Temporal selectivity declines in the aging human auditory cortex

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

[1] Department 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 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 other 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

Age-dependent broadening of tuning to temporal modulation rate in older humans' cortical sound representations highlights decreased temporal selectivity as an anatomically and functionally specific 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 to 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, 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). This has been interpreted as compensatory mechanism 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 recent more unspecific 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 MEG 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 (Lalwani et al., 2019). Accumulating evidence supports the theory of an age-related loss of inhibition in sensory cortices (Caspary et al., 2008).

A recent model of cortical processing (Chi et al., 2005) 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., 2019). However, those studies 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 applies an fMRI encoding and decoding approach 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 thus appears to

be accompanied by an anatomically and functionally specific broadening of temporal-rate tuning in auditory cortex.

Results

fMRI responses to sounds from a total of N=64 younger and older listeners were modelled using both model-based encoding (Santoro et al., 2014) and decoding (Santoro et al., 2017; Figure 1). In brief, sounds were decomposed into acoustic features, that is, the frequency-dependent spectro-temporal modulation content (Chi et al., 2005). Analyses were restricted to voxels in an anatomic mask of the auditory cortex. In a fourfold cross-validation procedure, we used ridge regression to derive single- and multi-voxel MTFs.

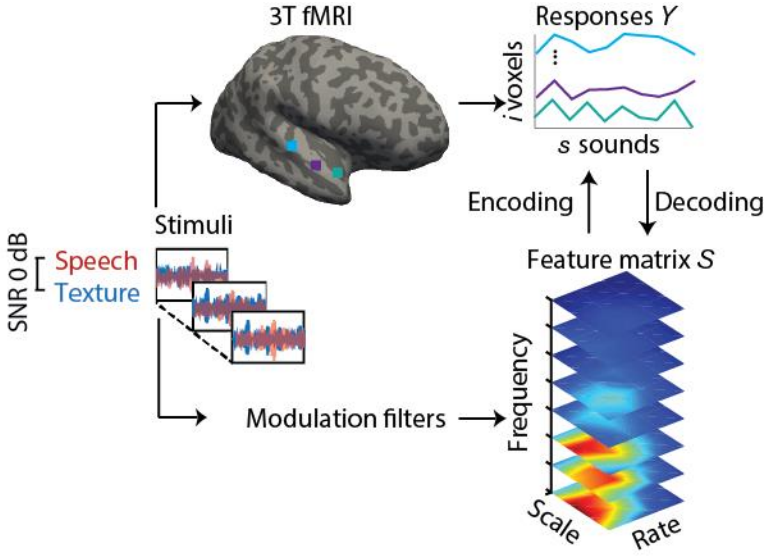


Figure 1. Experimental design and analysis pipeline. Participants listened to a story embedded in sound textures at an SNR of 0 dB while we acquired 3T-fMRI data. For audiograms of the participants see Figure 1 – figure supplement 1). The sounds were decomposed into their modulation spectrum (Chi et al., 2005; for modulation spectra of the sounds see Figure 1 – figure supplement 2). 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.

Encoding results

We compared three models of hemodynamic responses to sound. The models respectively describe the fMRI responses to the (1) speech stream (foreground), (2) textures (background), or (3) the mixture of both streams (for modulation spectra of the different streams see Figure 1 – figure supplement 2).

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

Sound identification accuracy was higher than chance level (0.5) for all models and age groups (Figure 2). The sound mixture model had the highest sound identification accuracies (Figure 2a) and thus described most accurately the fMRI responses to sounds. All subsequent encoding and decoding analyses were based on this mixture model. Young participants had significantly higher identification accuracies for mixture than old participants (Figure 2a). There was no significant correlation between chronological age and sound identification accuracy for mixture within the young or older group (Figure 2b).

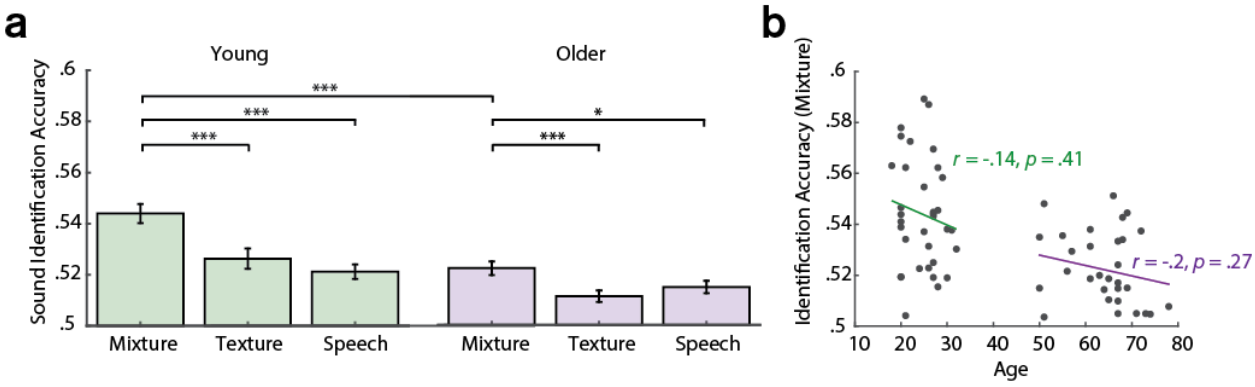
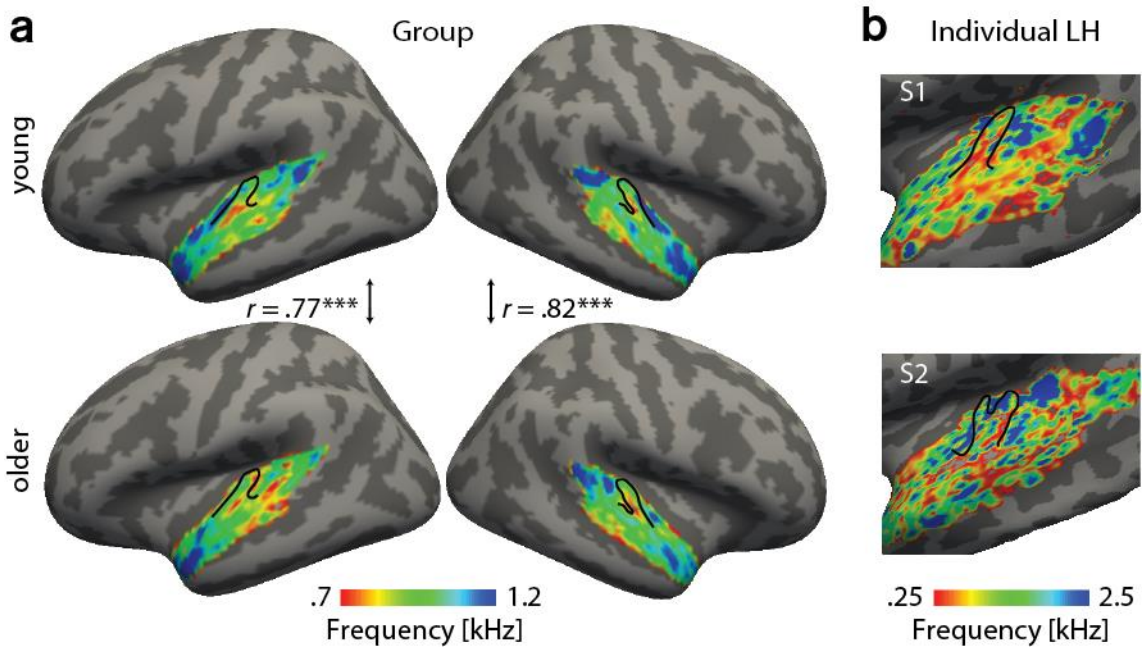


Figure 2. Sound identification accuracies for the different encoding models. (a) Accuracies for sound mixture, texture (background stream) and speech (foreground stream), for young (left, green) and older participants (right, violet). Bars indicate the mean (± standard error of the mean, SEM) of all 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) Identification accuracies for sound mixture were uncorrelated with chronological age within each age group.

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 (Figure 3). Tonotopic gradients were most obvious in the young group (Figure 3a, upper panel). Consistent with previous human fMRI data (Moerel et al., 2012), a low frequency region was observed in the central region of Heschl's gyrus (HG), presumably marking the boundary between the human homologue of primary auditory fields A1 and R (Figure 3a, red-yellow). This low frequency region was surrounded antero-medially and posteriorly by high frequency regions (Figure 3a, green-blue). The antero-medial high frequency areas clustered on the planum polare (PP). The posterior regions preferring high frequencies covered planum temporale (PT). We did not find significant age-group differences using two-sample voxel-wise *t* -tests (corrected for multiple comparisons). In fact,

129 tonotopic group maps were correlated, indicating that topography for best frequency was preserved

130 in older adults (Figure 3a).



131

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

138 Topographical best-feature maps for rate and scale were more complex. Similarly, we did not find

139 significant voxel-wise age-group differences using two-sample *t*-tests (corrected for multiple

140 comparisons). On the contrary, both for temporal and spectral modulations, group maps were

141 correlated between age groups (Figure 4). For best temporal modulation maps, we observed a medial-

142 to-lateral gradient for increasing temporal rate (Figure 4a). For best spectral modulation maps, the

143 locus of highest spectral resolution coincided with the low frequency region on Heschl's gyrus (Figure

144 3a, Figure 4b).

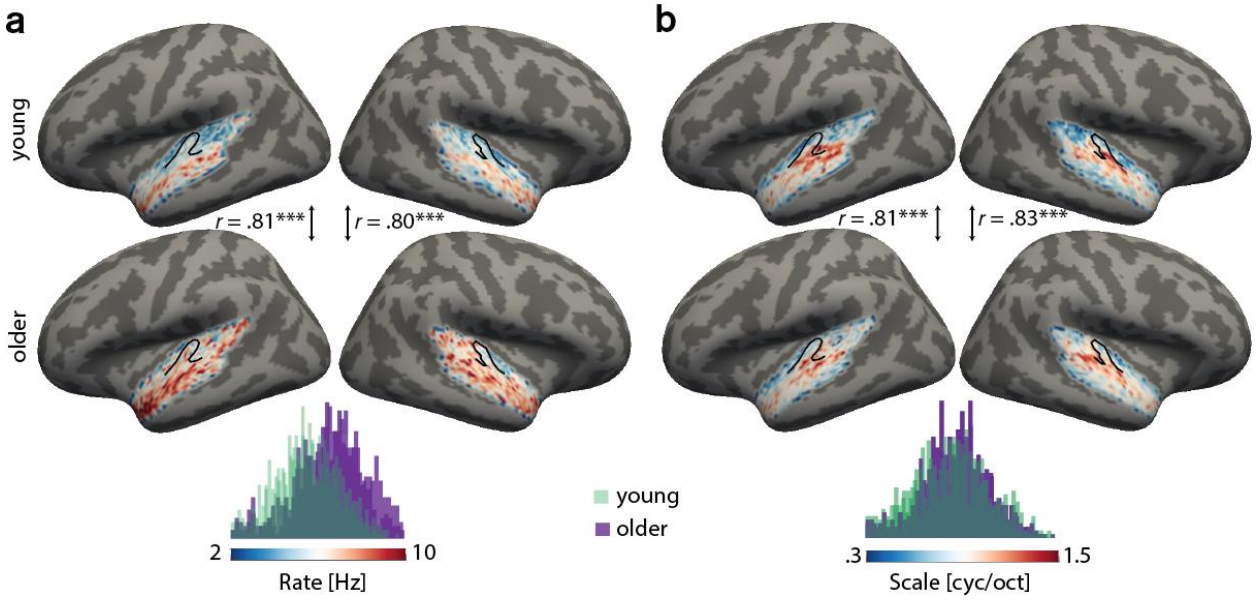


Figure 4. Group best-feature maps for temporal rate (a) and spectral scale (b). Group maps were obtained as median across subjects and shown on FreeSurfer's *fsaverage5* template. Black outlines indicate Heschl's gyrus. Topographic organization in young (upper panel) and older group (bottom panel) is 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. Note the shift of the distribution in older relative to young adults for temporal rate (a) but not spectral scale (b), indicating that older adults' voxel-wise preferred temporal rates are shifted towards higher modulation rates. *** $p < 0.005$.

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 (Figure 1). Reconstruction 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.

Our main hypothesis was that cortical sensitivity is highest for slow temporal modulations, based on previous observations of human (but not monkey) auditory cortex being most sensitive to the modulations present speech (Santoro et al., 2017, Erb et al., 2019a). Decoding yielded highest

accuracies at frequencies of 230–580 Hz (mean r = 0.47) and spectral scales of 0.25 cyc/oct (mean r = 0.47), irrespective of age (Figure 5 a,b), indicating that brain responses followed the frequency and spectral modulation content of the stimuli (Figure 1 – figure supplement 2 a). Conversely, for temporal modulations, reconstruction accuracies peaked at rates of 4–8 Hz (mean r = 0.4) in both age groups (Figure 5 a,b). Those peaks were not present in the stimuli (Figure 1 – figure supplement 2 a) and imply preferred processing of temporal rates in the speech-relevant range (Poeppel and Assaneo, 2020), corroborating our hypothesis (see also discussion).

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* age difference = 0.01, p = 0.004 [permutation test], Cohen's d = 0.68, Figure 5d, left). Note that removing the extreme case in the young group (outliers were defined as exceeding the grand average by ± 2 SD) for rate *SI* does not alter results qualitatively: It reduces the mean age-group difference to a still significant *SI* difference = 0.009, p = 0.026. Removing the two extreme cases in the older group for scale *SI* increases the mean age-group difference to a significant *SI* difference = 0.037, p = 0.024. A lower selectivity index is consistent with a broadening of temporal tuning in the aged auditory cortex.

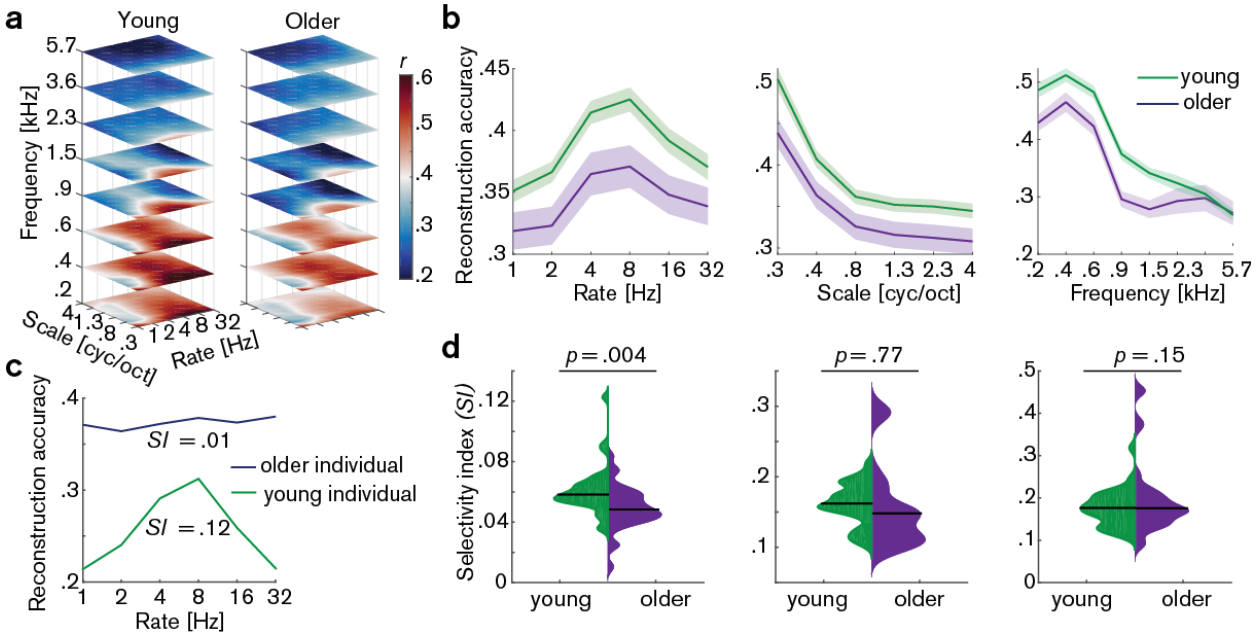


Figure 5. Decoding from auditory cortex. (a) Auditory cortical modulation transfer functions (MTFs) for the young and older group quantified by reconstruction accuracies (Pearson's *r* between predicted an actual acoustic features). (b) Mean ± SEM of the MTFs' marginal profiles for rate, scale and frequency for young (green) and older (violet) participants separately. (c) We quantified tuning selectivity by the selectivity index (see Eq. [7]). For illustration purposes, we show the marginal profiles for rate in a young (green) and older individual (violet) with extreme *SI* values. (d) *SI* was compared per acoustic dimension between age groups using an exact permutation test. Black line indicates the median *SI*. Black line indicates the median *SI*. Note that when removing the extreme case in the young group (> 2 SD away from the mean) for rate *SI*, the mean age-group difference remains significant (p = 0.026). Removing the two extreme cases in the older group for scale *SI* leads to a significant age-group difference (p = 0.024). Note the different scaling of the plots. See Figure 5 – figure supplement 1 for decoding from auditory cortex with ICA (AROMA)-cleaned data.

Careful artefact removal from fMRI data is particularly critical for age comparisons. Therefore, we re-ran the decoding analysis on ICA-cleaned data (using AROMA, Pruim et al., 2015). Irrespective of preprocessing without (Figure 5) or with ICA-cleaning (Figure 5–figure supplement 1), the key results remained unchanged, that is reconstruction accuracies peaking at temporal rates of 4 – 8 Hz as well as selectivity for temporal rates being higher in young than in older participants.

Also, this key result of a broadened tuning to temporal rate was corroborated by an even simpler measure of selectivity: The variance in the reconstruction accuracy profiles across the six temporal rate bins (1–32 Hz, 'rate variance') was higher in the young than in the older participants (mean difference in variance = 3.7×10^{-4}, $p < 0.001$ [permutation test], Cohen's $d = 0.93$, Figure 6, a, left panel).

In the older group, the rate variance was significantly correlated with the rate selectivity index and chronological age but was uncorrelated with hearing loss (Figure 6, b). Note however that the rate selectivity index (see above) in the older group was not significantly correlated with chronological age (Spearman's rho = -0.2, $p = 0.289$), nor with hearing loss (Spearman's rho = 0.34, $p = 0.062$).

In a control analysis, we ensured that these results cannot be explained by the overall slightly worse model fits in the older group as quantified by sound identification accuracy (as evident in Figure 2): In a multiple regression predicting older participants' variance in rate from chronological age with covariates PTA and sound identification accuracy, neither PTA ($t(26) = 0.73$, $p = 0.473$, permutation test) nor sound identification accuracy ($t(26) = 1.02$, $p = 0.312$) were significant, while age was significant ($t(26) = -2.8$, $p = 0.003$; *adjusted* $R^2 = 0.39$). Age remained significant when excluding two multivariate outliers identified based on Cook's distance ($t(24) = -2.25$, $p = 0.021$).

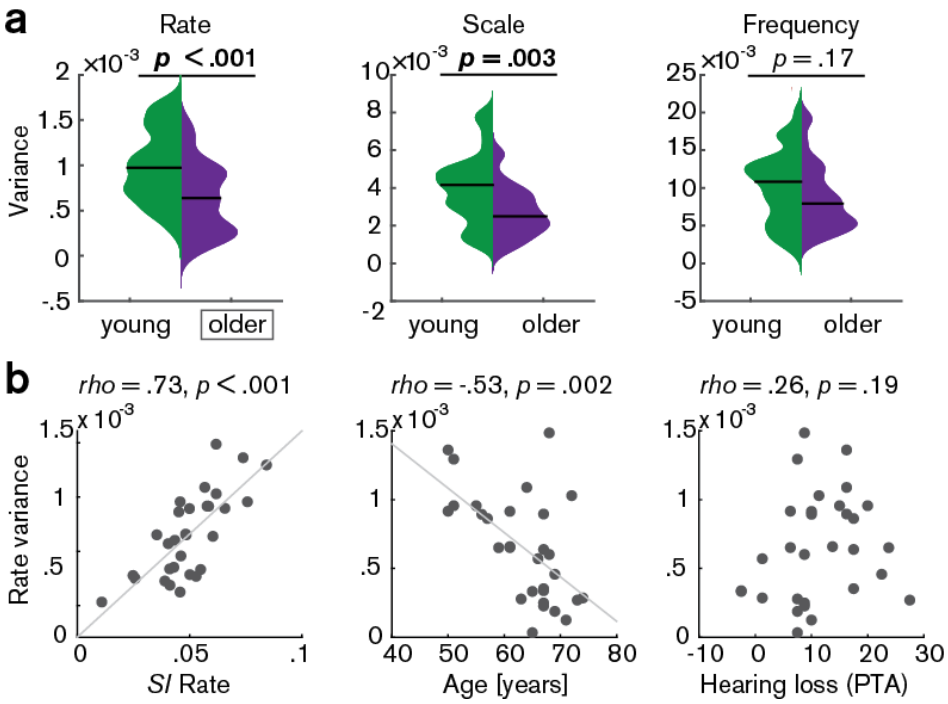


Figure 6. (a) Variance of the MTFs' marginal profiles compared between age groups. *P*-values are based on an exact permutation test. (b) Amongst the older group, the variance of the rate profiles ('rate variance') correlates with the *SI* for rate (left panel) and chronological age (middle panel), but not with hearing loss (right panel; Spearman's correlation coefficient). *SI*: selectivity index, PTA: pure tone average.

Feature reconstruction from auditory subregions. To test the differential sensitivity of auditory cortical subregions to acoustic features, we ran the decoding analyses in the following regions of interest (ROIs) derived from the FreeSurfer atlas (Destrieux et al., 2010): Heschl's gyrus and sulcus (HG/HS), planum polare (PP), planum temporale (PT), superior temporal gyrus (STG) and superior temporal sulcus (STS, Figure 7 a – f)). As control regions, we chose primary visual cortex (V1, calcarine sulcus) as another primary sensory area, and two higher-level areas, namely the middle frontal (MFG) and superior parietal gyrus (SPG; Figure 7 a, g – i).

A decrease in the rate selectivity index with age was confirmed for HG/HS (mean *SI* difference = 0.012, $p = 0.009$ [permutation test]) and PT (mean *SI* difference = 0.016, $p = 0.005$ [permutation test]; see Figure 7 b, c). Removing the extreme cases in the young group exceeding the grand average by ±

225 2 SD reduces the mean *SI* age difference to 0.012 (p = 0.008) in HG/HS, and to 0.012 (p = 0.04) in PT.
226 The alternative measure of selectivity, variance of the rate profiles, was also significantly higher in
227 young than older adults in HG/HS (variance difference = 4 x 10^{-4}, p = 0.0004) and PT (variance
228 difference = 5 x 10^{-4}, p = 0.0003).

229 Note that reconstruction accuracies in all ROIs but the control regions (V1, SPG, MFG) were
230 significant (permutation tests; for the z-scored reconstruction accuracies in an exemplary region
231 HG/HS see Figure 7–figure supplement 1). In contrast, there were no such age differences in temporal
232 selectivity in PP, STG, STS or the control regions V1, SPG and MFG (Figure 7 d – i).

233 Qualitatively, this observation was confirmed by a control analysis, where we derived topographic
234 maps of the rate selectivity index from the encoding models. To this end, we marginalized MTFs for
235 rate and calculated the *SI* for a given voxel. The topographic distribution confirmed highest voxel-wise
236 *SI* for rate in area PT of the younger group (Figure 7–figure supplement 2).

237 These results indicate that age-related changes in temporal selectivity selectively occur in primary
238 auditory cortex and the planum temporale, but not surrounding belt areas. For spectral scale, we also
239 found a decreased *SI* in older adults in HG/HS (*SI* difference = 0.012, p = 0.008) and PT (*SI* difference =
240 0.016, p = 0.021), but none of the other ROIs (Figure 7–figure supplement 3). For frequency, we did not
241 find any age differences in selectivity index at all (Figure 7–figure supplement 3).

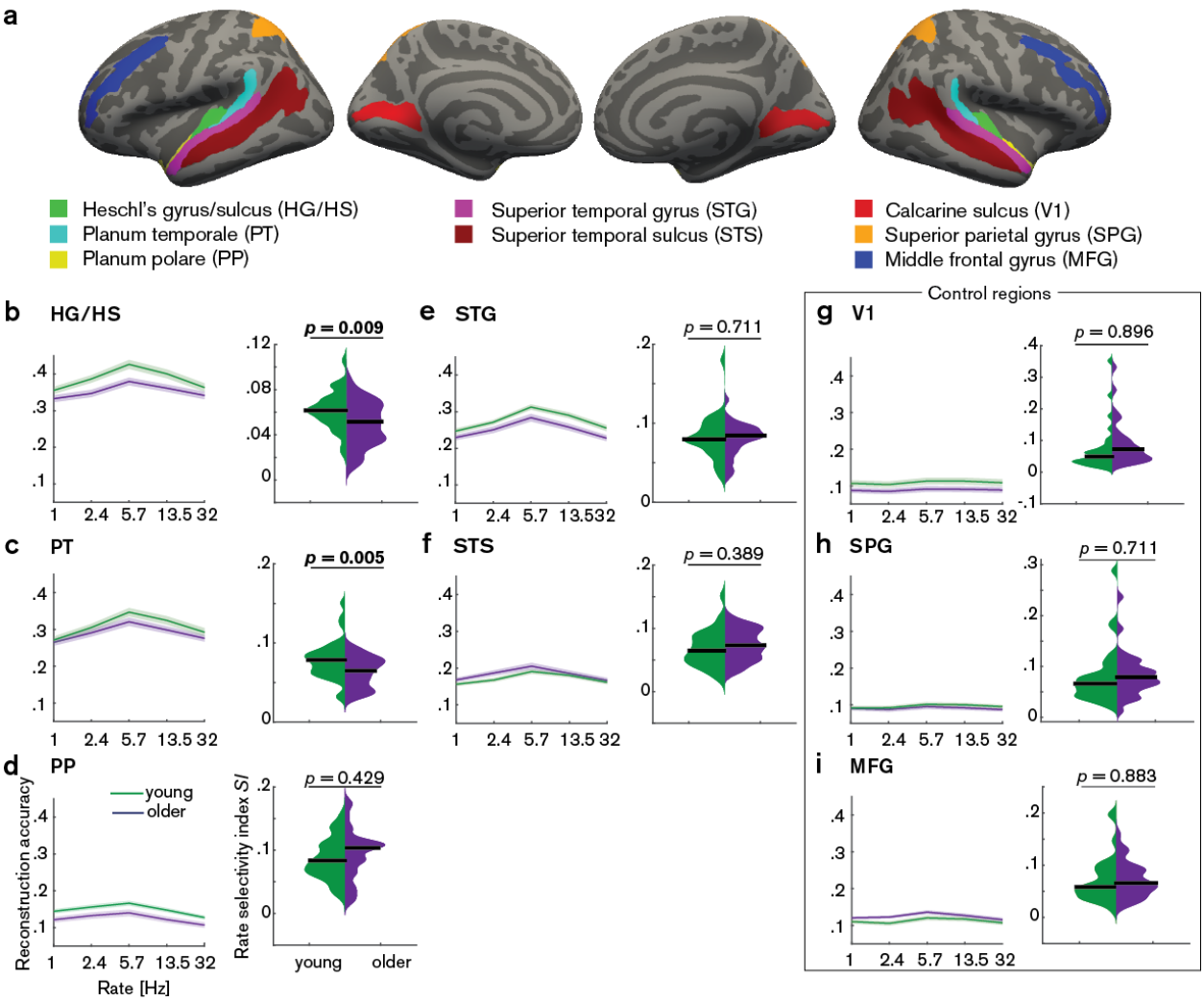


Figure 7. Decoding of temporal rate from regions of interest (ROIs). (a) ROIs derived from the FreeSurfer atlas (Destrieux et al., 2010) are displayed on FreeSurfer's *fsaverage* template. Mean (± SEM) reconstruction accuracy profiles (left) and selectivity index (right) per age group for temporal rate in (b) Heschl's gyrus and sulcus, (c) planum temporale, (d) planum polare, (e) superior temporal gyrus, (f) superior temporal sulcus and the control regions (g) calcarine sulcus, (h) superior parietal gyrus, (i) middle frontal gyrus. The selectivity index (*SI*) was compared per ROI between age groups using an exact permutation test. When removing young-subject *SI* exceeding the grand average by ± 2 SD, the *SI* age difference remains significant in HG/HS ($p = 0.008$), and in PT ($p = 0.04$). Consistently, our alternative measure of selectivity, variance across rate bins, was significantly higher in young than older adults in HG/HS ($p = 0.0004$) and PT ($p = 0.0003$). See Figure 7 – figure supplement 1 for the reconstruction accuracies z-scored with respect to the empirical null distribution in an exemplary region HG/HS. See Figure 7–figure supplement 2 for group maps of the rate selectivity index obtained from encoding models. See Figure 7–figure supplement 3 for decoding of spectral scale and frequency from ROIs.

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, broadening of rate tuning correlates with chronological 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 (Figure 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

278 and van Dijk, 2012, Moerel et al., 2012, 2014) and non-human primates (Merzenich and Brugge, 1973,
279 Morel et al., 1993, Kosaki et al., 1997, Rauschecker et al., 1997, Bendor and Wang, 2008, Baumann et al.,
280 2013, Joly et al., 2014, Baumann et al., 2015). Notably, tonotopic maps in younger versus older adults
281 were highly correlated, indicating topographical stability of tonotopic organization across age groups.

282 The well-demonstrated existence of tonotopic maps notwithstanding, a topographic organization
283 for temporal or spectral modulations in auditory cortex is less clear (Schonwiesner and Zatorre, 2009).
284 Here, in both young and older adults we observed a medial-to-lateral low-to-high gradient for
285 temporal modulations across the superior temporal plane (Figure 4a), while for spectral modulations,
286 the locus with the highest resolution (~1.5 cyc/oct) was observed in lateral Heschl's gyrus (Figure 4b).
287 A topographic representation for modulation rate has previously been observed in the primate
288 auditory cortex (Baumann et al., 2015) where fast temporal acoustic information was preferably
289 encoded in caudal auditory regions (Camalier et al., 2012, Kusmierek and Rauschecker, 2014) and slow
290 temporal information in rostral areas R and RT .

291 Whereas earlier human fMRI studies had failed to show a clear topography for modulation rate
292 (Giraud et al., 2000, Schonwiesner and Zatorre, 2009, Overath et al., 2012, Leaver and Rauschecker,
293 2016), recent human fMRI (Santoro et al., 2014) and electrocorticography (ECoG) studies (Hullett et al.,
294 2016) proposed the presence of a posterior-to-anterior high-to-low rate gradient in human auditory
295 cortex. However, comparison of those studies is hampered by the differences in coverage and ranges
296 of temporal modulations: While Hullett et al. (2016) examined representation of slow temporal
297 modulation rates present in speech (1–2 Hz) only along the STG with ECoG, Santoro at al. (2014)
298 analyzed rates of 1–27 Hz, whereas Baumann et al. (2015) presented rates of up to 512 Hz with more
299 comprehensive coverage using fMRI.

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

302 While previous studies had presented single natural sounds (Erb et al., 2019a), or even synthetic
303 sounds such as AM stimuli (Baumann et al., 2015) or dynamic ripples (Schonwiesner and Zatorre,
304 2009), here, we presented a continuous stream of speech embedded in an acoustically rich
305 background of sound textures. As an important foundation for the main conclusions of this
306 manuscript, we here demonstrate the capacity to derive meaningful tonotopic maps from natural
307 stimulus conditions, at conventional field strengths (3 T) and in special populations (older
308 participants).

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

314 Methodologically, our encoding approach relies on the assumption that acoustic features that
315 maximally activate single voxels are more accurately encoded. However, higher responses may not
316 necessarily mean better encoding. The second, complementary decoding analysis incorporates the
317 whole range of values rather than only the maximal values.. The reconstruction accuracy explicitly
318 quantifies the amount of information about a set of stimulus features that is available in a region of
319 interest. Accurate reconstruction of acoustic features in test sounds indicates that these features are
320 reproducibly mapped into distinct spatial patterns. As data from individual voxels are jointly
321 modelled, this combined analysis of signals from multiple voxels increases the sensitivity for stimulus
322 information that may be represented in patterns of responses, rather than in individual voxels.

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

324 Under the assumption that the reconstruction accuracy reflects tuning properties of neuronal
325 populations, the observed decrease in selectivity reveals a broadening of temporal tuning functions
326 with age. This finding provides mechanistic insights into the changes of temporal processing in the
327 aging auditory cortex: The broadened temporal tuning, reminiscent of an age-related decrease in
328 temporal precision in the midbrain (Anderson et al., 2012), may underlie the difficulties older adults
329 experience in speech comprehension in noisy environments (Anderson et al., 2011). While the present
330 study lacks more direct measures of encoding along the auditory pathway, it is important to note that
331 rate selectivity in the older listeners was statistically unrelated to their pure-tone audiograms (this
332 holds true for both presented measures of tuning, the selectivity index and the variance across
333 temporal rates [Figure 6]). Thus, the broadening of temporal tuning is not explicable merely by
334 peripheral hearing loss. Future studies need to carefully assess the consequences of altered cortical
335 temporal tuning on behavioral and psychoacoustic performance (Erb et al., 2019b, Flinker et al., 2019,
336 Holmes and Griffiths, 2019), a measure that was precluded by the use of natural sounds in the current
337 study.

338 The observation of decreased temporal selectivity may seem at odds with previous accounts of
339 elevated cortical responses to the temporal speech envelope in aging. Evidence for better
340 reconstruction of temporal envelopes from older listeners' MEG signals (Presacco et al., 2016) and
341 larger N1 amplitudes in older adults with widened auditory filters (Herrmann et al., 2016, Henry et al.,
342 2017) suggest an enhanced cortical envelope representation. However, an inflated cortical
343 representation of the speech envelope has been argued to be rather a maladaptive response as
344 compensation of the lack of temporal precision in the midbrain (Anderson et al., 2012).

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

347 modulation frequency, temporal coding is achieved through phase-locking to the stimulus envelope.
348 Fast temporal modulations (> 50 Hz) are typically encoded through the rate code, while neurons
349 tuned to slow temporal rates (< 50 Hz) mostly synchronize to the sounds' modulations (temporal
350 code, Sachs, 1984, Joris et al., 2004). It remains unclear how hemodynamic responses reflect these two
351 types of neuronal coding and how those are in turn related to the observed age-related broadening of
352 tuning functions. Evidence from electrophysiology suggests a change in coding strategy across
353 neuronal populations in A1 from less temporal coding towards more rate coding of temporal
354 modulations in aged animals (Ng and Recanzone, 2018).

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

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

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

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

379 **Conclusions**

380 The present study aimed to narrow the gap between recent progress in modelling the neural
381 processing of complex features from natural stimuli on the one hand, and potential senescent change
382 in these stimulus representations in the aging auditory cortex on the other hand. Although the large-
383 scale topographic organization of acoustic features appears preserved in the auditory cortex of older
384 compared to younger listeners, age-related differences in the marginal profiles of multi-voxel MTFs
385 were evident. Tuning to slow temporal rates which abounds in natural sounds and especially in
386 speech (Erb et al., 2019) was markedly sharper in young compared to older participants. Consistent
387 with previous findings in the macaque, these results suggest that temporal rate selectivity in auditory
388 cortex declines in normal aging. The specificity of this decline, confined to primary auditory and
389 adjacent areas and sparing cochleotopic representations, opens a new lead in the ongoing hunt for
390 tractable neurobiological signatures of older adults' widely observed deficits in speech
391 comprehension in noisy environments.

392 ## Materials and Methods

393 ### Participants

394 We invited a total of *n = 75* participants for scanning from which we had to exclude one participant
395 due to excessive movement, three participants due to incidental neurological findings, five
396 participants due to their inability to understand speech in noise, one participant due to broken
397 headphones, one participant due to age (43 years, i.e., could not be assigned to young or older group).
398 The remaining *n* = 64 participants were right-handed, young (*n* = 33; aged 18–32, mean 24.7 years18
399 female) and older (*n* = 31; aged 51–78, mean 63.8 years, 15 female) native German speakers.
400 Simulations (custom matlab code) showed that a two-sample permutation test at a conventional type
401 I error rate of 5 % with a sample size of *N* = 27 per group can detect medium to large effects of Cohen's
402 *d* = 0.75 (as can be expected for the neural fMRI measures under consideration, e.g. Alavash et al.,
403 2019)with a satisfactory power of 80 %. The power of our procedure reduces accordingly if the true
404 effect in the population is smaller but remains over 60 % also for true effects closer to Cohen's *d* = 0.5.

405 Younger participants had self-reported normal hearing. Older participants' hearing ranged from
406 normal hearing to mild hearing loss. The older participants were part of a cohort which is regularly
407 tested in the lab (e.g., Alavash et al., 2019). They were recruited based on their audiograms that had
408 been acquired within the two years prior to the experiment (for audiograms see Figure 1 – figure
409 supplement 1 a) and were excluded from the study if the pure-tone-average (PTA) of the better ear
410 exceeded 30 dB HL. On average, hearing thresholds are expected to increase by approximately 0.25–
411 1.7 dB HL per year at the age range of 48–79 years (Wiley et al., 2008). Thus, we considered 2-year-old
412 audiograms to be valid. All participants gave informed consent and were financially compensated or
413 received course credit. All procedures were approved by the local ethics committee of the University
414 of Lübeck (ethical approval AZ 16-107).

415 **Stimuli and task**

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

427 Participants were asked to listen to the story and answer three four-choice questions on its
428 semantic content after each run (24 questions in total); chance level was thus proportion correct =
429 0.25. Four-answer choice questions were displayed on a screen and participants responded via a
430 button box in their right hand. Both the young and older group performed on average above chance
431 (see Figure 1 – figure supplement 1b), but young participants performed significantly better than
432 older participants (mean proportion correct difference = 0.1, p = 0.009 (permutation test), Cohen's d =
433 0.65).

434 **Acquisition of MRI data**

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

437 Functional T2*-weighted data were acquired using an echo planar imaging sequence. We collected
438 eight runs of eight minutes each (8 runs à 519 volumes) using continuous scanning. The acquisition
439 parameters were as follows: repetition time (TR) = acquisition time (TA) = 947 ms, echo time (TE) = 28
440 ms, acceleration factor = 4, flip angle = 60°, field of view (FOV) = 200 x 200 mm, 52 slices; voxel size =
441 2.5 mm isotropic (whole-brain coverage). Field maps for intensity inhomogeneity correction were
442 acquired after every second run (TR = 610 ms, TE1 = 4.92, TE2 = 7.38, flip angle = 60°, FOV = 200 x 200
443 mm, 62 slices, voxel size = 2.5 mm isotropic).

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

449 **Preprocessing of MR data**

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

453 **Anatomical data preprocessing**

454 The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU)
455 using N4BiasFieldCorrection (ANTs 2.2.0) and used as T1w-reference throughout the workflow. The
456 T1w-reference was then skull-stripped using antsBrainExtraction.sh (ANTs 2.2.0), using OASIS as target
457 template. Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1), and the brain mask
458 estimated previously was refined with a custom variation of the method to reconcile ANTs-derived
459 and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle. Spatial

460 normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c was performed
461 through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of
462 both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter
463 (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9).

464 **Functional data preprocessing**

465 For each of the eight BOLD runs, the following preprocessing was performed. First, a reference
466 volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. A
467 deformation field to correct for susceptibility distortions was estimated based on a field map that was
468 co-registered to the BOLD reference, using a custom workflow of *fMRIPrep*. Based on the estimated
469 susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-
470 registration with the anatomical reference. The BOLD reference was then co-registered to the T1w
471 reference using bbregister (FreeSurfer) which implements boundary-based registration. Co-
472 registration was configured with nine degrees of freedom to account for distortions remaining in the
473 BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation
474 matrices, and six corresponding rotation and translation parameters) are estimated before any
475 spatiotemporal filtering using mcflirt (FSL 5.0.9). BOLD runs were slice-time corrected
476 using 3dTshift from AFNI 20160207. The BOLD time-series (including slice-timing correction when
477 applied) were resampled onto their original, native space by applying a single, composite transform to
478 correct for head-motion and susceptibility distortions. In a control analysis, ICA-based Automatic
479 Removal Of Motion Artifacts (AROMA) was used to generate a variant of data that is non-aggressively
480 denoised (Pruim et al., 2015). All analyses (see below) were performed in native space. Subsequently,
481 results (best feature maps) were resampled to surfaces on the *fsaverage5* template using mri_vol2surf
482 (FreeSurfer).

483 **Modulation representation**

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

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

496 Next, the auditory spectrogram was further analyzed by the cortical stage, where the modulation
497 content of the auditory spectrogram was computed through a bank of 2-dimensional filters selective
498 for a combination of spectral and temporal modulations. The filter bank performs a complex wavelet
499 decomposition of the auditory spectrogram. The magnitude of such decomposition yields a phase-
500 invariant measure of modulation content. The modulation selective filters have joint selectivity for
501 spectral and temporal modulations, and are directional, i.e. they respond either to upward or
502 downward frequency sweeps.

503 Filters were tuned to spectral modulation frequencies of Ω = [0.3, 0.4, 0.8, 1.3, 2.3, 4] cyc/oct,
504 temporal modulation frequencies of ω = [1, 2, 4, 8, 16, 32] Hz, and centre frequencies of f = [232, 367,
505 580, 918, 1452, 2297, 3633, 5746] Hz. Our rationale for this choice of values was to use a
506 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 Figure 1 – figure supplement 2). 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 temporal resolution of the feature matrix and the fMRI data.

Note that for the encoding analysis (see below), 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.

528 **Encoding and decoding models**

529 We applied two modelling approaches to fMRI data as described in Erb et al. (2019; Figure 1). In a
530 first *univariate encoding* approach, we calculated a modulation transfer function (MTF) for each
531 individual voxel. The MTF characterizes how faithfully the modulation content of the stimulus gets
532 transferred to the voxel. By assigning the feature value with the maximal response to each voxel, we
533 obtained maps of a voxel's best features across the auditory cortex.

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

539 **Univariate encoding analysis: Model estimation**

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

542 $$Y_i = S_{train} C_i + n \qquad (1)$$

543 where N_{train} is the number of sounds in the training set, and C_i is an $[F \times 1]$ vector of model
544 parameters, whose elements c_{ij} quantify the contribution of feature j to the overall response of voxel i.
545 Columns of matrices S_{train} and Y_i were converted to standardized z-scores. Therefore, Equation (1) does
546 not include a constant term. The solution to Equation (1) was computed using kernel ridge regression
547 (Hoerl and Kennard, 1970). The regularization parameter λ was selected independently for each voxel
548 via generalized cross-validation (Golub et al., 1979). The search grid included 25 values between 10^{-6}
549 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 within each participant 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 sounds on which the estimation was performed and 1038 test sounds. In this way, we obtained four estimates of each voxel's feature profile which were averaged across iterations.

Model evaluation

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

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

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

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

Note that $m = 1$ indicates a correct match; $m = 0$ indicates that the predicted activity pattern for stimulus s_i was least similar to the measured one among all stimuli. Normalized ranks (sound identification scores) were computed for all stimuli in the test set, and the overall model's accuracy was obtained as the mean of the sound identification scores across stimuli.

571 **Topographical best-feature maps**

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

577 $$tMTF(\omega) = \sum_{\Omega} \sum_{f} C(\omega, \Omega, \mathrm{f}) \qquad (4)$$

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

584 **Multivariate decoding analysis: Model estimation**

585 In the multivariate decoding analysis, we evaluated the fidelity with which regions of interest
586 (ROIs) in auditory cortex encode acoustic features by estimating decoders. In addition to a region
587 including the whole auditory cortex, we selected five ROIs based on anatomical criteria (using the
588 FreeSurfer labels): Heschl's gyrus/sulcus (HG/HS), planum temporale (PT), planum polare (PP), superior
589 temporal gyrus (STG) and sulcus (STS). In a control ROI analysis, we selected visual cortex (V1, calcarine
590 sulcus), middle frontal (MFG) and superior parietal gyrus (SPG). For each individual in each ROI, a linear
591 decoder was trained for every feature of the modulation space based on the training data only

592 (Santoro et al., 2017). Each stimulus feature S_j [$N_{train} \times 1$] was modelled as a linear transformation of the
593 multi-voxel response pattern Y_{train} [$N_{train} \times V$] plus a bias term b_j and a noise term n [$N_{train} \times 1$] as follows:

594 $$S_j = Y_{train} C_j + b_j 1 + n \qquad (5)$$

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

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

602 Decoders were estimated on the train runs (see above) and tested on the test runs. Given the
603 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
604 feature matrix activity $\hat{S}_{\text{test}}$ [$N_{test} \times F$] for the test sounds was calculated as:

605 $$\hat{S}_{\text{test}} = Y_{test}\, \tilde{C} \qquad (6)$$

606 Predictions for all features from the test sets were concatenated and decoders were assessed
607 individually by computing the Pearson's correlation coefficient (r) between a predicted and a given
608 actual stimulus feature. For decoding from the whole auditory cortex (ROIs), this resulted in 288 (125)
609 correlation coefficients, which represented the MTF. To obtain marginal profiles of the MTFs, we
610 averaged across the momentarily irrelevant dimensions, respectively.
611 Statistical significance of the reconstruction accuracy was assessed with permutation testing. For
612 each stimulus feature we computed the null-distribution of accuracies at the single-subject level by
613 randomly permuting (1, 000 times) the stimulus labels of the test sounds and computing the

correlation coefficient with the predicted features for each permutation. Note that to reduce computing time and resources, we limited the number of permutations to 1,000 whenever encoding and decoding models had to be estimated for each permutation at the single-subject level. For less computationally intense calculations (group comparisons), we used the default of 10,000 permutations.

To preserve the temporal correlations among sounds, the predictors were phase scrambled by using the fast Fourier transform (fft), randomly permuting the phase angles of the fft and calculating the inverse fft before convolution with the hemodynamic response function. 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 mean 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. Based on the marginal profiles of the MTFs, we quantified how selectively a brain region is tuned towards processing the acoustic feature with the highest reconstruction accuracy relative to all other features using a standard index

$$SI = \frac{\max\limits_{j\in[k]}(r_j) - \operatorname*{mean}\limits_{j\neq j_{max}}(r_j)}{\max\limits_{j\in[k]}(r_j) + \operatorname*{mean}\limits_{j\neq j_{max}}(r_j)} \qquad (7)$$

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 10,000 times.

Acknowledgements

This research was funded by an ERC consolidator Grant (ERC-CoG-2014-646696 "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 as well as the editors and reviewers for very constructive feedback 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 (https://osf.io/zbuah/).

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 116:660-669.

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 (2019a) Homology and Specificity of Natural Sound-Encoding in Human and Monkey Auditory Cortex. Cerebral cortex 29:3636-3650.

Erb J, Ludwig AA, Kunke D, Fuchs M, Obleser J (2019b) 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 1-10.

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.

Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. NeuroImage 50:1202-1211.

Joly O, Baumann S, Balezeau F, Thiele A, Griffiths TD (2014) Merging functional and structural properties of the monkey auditory cortex. Frontiers in neuroscience 8:198.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541-577.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proceedings of the National Academy of Sciences of the United States of America 97:11793-11799.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.

Koch U, Grothe B (1998) GABAergic and glycinergic inhibition sharpens tuning for frequency modulations in the inferior colliculus of the big brown bat. Journal of neurophysiology 80:71-82.

Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. The Journal of comparative neurology 386:304-316.

Kusmierek P, Rauschecker JP (2014) Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. Journal of neurophysiology 111:1671-1685.

Lalwani P, Gagnon H, Cassady K, Simmonite M, Peltier S, Seidler RD, Taylor SF, Weissman DH, Polk TA (2019) Neural distinctiveness declines with age in auditory cortex and is associated with auditory GABA levels. NeuroImage 201:116033.

Langers DR, van Dijk P (2012) Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. Cerebral cortex 22:2024-2038.

Leaver AM, Rauschecker JP (2016) Functional Topography of Human Auditory Cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 36:1416-1428.

Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. Journal of neurophysiology 87:2237-2261.

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

Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain research 50:275-296.

Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:14205-14216.

Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. Frontiers in neuroscience 8:225.

Moerel M, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:11888-11898.

761 Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of
762 auditory cortex in macaque monkeys. The Journal of comparative neurology 335:437-459.
763 Ng CW, Recanzone GH (2018) Age-Related Changes in Temporal Processing of Rapidly-Presented
764 Sound Sequences in the Macaque Auditory Cortex. Cerebral cortex 28:3775-3796.
765 Overath T, Zhang Y, Sanes DH, Poeppel D (2012) Sensitivity to temporal modulation rate and spectral
766 bandwidth in the human auditory system: fMRI evidence. Journal of neurophysiology
767 107:2042-2056.
768 Park DC, Polk TA, Park R, Minear M, Savage A, Smith MR (2004) Aging reduces neural specialization in
769 ventral visual cortex. Proceedings of the National Academy of Sciences of the United States of
770 America 101:13091-13095.
771 Peelle JE, Wingfield A (2016) The Neural Consequences of Age-Related Hearing Loss. Trends Neurosci
772 39:486-497.
773 Poeppel D, Assaneo MF (2020) Speech rhythms and their neural foundations. Nature reviews
774 Neuroscience.
775 Presacco A, Simon JZ, Anderson S (2016) Evidence of degraded representation of speech in noise, in
776 the aging midbrain and cortex. Journal of neurophysiology 116:2346-2355.
777 Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF (2015) ICA-AROMA: A robust
778 ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage 112:267-277.
779 Rauschecker JP, Tian B, Pons T, Mishkin M (1997) Serial and parallel processing in rhesus monkey
780 auditory cortex. The Journal of comparative neurology 382:89-103.
781 Recanzone G (2018) The effects of aging on auditory cortical function. Hearing research 366:99-105.
782 Sachs MB (1984) Neural coding of complex sounds: speech. Annual review of physiology 46:261-273.
783 Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (2014) Encoding of
784 natural sounds at multiple spectral and temporal resolutions in the human auditory cortex.
785 PLoS computational biology 10:e1003412.
786 Santoro R, Moerel M, De Martino F, Valente G, Ugurbil K, Yacoub E, Formisano E (2017) Reconstructing
787 the spectrotemporal modulations of real-life sounds from fMRI response patterns. Proceedings
788 of the National Academy of Sciences of the United States of America 114:4799-4804.
789 Schonwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in
790 the human auditory cortex measured with high-resolution fMRI. Proceedings of the National
791 Academy of Sciences of the United States of America 106:14611-14616.
792 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily
793 temporal cues. Science 270:303-304.
794 Snell KB, Mapes FM, Hickman ED, Frisina DR (2002) Word recognition in competing babble and the
795 effects of age, temporal processing, and absolute sensitivity. The Journal of the Acoustical
796 Society of America 112:720-727.
797 Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical
798 fields obtained with phase-encoding FMRI. PloS one 6:e17832.
799 Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in
800 human auditory cortex revealed by progressions of frequency sensitivity. Journal of
801 neurophysiology 91:1282-1296.
802 Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory
803 neurons obtained using natural sounds. The Journal of neuroscience : the official journal of the
804 Society for Neuroscience 20:2315-2331.
805 Vaden KI, Jr., Eckert MA, Dubno JR, Harris KC (2019) Cingulo-opercular adaptive control for younger
806 and older adults during a challenging gap detection task. Journal of neuroscience research.
807 Vaden KI, Jr., Kuchinsky SE, Ahlstrom JB, Dubno JR, Eckert MA (2015) Cortical activity predicts which
808 older adults recognize speech in noise and when. The Journal of neuroscience : the official
809 journal of the Society for Neuroscience 35:3929-3937.

Voytek B, Kramer MA, Case J, Lepage KQ, Tempesta ZR, Knight RT, Gazzaley A (2015) Age-Related Changes in 1/f Neural Electrophysiological Noise. The Journal of neuroscience : the official journal of the Society for Neuroscience 35:13257-13265.

Wiley TL, Chappell R, Carmichael L, Nondahl DM, Cruickshanks KJ (2008) Changes in hearing thresholds over 10 years in older adults. Journal of the American Academy of Audiology 19:281-292; quiz 371.

Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X (2010) Functional properties of 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]Department 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

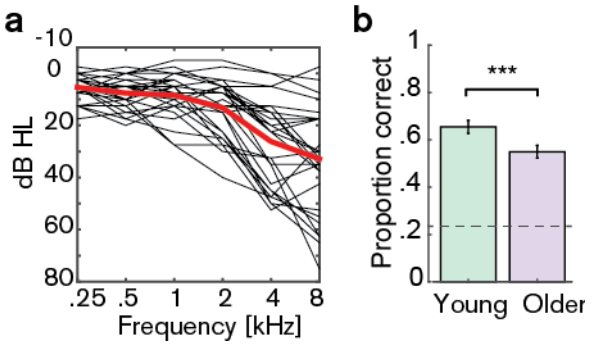
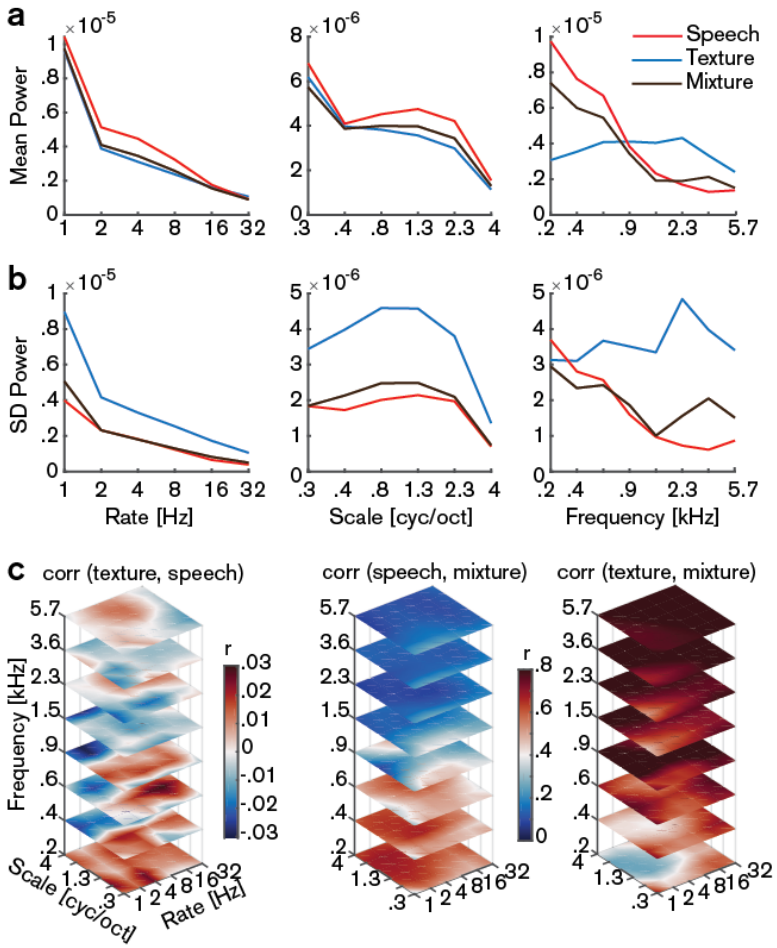
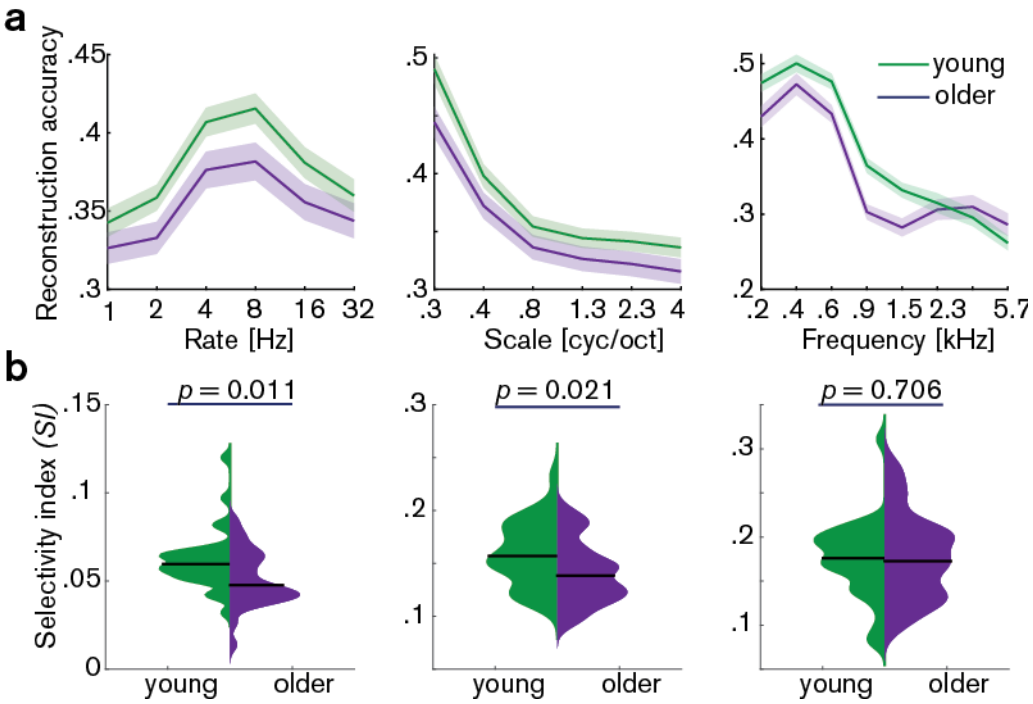


Figure 1 – figure supplement 1. (a) **Audiograms** for older participants (averaged over left and right ear). (b) **Average behavioral performance** (mean ± SEM) in the scanner in a four-choice task (3 questions on the story content after each run) for young (green) and older participants (violet). Chance 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 $d = 0.65$).



837 **Figure 1 – figure supplement 2. Modulation spectra of the stimuli.** (a) Average of the marginalized
838 modulation spectra for the stimulus streams separately and their mixture. (b) Standard deviation (SD) of the
839 marginalized modulation spectra. (c) Pairwise correlations (Pearson's *r*) between the different streams in the
840 modulation representation. While speech and texture are mostly uncorrelated over time, the speech–texture
841 mixture is highly correlated with the speech stream in the low frequency range and with textures in the high
842 frequency range. Note that the left plot has a different scaling than the middle and right plot.



843

844 **Figure 5 – figure supplement 1. Decoding from auditory cortex with ICA(AROMA)-cleaned data.** (a)
845 Mean ± SEM of the MTFs' marginal profiles for rate, scale and frequency for young (green) and older (violet)
846 participants separately. (b) The selectivity index *SI* (see Eq. [7]) was compared per acoustic dimension between
847 age groups using an exact permutation test. Note that irrespective of preprocessing (with or without ICA-
848 cleaning, Fig. 5), reconstruction accuracies peak at temporal rates of 4 – 8 Hz. Likewise, for AROMA-cleaned data
849 *SI* for temporal rates is higher in young compared to older participants. Removing the young-subject outlier for
850 rate *SI* (exceeding the grand average by ± 2 SD) does not alter the *SI* mean age difference qualitatively (p =
851 0.027).

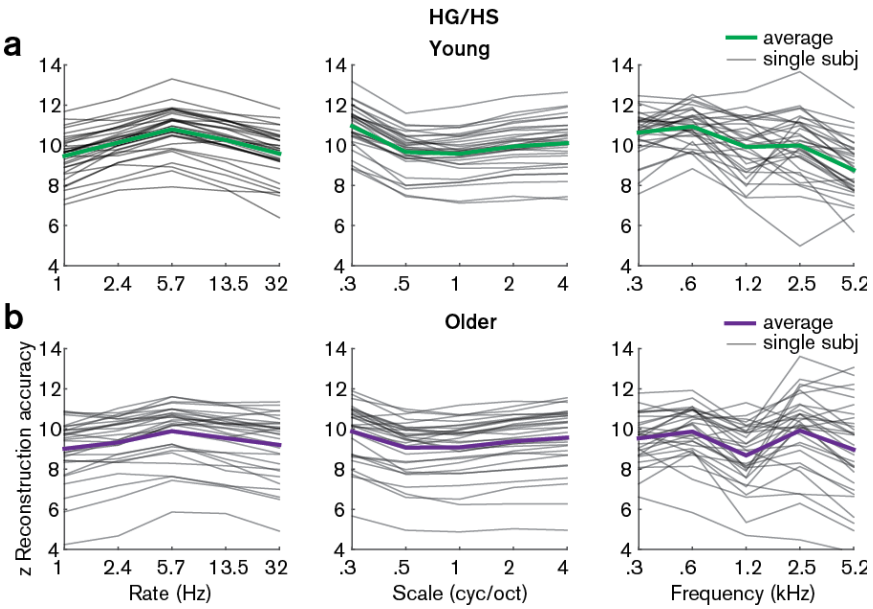


Figure 7 – figure supplement 1. Z-scored reconstruction accuracies from Heschl's gyrus/sulcus. Single subject and median reconstruction accuracy profiles for temporal rate, spectral scale and frequency z-scored with respect to the empirical null distribution for young (a) and older (b) participants. Note that all z-values are > 1.96 and are thus significant at $p < 0.05$.

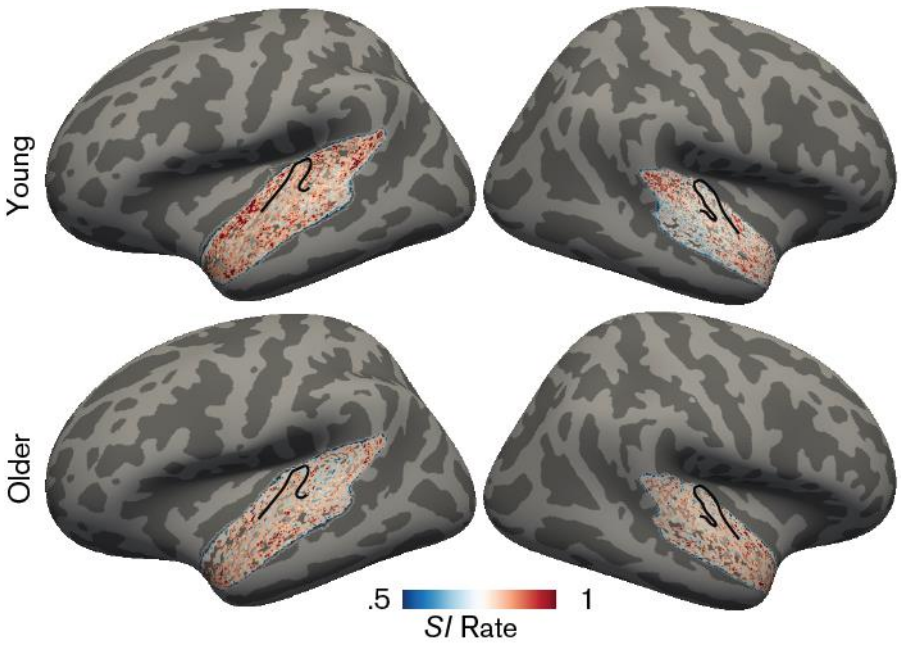
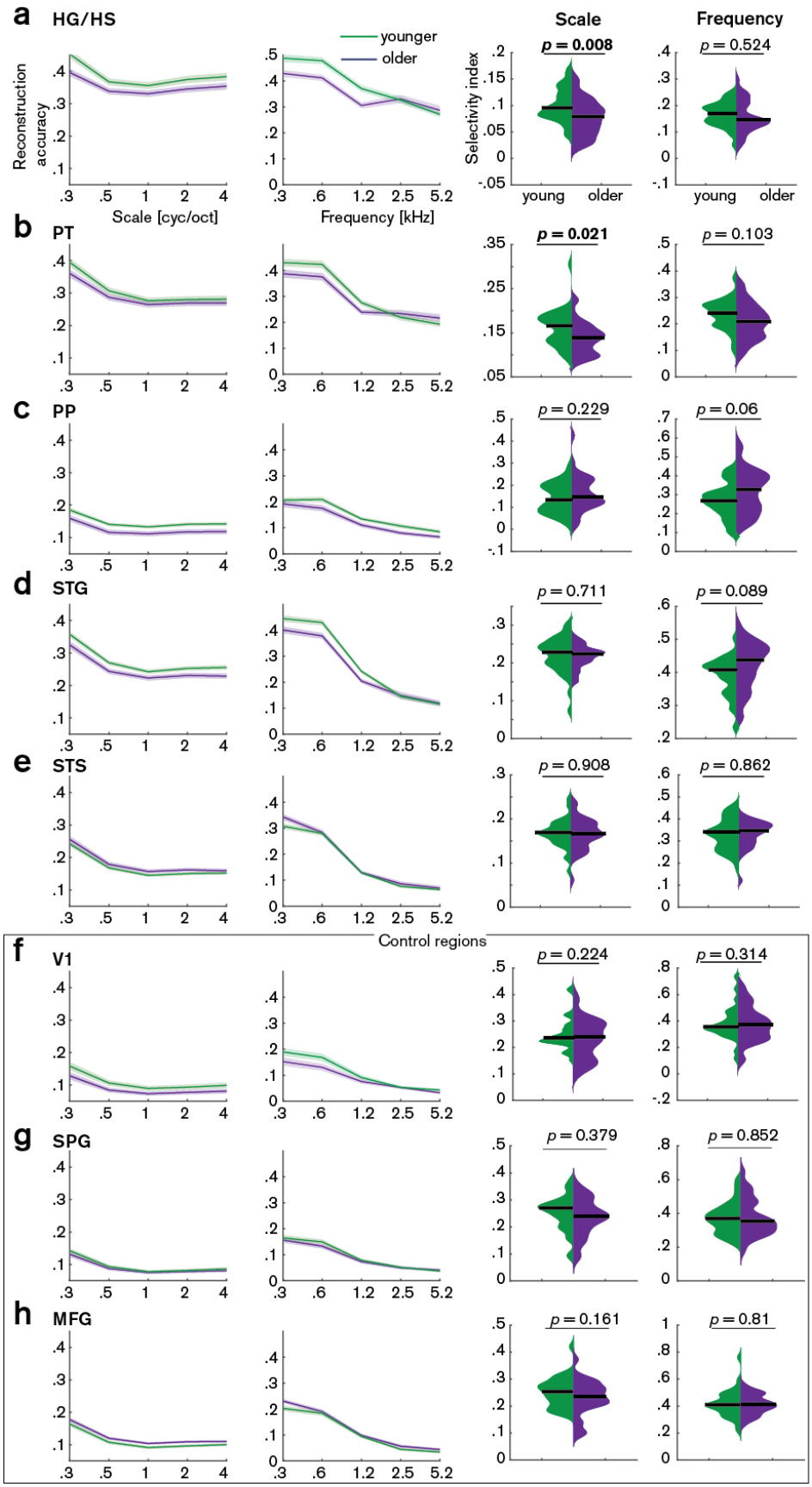


Figure 7 – figure supplement 2. Topographic maps for the rate selectivity index. Maps were derived by calculating the *SI* across the rate profile for a given voxel and averaged in the young group (upper panel) and older group (bottom panel). Higher *SI* (red) indicates higher selectivity. Black outlines indicate Heschl's gyrus.


a HG/HS
younger
older
Scale
Frequency
Reconstruction accuracy
Selectivity index
p = 0.008
p = 0.524
young
older
Scale [cyc/oct]
Frequency [kHz]
b PT
p = 0.021
p = 0.103
c PP
p = 0.229
p = 0.06
d STG
p = 0.711
p = 0.089
e STS
p = 0.908
p = 0.862
Control regions
f V1
p = 0.224
p = 0.314
g SPG
p = 0.379
p = 0.852
h MFG
p = 0.161
p = 0.81

861

862 **Figure 7 – figure supplement 3. Decoding of spectral scale and frequency in ROIs.** Mean (±SE)
863 reconstruction accuracy profiles (left) and selectivity index (right) per age group for spectral scale and frequency
864 in (a) Heschl's gyrus and sulcus (HG/HS), (b) planum temporale (PT), (c) planum polare (PP), (d) superior temporal
865 gyrus (STG), (e) superior temporal sulcus (STS) and the control regions (f) calcarine sulcus (V1), (g) superior
866 parietal gyrus (SPG) and middle frontal gyrus (MFG). Selectivity index was compared per ROI between age
867 groups using exact permutation test. Black line indicates the median *SI*. Note that removing the young outlier
868 participant (> 2 SD away from the mean) for scale in PT reduces the mean age difference to a still significant *SI*
869 difference = 0.015, $p = 0.026$.

870