated with hearing loss (pure tone average, PTA; Fig. 5d, right).

389 In a control analysis, we ensured that these results cannot be explained by the overall slightly
390 worse model fits in the older group as quantified by sound identification accuracy (as evident in Fig.
391 2): In a multiple regression predicting older participants' rate selectivity index from chronological age
392 with covariates PTA and sound identification accuracy, neither PTA ($t(26)$ = 1.21, p = 0.24) nor sound
393 identification accuracy ($t(26)$ = -0.52, p = 0.6) were significant, while age was significant ($t(26)$ = -2.09, p
394 = 0.04; *adjusted* R^2 = 0.11).

395 Also, this key result of a broadened tuning to temporal rate was corroborated by an even simpler
396 measure of selectivity: The variance in accuracy across the six temporal rate bins (1–32 Hz) was lower
397 in the older than the younger participants (mean difference in variance = 3.7 x 10^{-4}, p < 0.001
398 [permutation test], Cohen's d = 0.93, Fig. S3a, left panel). In the older group, the variance in rate was
399 significantly correlated with the rate selectivity index and chronological age but was uncorrelated
400 with hearing loss (Fig. S3b). A similar control analysis showed that in a multiple regression predicting
401 older participants' variance in rate from chronological age with covariates PTA and sound
402 identification accuracy, neither PTA ($t(26)$ = 0.88, p = 0.38) nor sound identification accuracy ($t(26)$ =
403 1.23, p = 0.22) were significant, while age was significant ($t(26)$ = -3.4, p = 0.002; *adjusted* R^2 = 0.39).

404

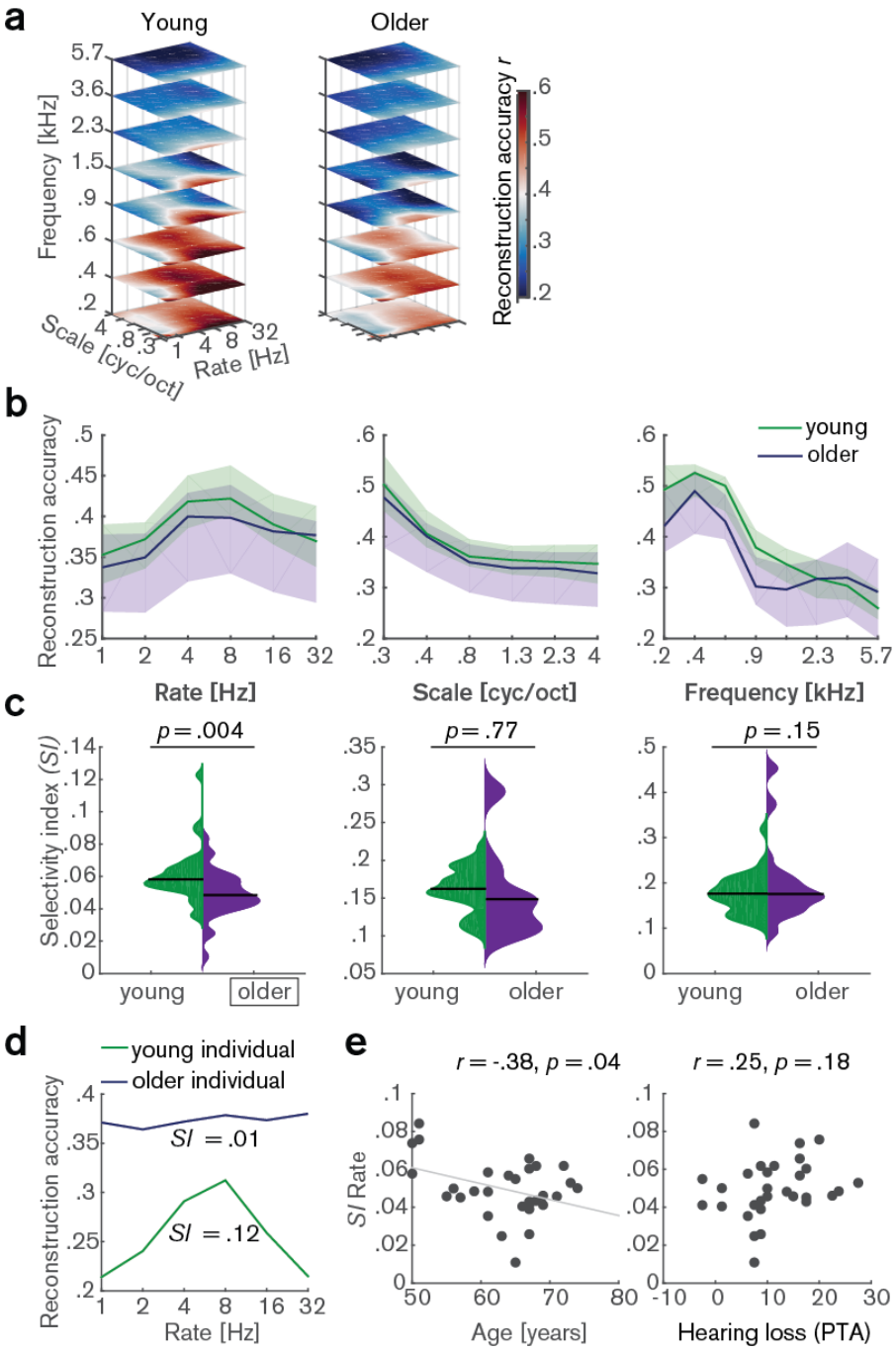


Figure 5. Decoding results for auditory cortex. (a) Auditory cortical MTFs for the young and older group and (b) median and IQR of the MTFs' marginal profiles for rate, scale and frequency. (c) We quantified tuning selectivity by the selectivity index (see Eq. [7]), which was compared per acoustic dimension between age groups using an exact permutation test. (d) For illustration purposes, we show the marginal profiles for rate for a young (green) and older individual (violet) with the most extreme *SI* values. (e) Correlation of *SI* for rate with PTA and chronological age in the older group.

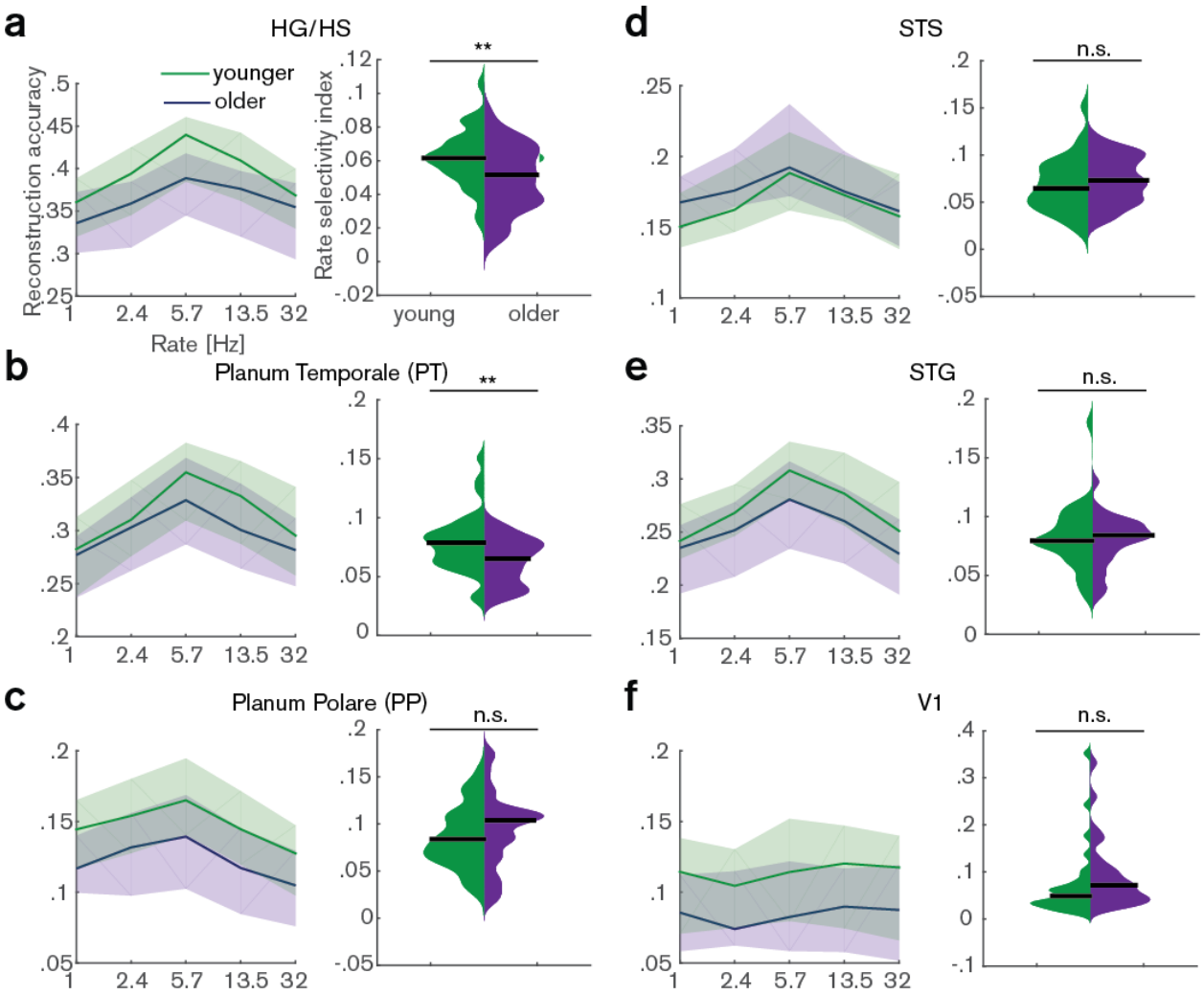
412 **Feature reconstruction from auditory subregions.** To test the differential sensitivity of auditory
413 cortical subregions to acoustic features, we ran the decoding analyses in the following regions of
414 interest (ROIs) derived from the FreeSurfer atlas (Destrieux et al., 2010): Heschl's gyrus and sulcus
415 (HG/HS), planum polare (PP), planum temporale (PT), superior temporal gyrus (STG) and superior
416 temporal sulcus (STS). As a control analysis, we tried to decode from primary visual cortex (V1,
417 calcarine sulcus).

418 Note that reconstruction accuracies in all ROIs but V1 were significant (permutation tests; for the z-
419 scored reconstruction accuracies in an exemplary region HG/HS see Fig. S4). A decrease in the
420 selectivity index with age was confirmed for HG/HS (mean *SI* difference = 0.012, p = 0.009
421 [permutation test]) and PT (mean *SI* difference = 0.016, p = 0.005 [permutation test]).

422 In contrast, there were no such age differences in temporal selectivity in PP (mean *SI* difference =
423 0.007, p = 0.42), STG (*SI* difference = 0.002, p = 0.72), STS (*SI* difference = 0.006, p = 0.38) or the control
424 region V1 (*SI* difference = 0.01, p = 0.89; Fig 6.).

425 Qualitatively, this observation was confirmed by a control analysis, where we derived topographic
426 maps of the rate selectivity index from the encoding models. To this end, we marginalized MTFs for
427 rate and calculated the *SI* for a given voxel. The topographic distribution confirmed highest voxel-wise
428 *SI* for rate in area PT of the younger group (Fig. S5).

429 These results indicate that age-related changes in temporal selectivity selectively occur in primary
430 auditory cortex and the planum temporale, but not surrounding belt areas. For spectral scale, we also
431 found a decreased *SI* in older adults in HG/HS (*SI* difference = 0.012, p = 0.01) and PT (*SI* difference =
432 0.016, p = 0.006), but none of the other ROIs (Fig. S6). For frequency, we did not find any age
433 differences in selectivity index at all (Fig. S6).



434

435 **Figure 6. Decoding of temporal rate from regions of interest (ROIs).** Median reconstruction accuracy
436 profiles (left) and selectivity index (right) per age group for temporal rate in (a) Heschl's gyrus and sulcus, (b)
437 planum temporale, (c) planum polare, (d) superior temporal sulcus, (e) superior temporal gyrus and (f) calcarine
438 sulcus. The selectivity index was compared per ROI between age groups using an exact permutation test. HG/HS:
439 Heschl's gyrus, STS: superior temporal sulcus, STG: superior temporal gyrus, V1: primary visual cortex. ** $p < 0.01$,
440 n.s. $p > 0.05$.

441

Discussion

The cross-sectional approach used encoding and decoding models of hemodynamic responses, to compare the spectro-temporal fidelity in auditory cortex of younger and older listeners. We pursued the long-standing but largely understudied question whether temporal and/or spectral modulations, which are key acoustic dimensions of behaviorally relevant signals such as speech and music, are represented in a less differentiated fashion in the aging auditory cortex.

We find the large-scale topographic organization of acoustic features entirely preserved in the aged auditory cortex, and best feature maps for frequency, scale and rate accordingly were correlated very highly between age groups.

However, age-related differences are evident in the cortical sensitivity to mainly temporal–modulation features as quantified by the decoding analysis. Decoding of temporal rate is most accurate at slow rates of 4–8 Hz irrespective of age, but the tuning to temporal rate is significantly sharper in young than in older participants. Also, amongst older adults, rate selectivity correlates with biological age. Importantly, the age-related rate selectivity difference was driven by primary auditory (HG) and adjacent area PT, while secondary auditory areas PP, STG, and STS did not show age differences.

Preserved topographic organization in aging auditory cortex

Using natural stimuli, we revealed topographic maps of acoustic feature preference for frequency, spectral and temporal modulations. Thus, we show that topographic maps of feature preference can be observed under approximately natural hearing situations. Tonotopic maps exhibited the well-established mirror-symmetric high-low-high frequency gradients across Heschl's gyrus and adjacent areas (Fig. 3) consistent with abundant evidence in human (Formisano et al., 2003, Talavage et al., 2004, Humphries et al., 2010, Woods et al., 2010, Da Costa et al., 2011, Striem-Amit et al., 2011, Langers

465 and van Dijk, 2012, Moerel et al., 2012, 2014) and non-human primates (Merzenich and Brugge, 1973,
466 Morel et al., 1993, Kosaki et al., 1997, Rauschecker et al., 1997, Bendor and Wang, 2008, Baumann et al.,
467 2013, Joly et al., 2014, Baumann et al., 2015). Notably, tonotopic maps in younger versus older adults
468 were highly correlated, indicating topographical stability of tonotopic organization across age groups.

469 The well-demonstrated existence of tonotopic maps notwithstanding, a topographic organization
470 for temporal or spectral modulations in auditory cortex is less clear (Schonwiesner and Zatorre, 2009).
471 Here, in both young and older adults we observed a medial-to-lateral low-to-high gradient for
472 temporal modulations across the superior temporal plane (Fig. 4a), while for spectral modulations, the
473 locus with the highest resolution (~1.5 cyc/oct) was observed in lateral Heschl's gyrus (Fig. 4b). A
474 topographic representation for modulation rate has previously been observed in the inferior colliculus
475 (IC) (Schreiner and Langner, 1988) and the auditory cortex (Baumann et al., 2015) where fast temporal
476 acoustic information was preferably encoded in caudal auditory regions (Camalier et al., 2012,
477 Kusmierek and Rauschecker, 2014) and slow temporal information in rostral areas R and RT (Liang et
478 al., 2002, Bendor and Wang, 2008).

479 Whereas earlier human fMRI studies failed to show a clear topography for modulation rate (Giraud
480 et al., 2000, Schonwiesner and Zatorre, 2009, Overath et al., 2012, Leaver and Rauschecker, 2016),
481 recent human fMRI (Santoro et al., 2014) and electrocorticography (ECoG) studies (Hullett et al., 2016)
482 proposed the presence of a posterior-to-anterior high-to-low rate gradient in human auditory cortex.
483 However, comparison of those studies is hampered by the differences in coverage and ranges of
484 temporal modulations: While Hullett et al. (2016) examined representation of slow temporal
485 modulation rates present in speech (1–2 Hz) only along the STG with ECoG, Santoro at al. (2014)
486 analyzed rates of 1–27 Hz, whereas Baumann et al. (2015) presented rates of up to 512 Hz with more
487 comprehensive coverage using fMRI.

488 **Functional mapping from natural sounds is feasible using conventional MR field strength in**
489 **aging populations**

490 While previous studies had presented single natural sounds (Erb et al., 2018), or even synthetic
491 sounds such as AM stimuli (Baumann et al., 2015) or dynamic ripples (Schonwiesner and Zatorre,
492 2009), here, we presented a continuous stream of speech embedded in an acoustically rich
493 background of sound textures. As an important foundation for the main conclusions of this
494 manuscript, we here demonstrate the capacity to derive meaningful tonotopic maps from natural
495 stimulus conditions, at conventional field strengths (3 T) and in special populations (older
496 participants).

497 Although an inherent caveat of the use of natural stimulus conditions is the difficulty of measuring
498 psychoacoustic performance, this experimental design has the advantage that it closely approximates
499 natural listening conditions (Hamilton and Huth, 2018). As synthetic sounds lack both the behavioral
500 relevance and the statistical structure of natural sounds, they activate auditory cortex differently than
501 natural stimuli (Theunissen and Elie, 2014) (Theunissen et al., 2000, Bitterman et al., 2008).

502 Methodologically, our encoding approach relies on the assumption that acoustic features that
503 maximally activate single voxels are more accurately encoded. However, higher responses may not
504 necessarily mean better encoding. The second, complementary decoding analysis is based on
505 measures of information rather than activation levels. The reconstruction accuracy explicitly quantifies
506 the amount of information about a set of stimulus features that is available in a region of interest.
507 Accurate reconstruction of acoustic features in test sounds indicates that these features are
508 reproducibly mapped into distinct spatial patterns. As data from individual voxels are jointly
509 modelled, this combined analysis of signals from multiple voxels increases the sensitivity for stimulus
510 information that may be represented in patterns of responses, rather than in individual voxels.

511 **Broadened temporal tuning in aging auditory cortex**

512 Under the assumption that the reconstruction accuracy reflects tuning properties of neuronal
513 populations, the observed decrease in selectivity reveals a broadening of temporal tuning functions
514 with age. This finding provides mechanistic insights into the changes of temporal processing in the
515 aging auditory cortex: The broadened temporal tuning, reminiscent of an age-related decrease in
516 temporal precision in the midbrain (Anderson et al., 2012), may underlie the difficulties older adults
517 experience in speech comprehension in noisy environments (Anderson et al., 2011). While the present
518 study lacks more direct measures of encoding along the auditory pathway, it is important to note that
519 rate selectivity in the older listeners was statistically unrelated to their pure-tone audiograms (this
520 holds true for both presented measures of tuning, the selectivity index [Fig.5e] and the variance across
521 temporal rates, [Fig. S3b]). Thus, the broadening of temporal tuning is not explicable merely by
522 peripheral hearing loss. Future studies need to carefully assess the consequences of altered cortical
523 temporal tuning on behavioral and psychoacoustic performance (Erb et al., 2019, Flinker et al., 2019,
524 Holmes and Griffiths, 2019), a measure that was precluded by the use of natural sounds in the current
525 study.

526 The observation of decreased temporal selectivity may seem at odds with previous accounts of
527 elevated cortical responses to the temporal speech envelope in aging. Evidence for better
528 reconstruction of temporal envelopes from older subjects' EEG signals (Presacco et al., 2016) and
529 larger N1 amplitudes in older adults with widened auditory filters (Herrmann et al., 2016, Henry et al.,
530 2017) suggest an enhanced cortical envelope representation. However, an inflated cortical
531 representation of the speech envelope has been argued to be rather a maladaptative response as
532 compensation of the lack of temporal precision in the midbrain (Anderson et al., 2012).

533 Neuronal coding of temporal modulations follows two distinct principles, that is rate and temporal
534 coding (Joris et al., 2004). While rate coding constitutes a variation of overall spike rate as a function of

535 modulation frequency, temporal coding is achieved through phase-locking to the stimulus envelope.
536 Fast temporal modulations (> 50 Hz) are typically encoded through the rate code, while neurons
537 tuned to slow temporal rates (< 50 Hz) mostly demonstrate synchronization to the sounds'
538 modulations (temporal code). It remains unclear how hemodynamic responses reflect these two types
539 of neuronal coding and how those are in turn related to the observed age-related broadening of
540 tuning functions. Evidence from electrophysiology suggests a change in coding strategy across
541 neuronal populations in A1 from less temporal coding towards more rate coding of temporal
542 modulations in aged animals (Ng and Recanzone, 2018).

543 Interestingly, the observed age-related decrease in selectivity to temporal modulations was
544 restricted to areas HG/HS and PT, the putative homologues of areas A1 and CL in the macaque (Kaas
545 and Hackett, 2000). Both areas have been shown to exhibit altered temporal encoding in older
546 macaques (Ng and Recanzone, 2018). In a direct comparison of the vector strengths (a metric of
547 periodicity of the neuronal response to a modulated signal) as a function of the inter-stimulus interval,
548 responses differed in young A1 and CL neurons such that A1 neurons had higher vector strength. This
549 is consistent with the notion that CL neurons process predominantly spatial (rather than temporal)
550 information. Therefore, so the theory, CL neurons would not necessarily need to pass on the temporal
551 coding of A1 neurons. This difference between belt and core areas was abandoned in the older
552 animals, where aged A1 were similar to both the aged and young CL neurons, indicating that the
553 temporal fidelity of A1 responses decreases with age.

554 However, a good temporal fidelity of A1 neurons is thought to be critical for temporal processing.
555 We speculate that such an age-related decline in temporal processing may be linked to an imbalance
556 of excitation and inhibition observed in aging (Voytek et al., 2015). Inhibitory (e.g. GABAergic or
557 glycinergic) neurotransmitters have been shown to increase response synchrony to modulated stimuli
558 in both the cochlear nucleus and the inferior colliculus (Koch and Grothe, 1998, Backoff et al., 1999),

559 for review see (Caspary et al., 2008), suggesting that inhibitory neurotransmission is of particular
560 importance to precise neural timing and, thus, the adequate tracking of temporal sound features.

561 An inherent challenge to age group comparisons is to distinguish any unspecific decrease of SNR in
562 the data of older participants (e.g., due to vascular changes that impact the BOLD response (Garrett et
563 al., 2017) or due to the typical increase in movement artefacts) from specific age-related sensory
564 processing changes. Most reassuringly, however, a multiple regression model taking into account
565 overall goodness of the decoding model fit (i.e., sound identification accuracy) did not alter the
566 observed relationship between age and temporal rate selectivity.

567 **Conclusions**

568 The present study aimed to narrow the gap between recent progress in modelling the neural
569 processing of complex features from natural stimuli on the one hand, and potential senescent change
570 in these stimulus representations in the aging auditory cortex on the other hand. Although the large-
571 scale topographic organization of acoustic features appears preserved in the auditory cortex of older
572 compared to younger listeners, age-related differences in the marginal profiles of multi-voxel MTFs
573 were evident. Tuning to slow temporal rates which abounds in natural sounds and especially in
574 speech (Erb et al., 2018) was markedly sharper in young compared to older participants. Consistent
575 with previous findings in the macaque, these results suggest that temporal rate selectivity in auditory
576 cortex declines in normal aging. The specificity of this decline, confined to primary auditory and
577 adjacent areas and sparing spectral as well as cochleotopic representations, opens a new lead in the
578 ongoing hunt for tractable neurobiological signatures of older adults' widely observed deficits in
579 speech comprehension in noisy environments.

Acknowledgements

This research was funded by an ERC consolidator Grant 646696 (AUDADAPT) to J.O and the German Research Foundation (DFG; OB 352/2-1). Martin Göttlich helped with MR sequences, Anne Herrmann, Malte Naujokat, Clara Mergner and Anne Ruhe helped acquire the data. We thank the members of the Auditory Cognition group for useful discussions on the present study.

Competing interests

The authors declare no competing interests.

Data availability

MRI data and custom code to reproduce all essential findings will be publicly available on the Open Science Framework (OSF) upon publication.

References

Alavash M, Tune S, & Obleser J (2019). Modular reconfiguration of an auditory-control brain network supports adaptive listening behavior. Proceedings of the National Academy of Sciences of the United States of America 116(2):660-669.

Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2012) Aging affects neural precision of speech encoding. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:14156-14164.

Anderson S, Parbery-Clark A, Yi HG, Kraus N (2011) A neural basis of speech-in-noise perception in older adults. Ear and hearing 32:750-757.

Backoff PM, Shadduck Palombi P, Caspary DM (1999) Gamma-aminobutyric acidergic and glycinergic inputs shape coding of amplitude modulation in the chinchilla cochlear nucleus. Hearing research 134:77-88.

Baumann S, Joly O, Rees A, Petkov CI, Sun L, Thiele A, Griffiths TD (2015) The topography of frequency and time representation in primate auditory cortices. eLife 4.

Baumann S, Petkov CI, Griffiths TD (2013) A unified framework for the organization of the primate auditory cortex. Front Syst Neurosci 7:11.

Bendor D, Wang X (2008) Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. Journal of neurophysiology 100:888-906.

Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature 451:197-201.

Camalier CR, D'Angelo WR, Sterbing-D'Angelo SJ, de la Mothe LA, Hackett TA (2012) Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in

primates. Proceedings of the National Academy of Sciences of the United States of America 109:18168-18173.
Caspary DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. The Journal of experimental biology 211:1781-1791.
Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. The Journal of the Acoustical Society of America 118:887-906.
Da Costa S, van der Zwaag W, Marques JP, Frackowiak RS, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:14067-14075.
Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53:1-15.
Erb J, Armendariz M, De Martino F, Goebel R, Vanduffel W, Formisano E (2018) Homology and Specificity of Natural Sound-Encoding in Human and Monkey Auditory Cortex. Cerebral cortex.
Erb J, Ludwig AA, Kunke D, Fuchs M, Obleser J (2019) Temporal Sensitivity Measured Shortly After Cochlear Implantation Predicts 6-Month Speech Recognition Outcome. Ear and hearing 40:27-33.
Erb J, Obleser J (2013) Upregulation of cognitive control networks in older adults' speech comprehension. Front Syst Neurosci 7:116.
Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ (2019) fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods 16:111-116.
Flinker A, Doyle WK, Mehta AD, Devinsky O, Poeppel D (2019) Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. Nat Hum Behav 3:393-405.
Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40:859-869.
Garrett DD, Lindenberger U, Hoge RD, Gauthier CJ (2017) Age differences in brain signal variability are robust to multiple vascular controls. Scientific reports 7:10149.
Gates GA, Mills JH (2005) Presbycusis. Lancet 366:1111-1120.
Giraud AL, 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 84:1588-1598.
Golub G, Heath M, Wahba G (1979) Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. Technometrics 21:215-223.
Gordon-Salant S, Yeni-Komshian GH, Fitzgibbons PJ, Barrett J (2006) Age-related differences in identification and discrimination of temporal cues in speech segments. The Journal of the Acoustical Society of America 119:2455-2466.
Hamilton LS, Huth AG (2018) The revolution will not be controlled: natural stimuli in speech neuroscience. Language, cognition and neuroscience.
Henry MJ, Herrmann B, Kunke D, Obleser J (2017) Aging affects the balance of neural entrainment and top-down neural modulation in the listening brain. Nature communications 8:15801.
Herrmann B, Henry MJ, Johnsrude IS, Obleser J (2016) Altered temporal dynamics of neural adaptation in the aging human auditory cortex. Neurobiology of aging 45:10-22.
Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55-67.
Holmes E, Griffiths TD (2019) 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. Scientific reports 9:16771.
Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE, Chang EF (2016) Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. The Journal of neuroscience : the official journal of the Society for Neuroscience 36:2014-2026.

661 Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex.
662 NeuroImage 50:1202-1211.
663 Joly O, Baumann S, Balezeau F, Thiele A, Griffiths TD (2014) Merging functional and structural
664 properties of the monkey auditory cortex. Frontiers in neuroscience 8:198.
665 Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev
666 84:541-577.
667 Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates.
668 Proceedings of the National Academy of Sciences of the United States of America 97:11793-
669 11799.
670 Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity.
671 Nature 452:352-355.
672 Koch U, Grothe B (1998) GABAergic and glycinergic inhibition sharpens tuning for frequency
673 modulations in the inferior colliculus of the big brown bat. Journal of neurophysiology 80:71-
674 82.
675 Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields
676 delineated by parvalbumin immunoreactivity in macaque monkeys. The Journal of
677 comparative neurology 386:304-316.
678 Kusmierek P, Rauschecker JP (2014) Selectivity for space and time in early areas of the auditory dorsal
679 stream in the rhesus monkey. Journal of neurophysiology 111:1671-1685.
680 Lalwani P, Gagnon H, Cassady K, Simmonite M, Peltier S, Seidler RD, Taylor SF, Weissman DH, Polk TA
681 (2019) Neural distinctiveness declines with age in auditory cortex and is associated with
682 auditory GABA levels. NeuroImage 201:116033.
683 Langers DR, van Dijk P (2012) Mapping the tonotopic organization in human auditory cortex with
684 minimally salient acoustic stimulation. Cerebral cortex 22:2024-2038.
685 Leaver AM, Rauschecker JP (2016) Functional Topography of Human Auditory Cortex. The Journal of
686 neuroscience : the official journal of the Society for Neuroscience 36:1416-1428.
687 Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency
688 modulations in the primary auditory cortex of awake primates. Journal of neurophysiology
689 87:2237-2261.
690 McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery:
691 evidence from sound synthesis. Neuron 71:926-940.
692 Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal
693 plane of the macaque monkey. Brain research 50:275-296.
694 Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex:
695 tonotopy, spectral tuning, and relation to voice sensitivity. The Journal of neuroscience : the
696 official journal of the Society for Neuroscience 32:14205-14216.
697 Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human
698 auditory cortical areas. Frontiers in neuroscience 8:225.
699 Moerel M, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Processing of
700 natural sounds: characterization of multipeak spectral tuning in human auditory cortex. The
701 Journal of neuroscience : the official journal of the Society for Neuroscience 33:11888-11898.
702 Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of
703 auditory cortex in macaque monkeys. The Journal of comparative neurology 335:437-459.
704 Ng CW, Recanzone GH (2018) Age-Related Changes in Temporal Processing of Rapidly-Presented
705 Sound Sequences in the Macaque Auditory Cortex. Cerebral cortex 28:3775-3796.
706 Overath T, Zhang Y, Sanes DH, Poeppel D (2012) Sensitivity to temporal modulation rate and spectral
707 bandwidth in the human auditory system: fMRI evidence. Journal of neurophysiology
708 107:2042-2056.

709 Park DC, Polk TA, Park R, Minear M, Savage A, Smith MR (2004) Aging reduces neural specialization in
710 ventral visual cortex. Proceedings of the National Academy of Sciences of the United States of
711 America 101:13091-13095.
712 Peelle JE, Wingfield A (2016) The Neural Consequences of Age-Related Hearing Loss. Trends Neurosci
713 39:486-497.
714 Presacco A, Simon JZ, Anderson S (2016) Evidence of degraded representation of speech in noise, in
715 the aging midbrain and cortex. Journal of neurophysiology 116:2346-2355.
716 Rauschecker JP, Tian B, Pons T, Mishkin M (1997) Serial and parallel processing in rhesus monkey
717 auditory cortex. The Journal of comparative neurology 382:89-103.
718 Recanzone G (2018) The effects of aging on auditory cortical function. Hearing research 366:99-105.
719 Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (2014) Encoding of
720 natural sounds at multiple spectral and temporal resolutions in the human auditory cortex.
721 PLoS computational biology 10:e1003412.
722 Santoro R, Moerel M, De Martino F, Valente G, Ugurbil K, Yacoub E, Formisano E (2017) Reconstructing
723 the spectrotemporal modulations of real-life sounds from fMRI response patterns. Proceedings
724 of the National Academy of Sciences of the United States of America 114:4799-4804.
725 Schonwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in
726 the human auditory cortex measured with high-resolution fMRI. Proceedings of the National
727 Academy of Sciences of the United States of America 106:14611-14616.
728 Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical
729 organization. Journal of neurophysiology 60:1823-1840.
730 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily
731 temporal cues. Science 270:303-304.
732 Snell KB, Mapes FM, Hickman ED, Frisina DR (2002) Word recognition in competing babble and the
733 effects of age, temporal processing, and absolute sensitivity. The Journal of the Acoustical
734 Society of America 112:720-727.
735 Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical
736 fields obtained with phase-encoding FMRI. PloS one 6:e17832.
737 Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in
738 human auditory cortex revealed by progressions of frequency sensitivity. Journal of
739 neurophysiology 91:1282-1296.
740 Theunissen FE, Elie JE (2014) Neural processing of natural sounds. Nature reviews Neuroscience
741 15:355-366.
742 Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory
743 neurons obtained using natural sounds. The Journal of neuroscience : the official journal of the
744 Society for Neuroscience 20:2315-2331.
745 Vaden KI, Jr., Eckert MA, Dubno JR, Harris KC (2019) Cingulo-opercular adaptive control for younger
746 and older adults during a challenging gap detection task. Journal of neuroscience research.
747 Vaden KI, Jr., Kuchinsky SE, Ahlstrom JB, Dubno JR, Eckert MA (2015) Cortical activity predicts which
748 older adults recognize speech in noise and when. The Journal of neuroscience : the official
749 journal of the Society for Neuroscience 35:3929-3937.
750 Voytek B, Kramer MA, Case J, Lepage KQ, Tempesta ZR, Knight RT, Gazzaley A (2015) Age-Related
751 Changes in 1/f Neural Electrophysiological Noise. The Journal of neuroscience : the official
752 journal of the Society for Neuroscience 35:13257-13265.
753 Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X (2010) Functional properties of
754 human auditory cortical fields. Front Syst Neurosci 4:155.

Supplementary Information for

Temporal selectivity declines in the aging human auditory cortex

Julia Erb[1], Lea-Maria Schmitt[1], Jonas Obleser[1]

[1]Institute of Psychology, University of Lübeck, Maria-Goeppert-Str. 9A, 23562 Lübeck, Germany

*Correspondence: Julia Erb, Maria-Goeppert-Str. 9a, 23562 Lübeck, Germany
Phone: +49 451 3101 3631, e-mail: julia.erb@uni-luebeck.de

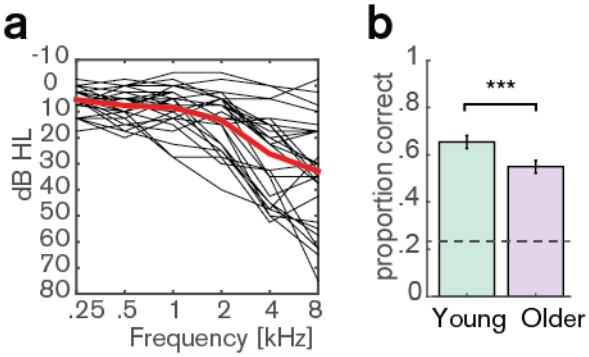


Figure S1. (a) **Audiograms** for older participants (averaged over left and right ear). (b) **Average behavioral performance** in the scanner in a four-choice task on the story content (3 questions after each run) for young (green) and older participants (violet). Chance level was 0.25 proportion correct. Young participants performed significantly better than older participants (mean proportion correct difference =0.1, $p = 0.009$ [permutation test], Cohen's $d = 0.65$).

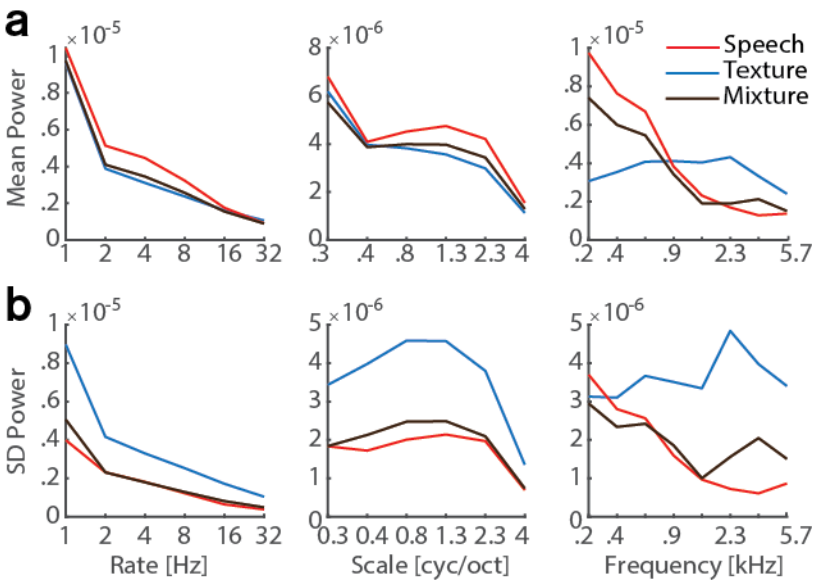


Figure S2. Modulation spectra of the stimuli. (a) Average of the marginalized modulation spectra for the stimulus streams separately and their mixture. (b) Standard deviation (SD) of the marginalized modulation spectra.

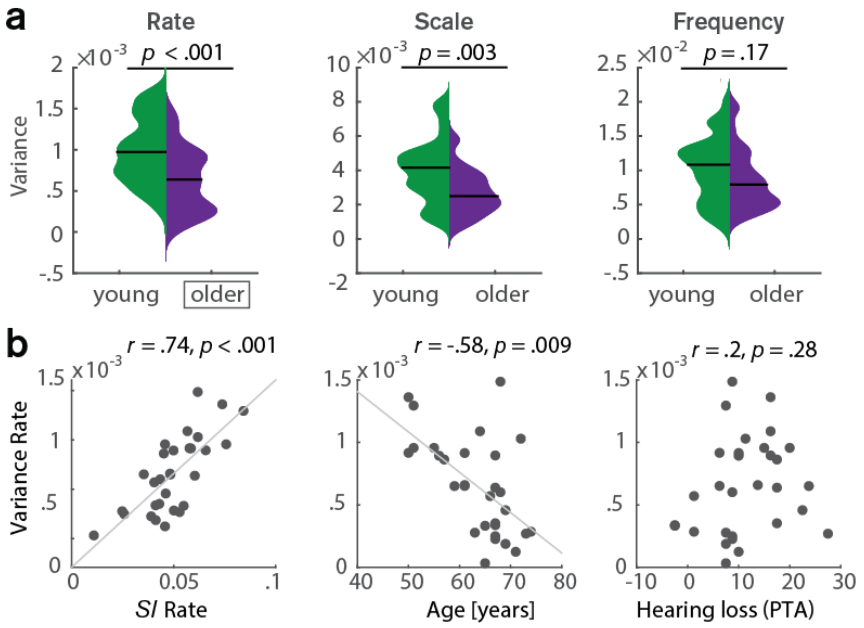


Figure S3. Decoding results. (a) Variance of the MTFs' marginal profiles compared between age groups. *P*-values are based on an exact permutation test. (b) Amongst the older group, variance in the rate profiles correlate with the selectivity index for rate and chronological age, but not hearing loss. PTA: pure tone average.

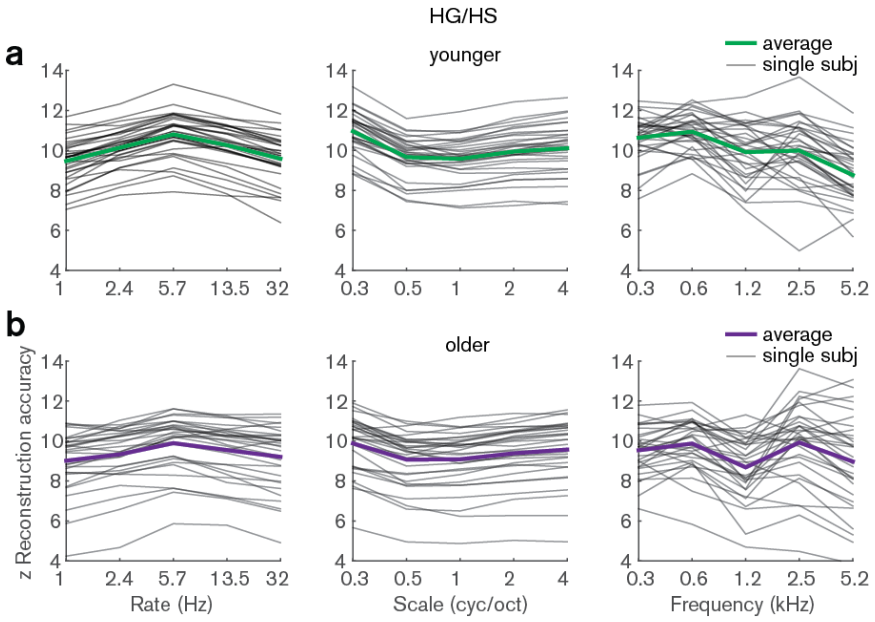


Figure S4. Z-scored reconstruction accuracies from Heschl's gyrus. Single subject and median reconstruction accuracy profiles for temporal rate, spectral scale and frequency z-scored with respect to the empirical null distribution for young (a) and older (b) participants. Note that all z-values are > 1.96 and are thus significant at $p < 0.05$.

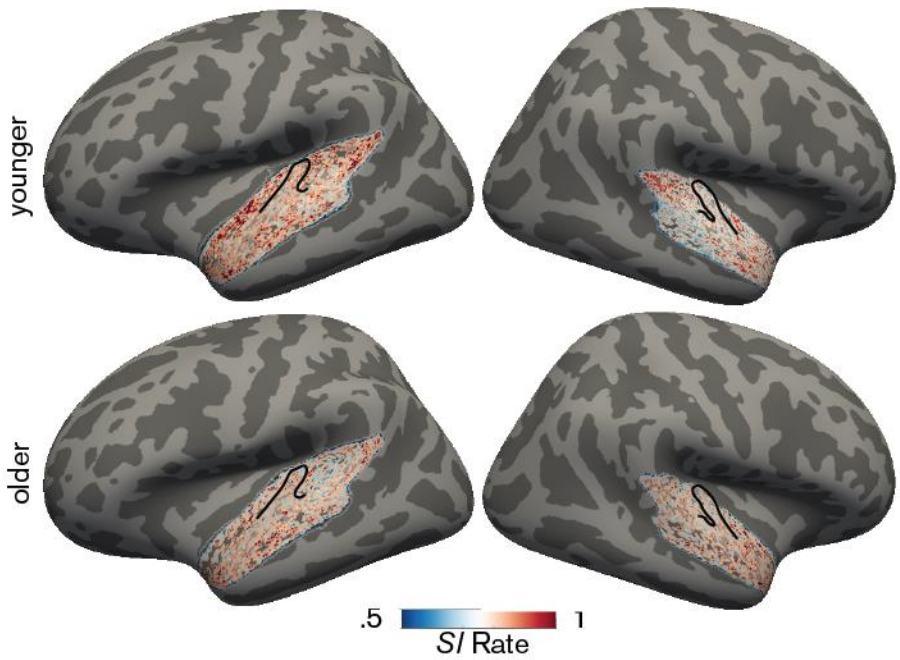


Figure S5. Topographic maps for rate selectivity index. Maps were derived by calculating the SI across the rate profile for a given voxel and averaged in the young group (upper panel) and older group (bottom panel). Black outlines indicate Heschl's gyrus.

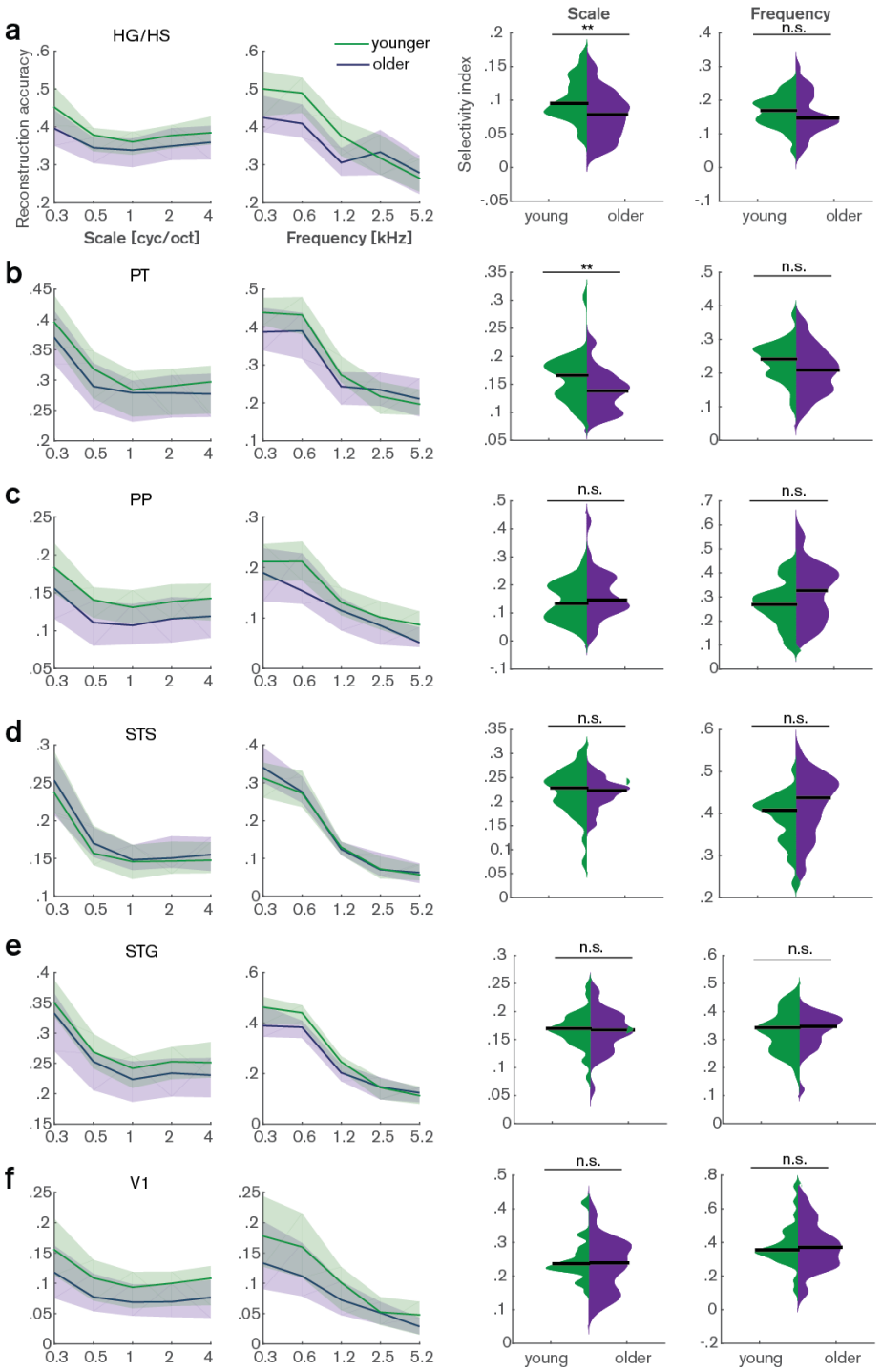


Figure S6. Decoding of spectral scale and frequency in ROIs. Median reconstruction accuracy profiles (left) and selectivity index (right) per age group for spectral scale and frequency in (a) Heschl's gyrus and sulcus, (b) planum temporale, (c) planum polare, (d) superior temporal sulcus, (e) superior temporal gyrus and (f) calcarine sulcus. Selectivity index was compared per ROI between age groups using exact permutation test. HG/HS: Heschl's gyrus, PT: planum temporale, PP: planum polare, STS: superior temporal sulcus, STG: superior temporal gyrus, V1: primary visual cortex. ** $p < 0.01$.