

Acoustic cue selection and discrimination under degradation: Differential contributions of the inferior parietal and posterior temporal cortices



Mathias Scharinger*, Molly J. Henry, Jonas Obleser

Max Planck Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:

Accepted 23 November 2014

Available online 3 December 2014

Keywords:

Auditory categorization
fMRI parietal cortex
Planum temporale
Acoustic cues
Listening strategies

ABSTRACT

Auditory categorization is a vital skill for perceiving the acoustic environment. Categorization depends on the discriminability of the sensory input as well as on the ability of the listener to adaptively make use of the relevant features of the sound. Previous studies on categorization have focused either on speech sounds when studying discriminability or on visual stimuli when assessing optimal cue utilization. Here, by contrast, we examined neural sensitivity to stimulus discriminability and optimal cue utilization when categorizing novel, non-speech auditory stimuli not affected by long-term familiarity. In a functional magnetic resonance imaging (fMRI) experiment, listeners categorized sounds from two category distributions, differing along two acoustic dimensions: spectral shape and duration. By introducing spectral degradation after the first half of the experiment, we manipulated both stimulus discriminability and the relative informativeness of acoustic cues. Degradation caused an overall decrease in discriminability based on spectral shape, and therefore enhanced the informativeness of duration. A relative increase in duration-cue utilization was accompanied by increased activity in left parietal cortex. Further, discriminability modulated right planum temporale activity to a higher degree when stimuli were spectrally degraded than when they were not. These findings provide support for separable contributions of parietal and posterior temporal areas to perceptual categorization. The parietal cortex seems to support the selective utilization of informative stimulus cues, while the posterior superior temporal cortex as a primarily auditory brain area supports discriminability particularly under acoustic degradation.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Auditory categorization is vital for human behavior in acoustic environments, where auditory stimuli need to be associated with behaviorally relevant meanings. Oftentimes, however, listening conditions are not optimal, and discrimination between two categories may be impeded by internal (e.g. hearing impairment) or external (e.g. background noise) noise that either affects overall discriminability or selectively targets specific stimulus dimensions (e.g. spectral detail; Tuomainen et al., 2013).

Discriminability of auditory stimuli of course depends on the distinctiveness of their acoustic properties. If, for instance, sounds have to be associated with distinct categories A or B, differing in pitch and duration, discriminability of a particular stimulus should improve with greater distance from a hypothetical maximally ambiguous point that could be categorized as A or B with equal likelihood. Stimuli close to this point should be harder to discriminate than stimuli far away from this point. Accordingly, Euclidean distance in acoustic space has been argued

to constitute an appropriate measure of perceptual distance for separable stimulus dimensions, such as pitch and duration (Nosofsky, 1985).

Discriminability may also deteriorate as a result of degradation. Degradation may simultaneously affect all available acoustic cues (e.g. by ambient noise), or selectively affect spectral cues (e.g. pitch), leaving temporal (duration) cues intact (Tuomainen et al., 2013). In this situation, the ability to make use of the most informative cue (i.e., duration, cf. Holt and Lotto, 2006) at the expense of other available cues should prove beneficial.

Previous functional brain imaging studies have investigated aspects of discriminability (Desai et al., 2008; Guenther et al., 2004) as well as aspects of optimal cue utilization in auditory categorization (Hill and Miller, 2010; Pugh et al., 1996; Shaywitz et al., 2001). Auditory categorization in general has been found to recruit the posterior part of the superior temporal gyrus (STG) and the planum temporale (PT; Desai et al., 2008; Griffiths and Warren, 2002; Guenther et al., 2004). Desai et al. (2008) found that activation in the (left) posterior STG (y-values in Talairach space between -30 and -40 , inferior to the planum temporale, cf. Westbury et al., 1999) correlated with the degree to which participants processed stimuli in a categorical way, that is, how readily they could label the respective stimulus. Note that for the remainder of this article, we refer to the posterior STG/STS if y-values of peak coordinates (in Montreal Neurology Institute [MNI] space) are < -15 .

* Corresponding author at: University of Leipzig, Institute of Psychology BioCognitive Incl. Biological Psychology, Neumarkt 9-19, D-04109 Leipzig, Germany.

E-mail address: mathias.scharinger@uni-leipzig.de (M. Scharinger).

Several studies on speech processing identified the posterior STG, including parts of the PT, as subserving categorical (speech-specific) processing (Chang et al., 2010; Jäncke et al., 2002; Turkeltaub and Coslett, 2010). While these studies suggest that the posterior STG is sensitive to aspects of discriminability (ambiguity) in auditory categorization, it is less clear whether activity in this region would scale with more fine-grained differences in discriminability between non-speech stimuli.

A recurring cortical site involved in selective stimulus cue utilization is the parietal cortex, comprising the inferior parietal lobe and intraparietal sulcus (Geng and Mangun, 2009; Hill and Miller, 2010; Pugh et al., 1996; Rinne et al., 2007; Shaywitz et al., 2001), situated within a larger network associated with executive function (Corbetta et al., 2000; Posner and Dehaene, 1994). Aside from its role in spatial auditory attention (Alain et al., 2007; Brunetti et al., 2008), the IPL has been found to support the utilization of informative acoustic cues in a given listening situation (e.g. Henry et al., 2013). Despite a strong focus on visual processing while investigating the IPL, some studies suggest that this region subserves modality-independent functions of attention switching and object representation (Cusack et al., 2000, 2010).

In this study, we were interested in the role of the posterior STG in responding to fine-grained differences in discriminability of non-speech stimuli, and in the role of the IPL for optimal acoustic cue utilization depending on varying listening situations. Using non-speech stimuli ensured that categories were not *overlearned*; that is, participants had no prior experience with our experimental stimuli. In contrast to previous studies, our stimulus distributions and manipulation allowed us to examine discriminability and acoustic cue utilization within one experimental paradigm. We hypothesized that if the posterior STG supports discriminability in auditory categorization, activation there should be sensitive to changes in discriminability of auditory, non-speech stimuli. Further, if the IPL subserves optimal cue utilization in a domain-general manner (Henry et al., 2013), shifts in informativeness of acoustic cues should lead to increased activation in this region.

Materials and methods

Participants

Thirty-six healthy volunteers were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences (all right-handed, 18 females, age range 20–31 years, mean age 25.7; standard deviation [SD] = 2.8 years). They were native speakers of German with no self-reported hearing impairments or neurological disorders. Participants gave written informed consent and received financial compensation for their participation. All procedures followed the guidelines of the local ethics committee (University of Leipzig) and were in accordance with the Declaration of Helsinki.

Stimuli

Auditory stimulus tokens differed in spectral and durational properties. The base for all stimuli was an inharmonic narrow-band sound composed of 17 components. The lowest component frequency was

500 Hz, and the frequencies of the additional 16 components were related to each other by a ratio of 1.15 (Fig. 1A, bottom; Goudbeek et al., 2009; Scharinger et al., 2014). Spectral manipulations constituted filtering the base sound with a band-pass filter (second-order infinite-impulse response, IIR) with a single frequency peak that was unique

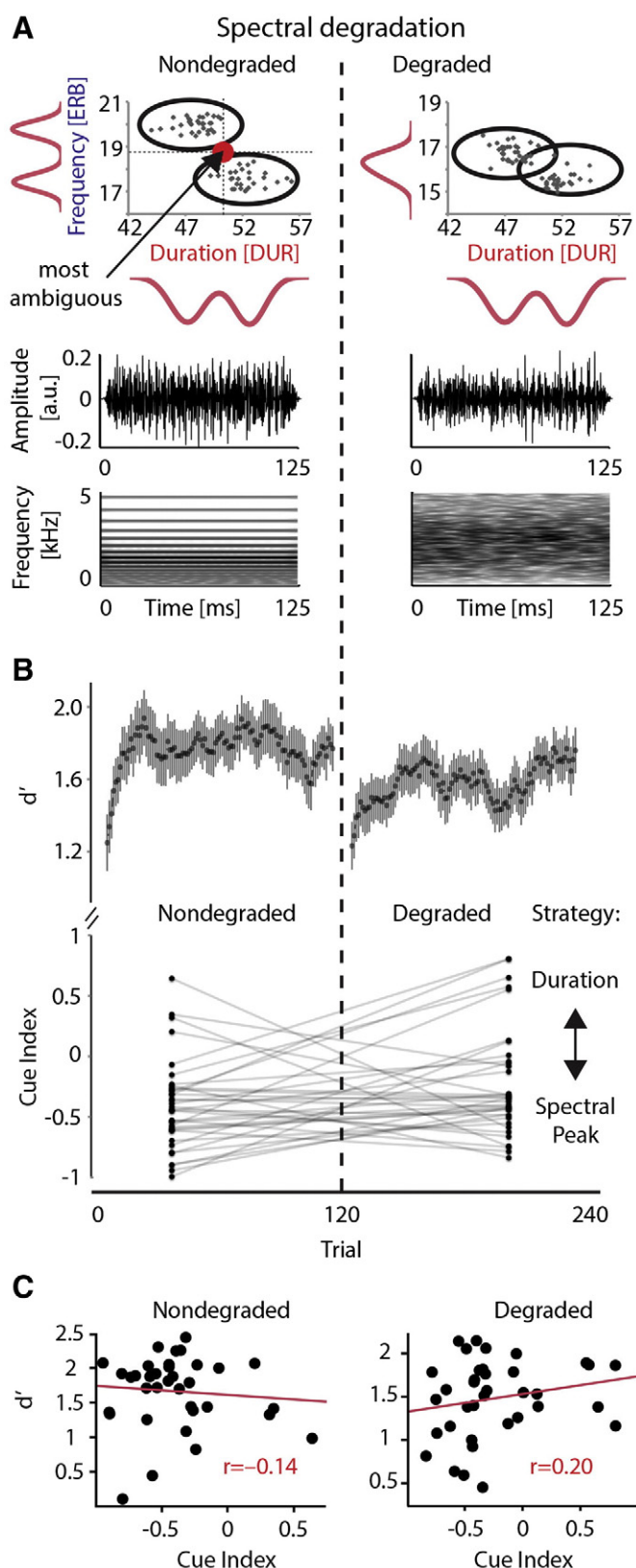


Fig. 1. Stimulus characteristics and behavioral results. **A.** Top: Representation of sounds with varying spectral peaks (ERB; y-axis) and durations (DUR; x-axis). Distributions are indicated by ellipses; black dots show exemplary distributions for a representative participant. Bottom: Acoustic properties of sounds in the nondegraded (left) and degraded (right) conditions. Duration and amplitude envelope were unaffected by degradation, while spectral properties were smeared. **B.** Behavioral performance results. Top: Perceptual sensitivity (d') over time, obtained from sliding windows over nondegraded and degraded trials per participant (window size = 20 trials, step size = 1 trial). Error bars show the standard error of the mean. Bottom: Comparison of cue indices across conditions; black dots show individual participant data. Mean cue index values between conditions are connected for each participant. **C.** Correlations of d' and cue index in the nondegraded (left) and degraded (right) conditions.

for each stimulus; bandwidth was 1/5 of the center frequency. We use the term ‘spectral peak’ to refer to the filters’ center frequency and its resulting spectral stimulus properties. Duration manipulations constituted changes to the sounds’ physical durations.

Sound categories, labeled “A” and “B”, differed in terms of the sounds’ spectral peak and duration (Fig. 1A): For each sound in categories A and B (each category was comprised of 1000 exemplars), spectral peaks and durations were randomly drawn from bivariate normal distributions, with equal standard deviations, σ , and means, μ , that differed between categories. Spectral peak frequencies were converted to equivalent rectangular bandwidth (ERB; approximating the bandwidths of the auditory filters in human hearing, cf. Glasberg and Moore, 1990), and sound durations were log 10-transformed for reasons of psycho-physical comparability (DUR; cf. Smits et al., 2006). Table 1 shows the respective means of categories A and B.

In the first half of the experiment (*nondegraded condition*), sound distributions did not overlap in spectral peak, but 1/3 of the sounds in categories A and B overlapped in duration (Fig. 1A, top). Separating categories more along the spectral than the duration dimension was intended to bias participants to focus on spectral peak as the most informative cue for categorization because maximum accuracy could only be achieved if categorization was based on spectral peak. Halfway through the experiment, spectral cues were degraded by applying four-band noise vocoding to the original stimuli (Drullman et al., 1994; Shannon et al., 1995; degraded condition). Noise-vocoding was applied to all stimuli, irrespective of category, as described in Rosen et al. (1999). In short, this procedure involves filtering the raw signal into frequency bands (here, four), extracting the amplitude envelope from each band and reapplying it to bandpass-filtered noise carriers, leading to a smearing of spectral detail. Precise vocoding settings were identical to those reported in Erb et al. (2012). Most important to the present study, degraded stimuli maintained their amplitude-envelope features and their original duration, while they showed a change in the location and spread of the spectral peak (Table 1; Figs. 1A, B). After degradation, participants were expected to categorize sounds on the basis of duration, as maximum accuracy after degradation was only achievable by utilizing duration cues.

All stimuli were normalized to equal root-mean-square (RMS) intensity and presented at ~60 dB SPL. Linear onset and offset ramps (5 ms) ensured that acoustic artifacts were minimized.

We took Euclidean distance of each stimulus to the overall median point (i.e., the most ambiguous point in the stimulus space) as a stimulus-based measure of discriminability (Fig. 1A). This median point corresponded to a stimulus that was never presented, and that could have been categorized as A or B with equal likelihood. Shorter distances to the median make individual sounds harder to categorize (less discriminable), while longer distances render them easier to categorize (more discriminable). The selection of sounds from Gaussian distributions ensured that no systematic differences in Euclidean distances existed between categories A and B ($t(1999) = 0.86$, $p = 0.43$, effect size $r_e = 0.02$). However, as expected, Euclidean distances to the median differed between the nondegraded and degraded distributions ($t(1999) = 49.50$, $p < 0.001$, $r_e = 0.74$; Fig. 1A), with smaller distances in the degraded than in the nondegraded condition, thus, categorization should be harder in the degraded than in the nondegraded condition.

Effect sizes are reported as $r_{\text{equivalent}}$ (throughout, r_e), which is equivalent to a Pearson product-moment correlation for two continuous variables, to a point-biserial correlation for one continuous and one dichotomous variable, and to the square root of η^2 (eta-squared) for ANOVAs. We used the by-item measure *discriminability* (i.e. Euclidean distance of each stimulus to the median point) as a parametric modulator in the first-level fMRI analyses reported below.

Experimental procedure

Participants were first familiarized with the categorization task and had to complete a short practice block consisting of 20 sounds that did not occur in the main experiment (10 from category A and 10 from category B). The subsequent main experiment was arranged in four runs (two per condition). In each run, 60 sound exemplars, randomly drawn from categories A and B with equal probability, were presented in a sparse imaging design (Hall et al., 1999). The sparse design was chosen in order to guarantee that sounds could be presented during silent periods between the acquisitions of echo-planar images (EPIs).

Sounds were presented on average 2 s after the offset of a preceding EPI volume acquisition (± 500 ms). Subsequently, a visual response prompt (green traffic light) was presented on a screen 3 s after the stimulus onset. Participants were then required to indicate whether the presented sound belonged to category A or category B by pressing one of two keys on a button box; button assignment was counterbalanced across participants. Following the response, participants received corrective feedback (“Correct”/“Incorrect”), which was displayed for 1 s in the middle of the screen. Seven seconds after the onset of an acoustic stimulus, a subsequent EPI volume ($TA = 2$ s) was acquired. Also, in each run, 15 silent trials for which no response was required occurred at random positions. This corresponded to 20% of all trials. The duration of the entire experiment with short breaks between runs was 50 min. Participants were not asked to focus on specific sound properties (e.g. frequency, duration). Participants were further told that they should maintain their categorization even though it would be possible that the quality of the sounds changed within the experiment. They were also instructed that sometimes, sound presentations would be missing (silent trials) in which case no response was required.

Imaging data acquisition

Functional MRI data were recorded with a Siemens VERIO 3.0-T MRI scanner equipped with a 12-channel head coil while participants performed the categorization task in supine position inside the scanner. Fifteen participants were additionally equipped with an MR-conform electroencephalography (EEG) cap for data acquisition reported elsewhere (Scharinger et al., 2014). For acoustic transmission, participants wore MR-compatible headphones (MR-confon GmbH, Magdeburg, Germany) together with in-ear hearing protection (Hearsafe Technologies GmbH, Cologne, Germany), reducing scanner noise by approximately 16 dB. A custom-made mirror and an LCD projector system were used to project the visual screen display.

Whole-brain EPIs (30 axial slices, thickness = 3 mm, gap = 1 mm) were collected every 9 s ($TE = 30$ ms; flip angle = 90° ; field of view = 192×192 mm; voxel size = $3 \times 3 \times 4$ mm). Seventy-five volumes

Table 1

Average ($\pm$ standard deviations in parentheses) of the sounds’ properties, spectral peak and duration in categories A and B. Degradation was achieved by 4-band noise-vocoding as described in the text.

Stimulus category	Nondegraded		Degraded	
	A	B	A	B
Spectral peak [ERB]	20.00 (± 0.35)	17.00 (± 0.36)	16.80 (± 0.34)	15.50 (± 0.27)
Spectral peak [Hz]	1739 (± 75)	1196 (± 60)	1166 (± 50)	984 (± 36)
Duration [DUR]	47.70 (± 1.28)	52.53 (± 1.32)	47.70 (± 1.28)	52.53 (± 1.32)
Duration [ms]	118 (± 15.2)	191 (± 25.6)	118 (± 15.2)	191 (± 25.6)

(sound and null trials) were acquired in each of the 4 sessions, yielding 300 volumes of interest in total. For anatomical localization and volume co-registration, high-resolution, 3D MP-RAGE T1-weighted scans were taken from the Max Planck Institute participant database. These scans had been collected on a 3 T Siemens TIM Trio scanner with a 12-channel head coil, on average 29 months prior to the experiment ($SD = 18$ months), and shared the following acquisition parameters: sagittal slices = 176, repetition time = 1300 ms, TE = 3.46 ms, flip angle = 10° , acquisition matrix = 256×240 , voxel size = $1 \times 1 \times 1$ mm.

In order to obtain better image distortion correction on the basis of voxel-displacement-maps (Hutton et al., 2002; Jezzard and Balaban, 1995), field maps (30 axial slices, thickness = 3 mm, gap = 1 mm, repetition time = 488 ms, TE1 = 4.92 ms, TE2 = 7.38 ms, flip angle = 60° , field of view = 192×192 mm, voxel size = $3 \times 3 \times 3$ mm) were recorded prior to the functional volume acquisition.

Analysis of behavioral data

Our behavioral dependent measures were *overall performance* and *cue utilization*. Overall performance was estimated by d' , a measure of perceptual sensitivity that is independent of response bias. Perceptual sensitivity, d' , was calculated from proportions of hits and false alarms (Macmillan and Creelman, 2005), where hits were defined as “category-A” responses to category-A stimuli, and false alarms were defined as “category-A” responses to category-B stimuli. Perceptual sensitivity was calculated separately for each experimental run; d' values were then averaged across blocks for each participant separately for the nondegraded and degraded conditions.

The *cue utilization* measure quantified the degree to which individual participants relied on the spectral vs. durational stimulus aspects in the nondegraded and degraded conditions. To this end, we first calculated logistic regressions for the nondegraded and degraded conditions of the experiment with category-A responses as the dependent measure and spectral and duration stimulus values as independent measures. The slope of the logistic function, expressed by absolute β , indicated the degree to which spectral peak or duration influenced the categorical response ($\beta_{\text{spectral peak}}$; β_{duration} ; Gougoux et al., 2009; Scharinger et al., 2013). Second, for each condition of the experiment, we expressed the bias for spectral versus durational cue utilization by a *cue index* that was calculated as shown below.

$$\text{Cue index} = \frac{|\beta_{\text{duration}}| - |\beta_{\text{frequency}}|}{|\beta_{\text{duration}}| + |\beta_{\text{frequency}}|} \quad (1)$$

According to Eq. (1), a positive cue index reflects the tendency to use durational cues more than spectral cues, while a negative cue index reflects the tendency to use spectral cues more than durational cues.

Analysis of imaging data

Functional (T2*-weighted) and structural (T1-weighted) images were processed using Statistical Parametric Mapping (SPM8; Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College of London). Functional images for each run were first realigned using the 6-parameter affine transformation in translational (x, y, and z) and rotational (pitch, roll, and yaw) directions to reduce individual movement artifacts (Ashburner and Good, 2003). Subsequently, a mean image of each run was used to estimate unwarping parameters, together with voxel-displacement-maps (VDMs) obtained from individually recorded field maps in order to account for magnetic field deformations (Hutton et al., 2002; Jezzard and Balaban, 1995). Participants' structural images were manually pre-aligned to a standardized EPI template (Ashburner and Friston, 2004) in MNI space (Montreal Neurological Institute) in order to improve co-registration and normalization accuracy. Next, functional

images were co-registered to the corresponding participants' structural images and normalized to MNI space. Normalization was based on segmented structural T1-images (gray matter, white matter, and cerebrospinal fluid) and used a 12-parameter affine transformation, where the parameters constitute a spatial transformation matrix obtained from the co-registration algorithm. Functional images were then smoothed using an 8-mm full-width half-maximum Gaussian kernel and subsequently used for first- and second-level general linear model (GLM) analyses.

At the first level, a general linear model was estimated for each participant using a first-order finite impulse response (FIR; window = 2 s) as the basis function and high-pass filtered with a cut-off of 128 s. The design matrix included regressors for 1) *sound trials* (corresponding to volumes following sound presentations), 2) the mean-centered single-trial parametric modulator *median distance*, and 3) *silent trials* (corresponding to volumes following null trials), specified separately for each of four runs (two nondegraded and two degraded). Experimental runs were included as regressors of no interest (one for each run). Nondegraded and degraded trials were thus modeled in one design matrix at the first level. Six additional regressors of no-interest accounted for movement artifacts in translational (x, y, and z) and rotational (pitch, roll, and yaw) directions.

Resulting T-maps were restricted to gray- and white matter obtained from group averages based on individual T1-weighted scans. Still at the first level, we calculated the contrasts of *sound trials* and parametric modulator *median distance* against the implicit baseline (mean activation). Contrasts were based on means for the first two runs (nondegraded condition), and on means for the last two runs (degraded condition). Conditions were compared by means of the nondegraded > degraded and degraded > nondegraded contrasts for sound trials and median distance.

At the second level, all contrasts were compared against zero using one-sample *t*-tests. Additionally, sound-trial contrasts (against implicit baseline) from the first level were correlated with cue indices across participants for the nondegraded and degraded conditions. This was done in order to examine how individual cue utilization modulated blood oxygenation level dependent (BOLD) responses. In a separate analysis, the impact of degradation on the coupling of BOLD to behavioral cue utilization was assessed by correlating, on the second level, the contrast images for the first-level degraded–nondegraded difference with the degraded–nondegraded difference in the cue index. Thereby, we obtained a within-participant measure of cue-utilization change.

When testing for brain regions that were involved in the auditory categorization task generally, we applied a family-wise error (FWE) corrected threshold of $p < 0.01$ (cluster-wise, based on Gaussian random fields) at the second level. For all other second-level analyses, we used an uncorrected threshold of $p < 0.005$ combined with a cluster extent of 22 voxels, which corresponds to a whole-brain alpha of $p < 0.05$, as determined using a MATLAB-implemented Monte Carlo simulation (Slotnick et al., 2003) with a smoothing of 8 mm for all comparisons (full width half maximum of the Gaussian smoothing kernel).

In order to illustrate significant effects and interactions, regression beta values were extracted from regions of interest (ROIs). These regions were defined using the SPM toolbox MarsBaR (Brett et al., 2002) as spheres with 5 mm radii and centers corresponding to the peak coordinates (in MNI space) identified in the whole-brain analyses (see Results). Determination of anatomical locations was based on the Automated Anatomical Labeling Atlas (AAL; Tzourio-Mazoyer et al., 2002), and (for areas involving the planum temporale) the Westbury Atlas (Westbury et al., 1999).

Results

Behavioral data

Participants categorized the sounds above chance level (average $d' = 1.58$, *t*-test against zero $t(35) = 23.12$, $p < 0.001$, $r_e = 0.97$).

Perceptual sensitivity differed between conditions (nondegraded: $d' = 1.67 \pm \text{standard error of the mean [SEM]} = 0.09$ vs. degraded: $d' = 1.49 \pm 0.07$; $t(35) = 2.10$, $p = 0.04$, $r_e = 0.33$). Average cue indices imply that participants relied less on spectral peak (i.e., more on duration) cues in the degraded (average cue index = -0.24 ± 0.07) than in the nondegraded (cue index = -0.41 ± 0.06) condition. However, the preference for making category membership decisions based on spectral peak cues in the first, nondegraded condition was not entirely abandoned in the second, degraded condition (cue index difference $t(35) = 1.90$, $p = 0.07$, $r_e = 0.31$; Fig. 1B). Correlations between d' and cue index were not significant. Their signs, however, suggested improved performance when relying on spectral peak in the nondegraded condition ($r = -0.14$, $t(35) = 0.55$, $p = 0.58$) and when relying on duration in the degraded condition ($r = 0.20$, $t(35) = 1.21$, $p = 0.23$; Fig. 1C).

Brain imaging results

Sound activation in the fronto-parietal network

Sound discrimination drove a bilateral fronto-parietal network, with activations in the inferior parietal lobule (IPL, Brodmann area [BA] 40), insula (BA 13), anterior prefrontal cortex (BA 10), and mid- and anterior cingulate (BA 32 and BA 24). In the nondegraded condition, there was additional activation in the right inferior frontal gyrus (IFG, BA 46), left thalamus (medial dorsal nucleus) and right cerebellum, while in the degraded condition, there were further clusters in the left precentral gyrus (BA 6) and right PT (46–65% within-PT probability according to Westbury et al., 1999; Fig. 2A). A direct test of activation differences between the nondegraded and degraded conditions revealed most notably clusters in the left dorso-lateral prefrontal cortex and right thalamus for the contrast nondegraded > degraded, and clusters in the right insula and STG (BA 21) for the contrast degraded > nondegraded (Fig. 2B). A full list of significant clusters is given in Table 2.

Discriminability effects in temporal and frontal cortices

A test for sensitivity to acoustic distance to the overall median point (i.e., the most ambiguous point in the stimulus space) revealed a cluster in the left IPL (BA 40) in the nondegraded condition, and several clusters in the bilateral STG and middle temporal gyrus (MTG, comprising BA 21, 22, 37 and 38, in the vicinity of PT with 5–25% within-PT probability) in the degraded condition (Fig. 3A). The reverse contrast (showing more activation for smaller distances from the most ambiguous point) yielded clusters in the frontal (insula) and anterior prefrontal cortices (SFG, BA 10) in the nondegraded condition, and clusters in the right mid-cingulate (BA 32) and left insula (BA 13) in the degraded condition (Fig. 3B).

When directly testing for changes in sensitivity to acoustic median distance in the degraded condition versus the non-degraded condition, a region corresponding to the right posterior STG/STS in the vicinity of the planum temporale (with 5–25% within-PT probability)/right MTG was identified. Beta values extracted from this area indicated that the observed difference was driven by increased sensitivity to acoustic median distance in the degraded condition (Fig. 3C). This is notable as degradation led to overall *reductions* in discriminability (i.e., acoustic distance to the median), yet the right pSTG/STS/PT/MTG reacted to this by increasing sensitivity to this acoustic parameter.

Effects of cue index in the prefrontal and parietal cortices

In the nondegraded condition, BOLD activity increased with increasing utilization of spectral cues in the left posterior cingulate cortex and left anterior prefrontal cortex (BA 10). In the degraded condition, BOLD activity increased with an increasing preference for duration cues (or to less reliance on spectral cues) in the bilateral orbito-frontal cortex (BA 11), left precentral gyrus (BA 6) and right IPL (BA 40, extending into the supramarginal gyrus, Fig. 4A). Differences in the correlations between conditions yielded a cluster in the left parietal cortex. This cluster extended rostrally into the postcentral sulcus and gyrus, dorsally into the intraparietal sulcus, and caudally into the inferior

parietal lobe (IPL), with a further peak in the postcentral gyrus (BA 3/4). We refer to the entire area as IPL; for a similar labeling, see Livesey et al. (2007); Table 3 and Fig. 3.

Beta values extracted from the significant cluster showed that the cue index exerted a stronger effect on IPL activation in the degraded as compared to the nondegraded condition (Fig. 4B). Using a more lenient threshold ($p < 0.05$), the interaction was also seen in a homologue cluster in the contra-lateral hemisphere and additional clusters in the mid-cingulate (BA 31), supplementary motor area (SMA) and right middle temporal gyrus (MTG, BA 22; illustrated in Fig. 4B). All clusters for the cue index effects are given in Table 3.

Discussion

The most important finding of this study is that different aspects of auditory categorization – discriminability of sensory input and utilization of acoustic cues – are differentially supported by temporal and parietal areas: Changes in cue utilization implicated the left parietal cortex (IPL), while differences in discriminability were correlated with activity in the right pSTG/STS/PT when the acoustic space was compressed due to spectral degradation. We will turn to a more detailed discussion regarding the involvement of these areas during auditory categorization in the subsequent sections.

Cue utilization draws on the parietal attention network

Stronger preferences for duration cues under degradation correlated with increased activity in the inferior parietal lobule (BA 40), including parts of the supramarginal gyrus. Thus, the more likely an individual was to use duration cues under spectral degradation (expressed by a more positive cue index), the larger the activation in the left IPL. Further, the coupling of duration-cue utilization and BOLD activity in the degraded (compared to the nondegraded) condition was stronger in an overlapping cluster, including parts of the inferior parietal lobe (IPL, Fig. 4B). Since this comparison was based on the within-participant difference of cue indices between the nondegraded and degraded conditions, this finding suggests that the IPL supported the change in cue utilization.

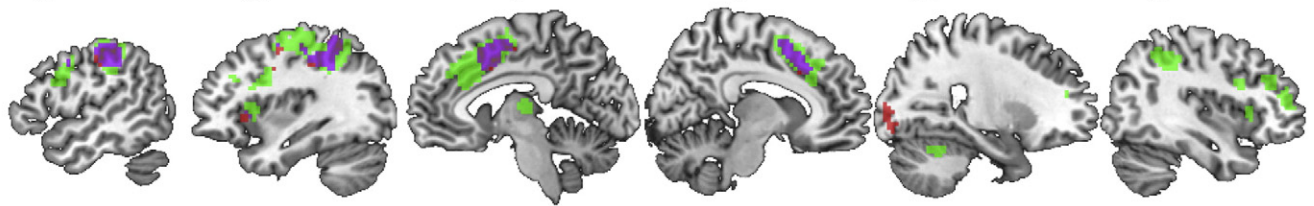
The IPL as part of the fronto-parietal executive network (Corbetta et al., 2000; Posner and Dehaene, 1994) has repeatedly been found to be engaged in situations that required a flexible deployment of neural resources to informative stimulus features (Geng and Mangun, 2009; Gillebert et al., 2012; Hill and Miller, 2010; Schultz and Lennert, 2009). In our experiment, before spectral degradation, the most informative acoustic cue for categorization was spectral peak, while stimulus duration could be used as a secondary cue. However, participants could perform best if they assigned stimulus duration more weight after spectral degradation. The cue index clearly demonstrated that participants in fact mostly relied on spectral peak in the nondegraded condition. In contrast, this reliance was reduced under spectral degradation in favor of duration cues.

One interpretation of co-varying IPL activation and duration utilization under degradation is that the IPL supported the relative change in cue weighting, assigning more importance to duration than before. Importantly, we cannot claim that the IPL would specifically be sensitive to duration cues – under our hypothesis, IPL should be similarly involved if participants had to assign more weight to spectral cues, a prediction that needs to be tested in future research. Note further that degradation alone could not have been responsible for IPL activation: The comparison between sound activation in the degraded versus nondegraded condition did not yield any parietal clusters. Even though degradation resulted in more effortful processing (as suggested by decreasing perceptual sensitivity), only the cue index correlation yielded significant clusters in the IPL. Furthermore, the fact that there were no significant correlations between d' and cue indices in the nondegraded and degraded conditions suggests that the coupling of IPL activity with change of cue utilization is not simply based on more successful categorization or more positive feedbacks and less errors. Finally, we

A Auditory categorization

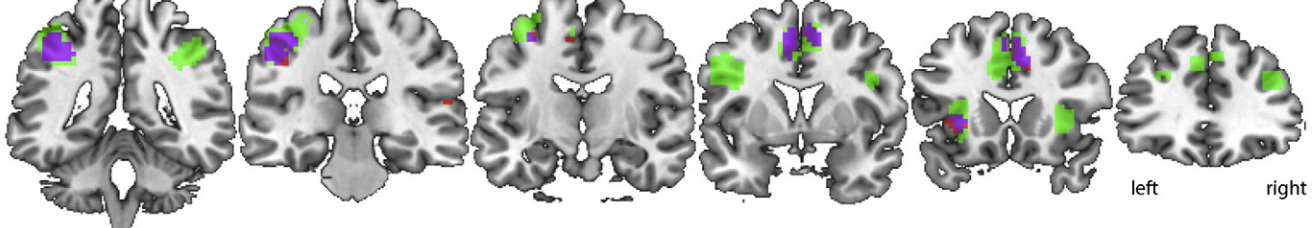
left – right

x: -51 -38 -8 8 24 39



caudal – rostral

y: -41 -26 -11 4 19 34



ventral – dorsal

z:

-2

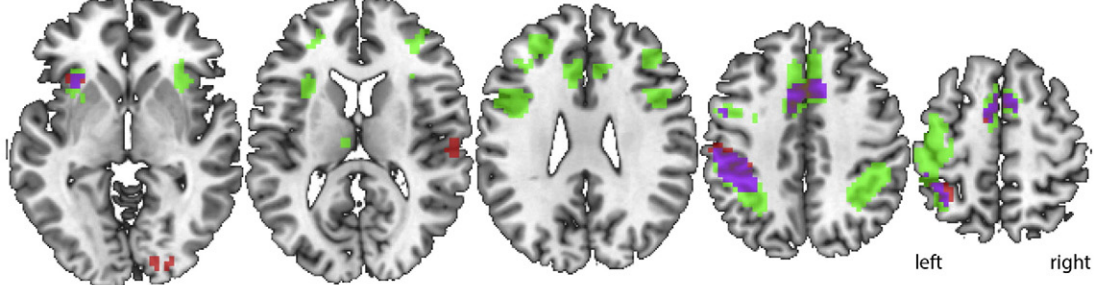
13

26

43

56

nondegraded
degraded
co-activation
nondegraded/
degraded



B Condition comparison

x=51

x=12

y=-15

y=13

z=30

nondegraded > degraded
degraded > nondegraded

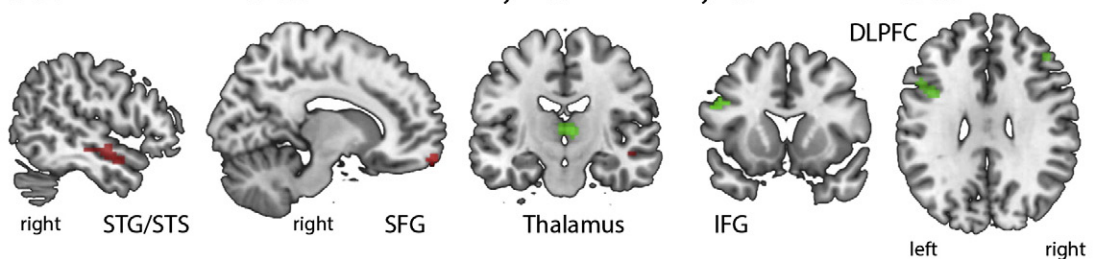


Fig. 2. A. Overall activation by auditory categorization in the nondegraded (green) and degraded (red) conditions ($p < 0.01$, FWE-corrected). Areas of co-activation are illustrated in violet. B. Condition differences for the contrasts nondegraded > degraded (green) and degraded > nondegraded (red, $p < 0.005$, extent threshold $k > 22$). Abbreviations: STG/STS – superior temporal gyrus/sulcus, SFG – superior frontal gyrus, IFG – inferior frontal gyrus, DLPFC – dorsolateral prefrontal cortex.

re-calculated the cue index/BOLD correlation differences between degraded and nondegraded conditions, regressing out d' . The results of this analysis revealed an almost identical (even slightly bigger) cluster to the one illustrated in Fig. 4B and Table 3. We therefore suggest that the IPL was not engaged due to task difficulty per se, but due to the necessity to change cue utilization after degradation.

These results add to previous research showing the IPL's involvement in auditory processing (Gaab et al., 2006; Husain et al., 2006; Jacquemot et al., 2003; Kiefer et al., 2008; Obleser et al.,

2012). Furthermore, our findings suggest that the IPL is a rather domain-general area with respect to cue utilization: While most previous studies on the parietal cortex focused on the visual domain (Corbetta et al., 2000; Yantis, 1993, 2008), the present data indicate that the parietal cortex also supports the utilization of informative acoustic cues. This is in line with studies that have provided evidence for a more modality-independent function of the IPL and inferior parietal sulcus (IPS) with regard to object representation and attention switching (Cusack et al., 2000, 2010).

Table 2

Significant BOLD activation in the sound vs. baseline contrast for the nondegraded and degraded conditions (thresholded at $p < 0.01$, FWE corrected). Comparison between conditions are thresholded more liberally (at $p < 0.005$, with $k > 22$). Peak activations are given in MNI coordinates. Abbreviations: IPL – inferior parietal lobule, MFG – middle frontal gyrus, APFC – anterior prefrontal cortex, IFG – inferior frontal gyrus, ling – lingual gyrus, STG – superior temporal gyrus, STS – superior temporal sulcus, SFG – superior frontal gyrus, PT – planum temporale, DLPFC – dorso-lateral prefrontal cortex.

Contrast	Region	MNI coordinates			Z-value	Size (voxels)
Auditory categorization (nondegraded)	l. Cingulate	–6	5	49	6.93	643
	l. IPL	–48	–34	43	6.92	1020
	r. IPL	42	–37	46	6.7	192
	l. Insula	–30	14	1	6.61	120
	r. Insula	30	20	–2	6.5	101
	r. APFC	33	41	16	6.21	141
	l. Thalamus	–9	–19	7	6.11	39
	l. APFC	–30	47	25	6.02	107
	r. IFG	48	11	22	5.93	87
	r. Cerebellum	21	–55	–23	5.73	56
Auditory categorization (degraded)	l. IPL	–45	–34	43	6.83	400
	r. Cingulate	3	14	52	6.14	249
	r. Ling.	24	–91	1	5.84	51
	l. Insula	–33	20	1	5.63	29
	l. Precentral	–42	–1	37	5.56	31
	l. MFG	–33	–7	52	5.47	28
	r. PT	60	–16	4	5.42	31
Direct comparison Nondegraded > degraded	r. MFG	42	35	28	3.44	51
	r. Thalamus	6	–16	4	3.38	34
	l. DLPFC	–42	8	28	3.38	64
	r. Cerebellum	3	–55	–35	3.3	23
Degraded > nondegraded	r. Insula	27	5	–20	3.41	39
	r. SFG	12	62	–17	3.41	22
	r. STG/STS	51	–1	–17	3.12	58

Discriminability under degradation involves the planum temporale

The right posterior temporal cortex in the vicinity of the PT was most sensitive to stimulus discriminability. This effect also became stronger once degradation was introduced (Fig. 3A), which is remarkable as discriminability was acoustically reduced by spectral degradation. We expected that discriminability should decrease with decreasing distance to the median position in the two-dimensional acoustic space where the assignment of either category label, A or B, was equally likely. The behavioral results provided supporting evidence for this: Spectral degradation lead to an overall reduction of median distances, and thus, discriminability. This was accompanied by a deterioration of performance, as seen in significantly lower d' values in the degraded compared to the nondegraded condition. Nevertheless, the posterior STG/STS and PT showed *increased* sensitivity to this acoustic parameter under degradation, which suggests a role for these areas in supporting discrimination in a compressed acoustic space.

Our finding is in line with a thread of previous results illustrating the role of the posterior temporal cortex for categorical speech sound processing and illustrating the role of the PT in auditory categorization. The PT and surrounding posterior temporal regions have been assigned an important role in forming and discriminating auditory categories (Desai et al., 2008; Griffiths and Warren, 2002; Guenther et al., 2004; Husain et al., 2006; Obleser and Eisner, 2009). In speech processing, the PT and pSTG/STS seem to be particularly sensitive to the ability to assign category labels (Chang et al., 2010; Dehaene-Lambertz et al., 2005; Desai et al., 2008; Turkeltaub and Coslett, 2010). For instance, Dehaene-Lambertz et al. (2005) showed that if acoustic stimuli could be assigned speech-relevant labels, activity was increased in the posterior parts of the left STG, compared to a situation where no such labeling was possible. In a similar vein, Chang et al. (2010) compared within-

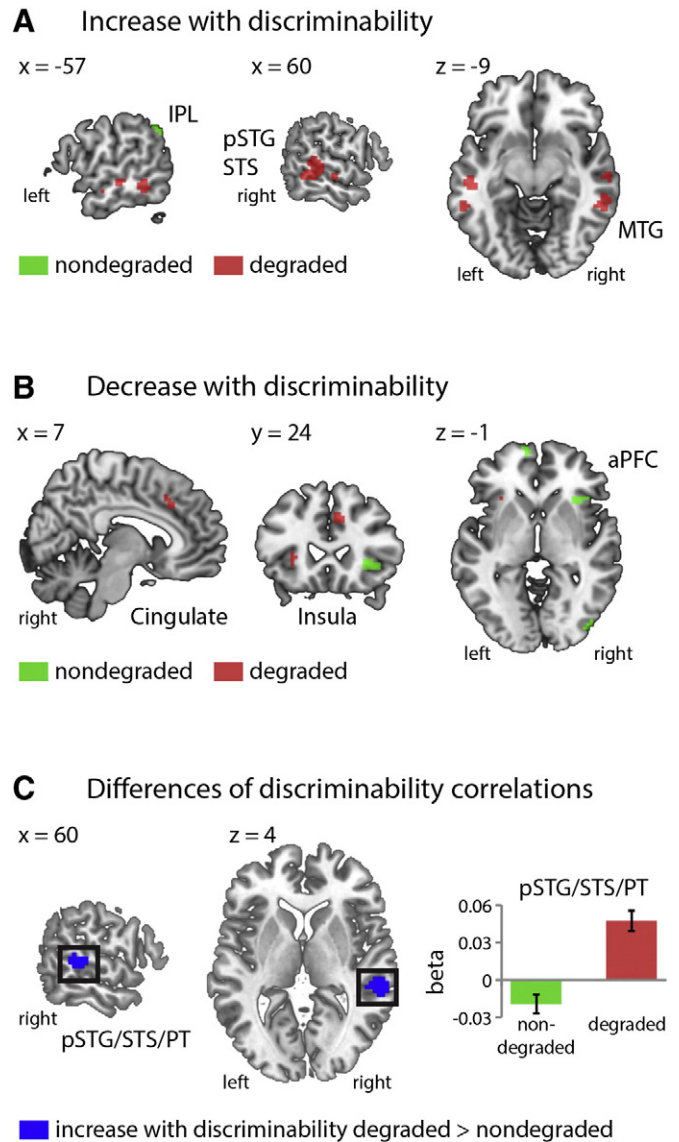


Fig. 3. Illustration of clusters showing sensitivity to discriminability (i.e., acoustic distance to median) under nondegraded (green) and degraded (red) conditions. A. Increasing activation with increasing discriminability. B. Decreasing activation with increasing discriminability. C. Effect of degradation on the sensitivity to discriminability (blue, $p < 0.005$, extent threshold $k > 22$). Abbreviations: IPL – inferior parietal lobule, aPFC – anterior prefrontal cortex, pSTG/STS – posterior superior temporal gyrus/sulcus, MTG – middle temporal gyrus, PT – planum temporale.

and across category discrimination along an acoustic continuum (from /ba/ to /ga/). They found that the pSTG was more sensitive to identical across-category than within-category contrasts, and least responsive to stimuli at ambiguous positions within the acoustic continuum.

Altogether, the present experiment provides evidence for the key role posterior temporal regions (posterior to primary auditory cortex) play in effortful, that is, acoustically challenging categorization situations. The challenge here arose from spectral degradation, which can occur in digital communication devices and poses the most drastic challenge for cochlear implant users. As argued above, spectral degradation essentially resulted in a compression of acoustic space, as expressed by the reduced distances of degraded stimuli to a maximally ambiguous median stimulus. Thus, posterior STG/STS and PT involvement is here somewhat more critical than in ideal listening situations with full spectral detail, where the acoustic space is more spread out.

Interestingly, the left IPL showed a complementary, if not opposite, pattern: the IPL was also sensitive to acoustic median distance, but

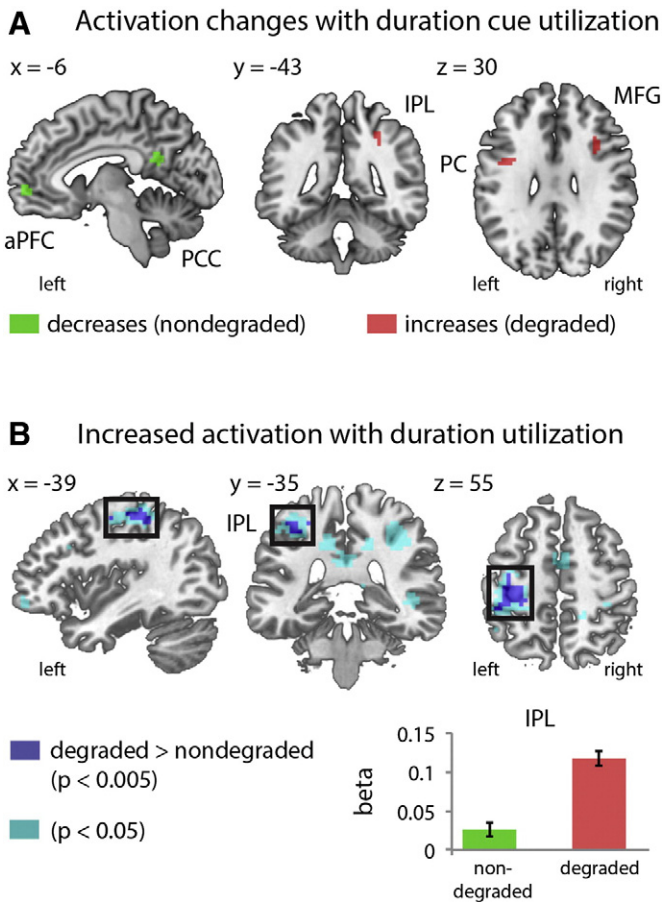


Fig. 4. A. Illustration of clusters resulting from the cue index correlations in the nondegraded (green) and degraded (red) conditions ($p < 0.005$, extent threshold $k > 22$). B. Condition differences of these correlations (blue/magenta). Abbreviations: aPFC – anterior prefrontal gyrus, PCC – posterior cingulate cortex, IPL – inferior parietal lobule, PC – precentral gyrus, MFG – middle frontal gyrus.

only in the nondegraded condition of our experiment (Fig. 3A). This activation might be best explained by a stronger weighting of duration cues for sounds further away from the ambiguous median location. Resulting from our stimulus design, sound durations at the periphery of the nondegraded distribution differed more from the median location than spectral peaks, and therefore contributed more to the acoustic median distance (Fig. 1A). Thus, even though spectral peak was the overall more informative cue in the nondegraded condition, for sounds with large median distances, duration might have been more informative, and participants needed to assign stronger weights to duration cues.

Effortful processing in the cingulo-opercular network

Lastly, less discriminable stimuli (with shorter distances to the most ambiguous median location) were associated with increased activity in the cingulo-opercular network (here, mid-cingulate cortex and bilateral anterior insula). Furthermore, BOLD activity under degradation showed additional clusters in the anterior cingulate cortex. There is by now a long list of studies highlighting the cingulo-opercular network's role in diverse scenarios of effortful cognitive processing and decision making (e.g., Eckert et al., 2009; Engström et al., 2013; Erb et al., 2012; Mulert et al., 2008; Vaden et al., 2013). Highlighting the link of the non-speech categorization challenges and speech processing, our results also corroborate the findings of Eckert et al. (2009), Erb et al. (2013), and Vaden et al. (2013). All these studies showed that the cingulo-opercular network also supports the comprehension of, and adaptation to degraded speech

and have underlined the role of the cingulo-opercular network in recognizing words in difficult listening conditions.

Conclusions

In this study, we were interested in the functional neural organization of two important aspects of auditory categorization: discriminating stimuli and utilizing most informative stimulus cues. The present fMRI data showed that, first, changes in acoustic cue utilization in response to acoustic degradation triggered activation increases in the IPL. Second, this was in contrast to activity in the posterior temporal cortex (including parts of the planum temporale), which scaled with stimulus discriminability under degradation. Taken together, these findings extend previous research on parietal versus posterior temporal cortex and provide support for their more general involvement in perceptual categorization.

Acknowledgments

This research has been funded by the Max Planck Society Germany (Max Planck Research Group grant "Auditory Cognition" to JO). We wish to express our thanks to Silvie Neubert and Dunja Kunka for their laboratory support and to Julia Erb and Björn Herrmann for their help and discussions regarding methodological issues.

References

- Alain, C., He, Y., Grady, C., 2007. The contribution of the inferior parietal lobe to auditory spatial working memory. *J. Cogn. Neurosci.* 20, 285–295.
- Ashburner, J., Friston, K.J., 2004. Computational neuroanatomy. In: Frackowiak, R.S., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C., Zeki, S. (Eds.), *Human Brain Function*. Academic Press, Amsterdam, pp. 655–672.
- Ashburner, J., Good, C.D., 2003. Spatial registration of images. In: Tofts, P. (Ed.), *Qualitative MRI of the Brain: Measuring Changes Caused by Disease*. John Wiley & Sons, Chichester, UK, pp. 503–531.
- Brett, M., Anton, J.-L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Brunetti, M., Della Penna, S., Ferretti, A., Del Gratta, C., Cianflone, F., Belardinelli, P., Caulo, M., Pizzella, V., Olivetti Belardinelli, M., Romani, G.L., 2008. A frontoparietal network for spatial attention reorienting in the auditory domain: a human fMRI/MEG study of functional and temporal dynamics. *Cereb. Cortex* 18, 1139–1147.
- Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432.
- Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297.
- Cusack, R., Carlyon, R.P., Robertson, I.H., 2000. Neglect between but not within auditory objects. *J. Cogn. Neurosci.* 12, 1056–1065.
- Cusack, R., Mitchell, D.J., Duncan, J., 2010. Discrete object representation, attention switching, and task difficulty in the parietal lobe. *J. Cogn. Neurosci.* 22, 32–47.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., Dehaene, S., 2005. Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33.
- Desai, R., Liebenthal, E., Waldron, E., Binder, J.R., 2008. Left posterior temporal regions are sensitive to auditory categorization. *J. Cogn. Neurosci.* 20, 1174–1188.
- Drullman, R., Festen, J.M., Plomp, R., 1994. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064.
- Eckert, M.A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., Dubno, J.R., 2009. At the heart of the ventral attention system: the right anterior insula. *Hum. Brain Mapp.* 30, 2530–2541.
- Engström, M., Landtblom, A.-M., Karlsson, T., 2013. Brain and effort: brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Front. Hum. Neurosci.* 7.
- Erb, J., Henry, M.J., Eisner, F., Obleser, J., 2012. Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia* 50, 2154–2164.
- Erb, J., Henry, M.J., Eisner, F., Obleser, J., 2013. The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *J. Neurosci.* 33, 10688–10697.
- Gaab, N., Gaser, C., Schlaug, G., 2006. Improvement-related functional plasticity following pitch memory training. *Neuroimage* 31, 255–263.
- Geng, J.J., Mangun, G.R., 2009. Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. *J. Cogn. Neurosci.* 21, 1584–1601.
- Gillebert, C.R., Dyrholm, M., Vangkilde, S., Kyllingsbæk, S., Peeters, R., Vandenberghe, R., 2012. Attentional priorities and access to short-term memory: parietal interactions. *Neuroimage* 62, 1551–1562.
- Glasberg, B.R., Moore, B.C., 1990. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138.

Table 3

Significant clusters from acoustic median distance and cue index correlations (thresholded at $p < 0.005$, $k > 22$) in the nondegraded and degraded conditions, together with comparisons between conditions. Peak activations are given in MNI coordinates. Abbreviations: IPL – inferior parietal lobule, MFG – middle frontal gyrus, IFG – inferior frontal gyrus, STG – superior temporal gyrus, STS – superior temporal sulcus, SFG – superior frontal gyrus, PT – planum temporale, TP – temporal pole, SMA – supplementary motor area, PCC – posterior cingulate cortex, Orb – orbito-frontal cortex, APFC – anterior prefrontal cortex.

Contrast	Region	MNI coordinates			Z-value	Size (voxels)
Increase with discriminability						
Nondegraded	l. IPL	–57	–55	37	3.36	29
Degraded	r. pSTG/STS/PT	60	–34	4	3.8	195
	l. pSTG/STS	–54	–25	–8	3.46	61
	l. MTG	–57	–46	–11	3.17	63
	r. TP	36	8	–17	3.02	26
Decrease with discriminability						
Nondegraded	l. APFC	–9	62	1	3.32	45
	r. Insula	33	23	1	3.29	31
Decrease with discriminability						
Degraded	r. Cingulate	9	23	34	3.48	34
	l. SMA	–3	17	49	3.2	28
	l. Insula	–30	20	7	3.13	43
Nondegraded > degraded					n.s.	
Degraded > nondegraded	r. pSTG/STS/PT	60	–34	4	3.61	52
	r. MTG	36	–4	–17	3.19	69
Cue Index						
Increase with spectral peak utilization	l. PCC	–6	–49	22	3.44	41
Nondegraded	l. APFC	–6	56	–2	3.37	27
Increase with duration utilization						
Degraded	r. MFG	33	14	31	3.84	32
	l. Precentral	–39	–1	28	3.62	27
	l. MFG	–30	38	–8	3.61	37
	r. IPL	27	–43	46	3.57	50
	r. Orb	45	50	–8	3.13	34
Nondegraded > degraded					n.s.	
Degraded > nondegraded	l. IPL	–39	–35	55	3.62	98

- Goudbeek, M., Swingle, D., Smits, R., 2009. Supervised and unsupervised learning of multi-dimensional acoustic categories. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1913–1933.
- Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., Zatorre, R.J., 2009. Voice perception in blind persons: a functional magnetic resonance imaging study. *Neuropsychologia* 47, 2967–2974.
- Grieffs, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353.
- Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S., Tourville, J.A., 2004. Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear. Res.* 47, 46–57.
- Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223.
- Henry, M.J., Herrmann, B., Obleser, J., 2013. Selective attention to temporal features on nested time scales. *Cereb. Cortex* 1–10 <http://dx.doi.org/10.1093/cercor/bht240>.
- Hill, K.T., Miller, L.M., 2010. Auditory attentional control and selection during cocktail party listening. *Cereb. Cortex* 20, 583–590.
- Holt, L.L., Lotto, A.J., 2006. Cue weighting in auditory categorization: implications for first and second language acquisition. *J. Acoust. Soc. Am.* 119, 3059–3071.
- Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., Horwitz, B., 2006. Neural bases of categorization of simple speech and nonspeech sounds. *Hum. Brain Mapp.* 27, 636–651.
- Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., Turner, R., 2002. Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage* 16, 217–240.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., Dupoux, E., 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 23, 9541–9546.
- Jäncke, L., Wüstenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. *Neuroimage* 15, 733–746.
- Jezzard, P., Balaban, R.S., 1995. Correction for geometric distortion in echo planar images from B0 field variations. *Magn. Reson. Med.* 34, 65–73.
- Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., Hoenig, K., 2008. The sound of concepts: four markers for a link between auditory and conceptual brain systems. *J. Neurosci.* 28, 12224–12230.
- Livesey, A.C., Wall, M.B., Smith, A.T., 2007. Time perception: manipulation of task difficulty dissociates clock functions from other cognitive demands. *Neuropsychologia* 45, 321–331.
- Macmillan, N.A., Creelman, C.D., 2005. *Detection Theory: A User's Guide*. Erlbaum, Mahwah, NJ.
- Mulert, C., Seifert, C., Leicht, G., Kirsch, V., Ertl, M., Karch, S., Moosmann, M., Lutz, J., Möller, H.-J., Hegerl, U., Pogarell, O., Jäger, L., 2008. Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *Neuroimage* 42, 158–168.
- Nosofsky, R., 1985. Overall similarity and the identification of separable-dimension stimuli: a choice model analysis. *Percept. Psychophys.* 38, 415–432.
- Obleser, J., Eisner, F., 2009. Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19.
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., Maess, B., 2012. Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383.
- Posner, M.I., Dehaene, S., 1994. Attentional networks. *Trends Neurosci.* 17, 75–79.
- Pugh, K.R., Offywitz, B.A., Shaywitz, S.E., Fulbright, R.K., Byrd, D., Skudlarski, P., Shankweiler, D.P., Katz, L., Constable, R.T., Fletcher, J., Lacadie, C., Marchione, K., Gore, J.C., 1996. Auditory selective attention: an fMRI investigation. *Neuroimage* 4, 159–173.
- Rinne, T., Stecker, G.C., Kang, X., Yund, E.W., Herron, T.J., Woods, D.L., 2007. Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *NeuroReport* 18, 1311–1314.
- Rosen, S., Faulkner, A., Wilkinson, L., 1999. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J. Acoust. Soc. Am.* 106, 3629–3636.
- Scharinger, M., Henry, M.J., Obleser, J., 2013. Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Mem. Cogn.* 41, 752–768.
- Scharinger, M., Herrmann, B., Nierhaus, T., Obleser, J., 2014. Simultaneous EEG-fMRI brain signatures of auditory cue utilization. *Front. Neurosci.* 8, 137.
- Schultz, J., Lennert, T., 2009. BOLD signal in intraparietal sulcus covaries with magnitude of implicitly driven attention shifts. *Neuroimage* 45, 1314–1328.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wyganski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Shaywitz, B.A., Shaywitz, S.E., Pugh, K.R., Fulbright, R.K., Skudlarski, P., Mencl, W.E., Constable, R.T., Marchione, K.E., Fletcher, J.M., Klorman, R., Lacadie, C., Gore, J.C., 2001. The functional neural architecture of components of attention in language-processing tasks. *Neuroimage* 13, 601–612.
- Slotnick, S.D., Moo, L.R., Segal, J.B., Hart Jr., J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res. Cogn. Brain Res.* 17, 75–82.
- Smits, R., Sereno, J., Jongman, A., 2006. Categorization of sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 733–754.
- Tuomainen, J., Savela, J., Obleser, J., Aaltonen, O., 2013. Attention modulates the use of spectral attributes in vowel discrimination: behavioral and event-related potential evidence. *Brain Res.* 1490, 170–183.
- Turkeltaub, P.E., Coslett, H.B., 2010. Localization of sublexical speech perception components. *Brain Lang.* 114, 1–15.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- Vaden, K.I., Kuchinsky, S.E., Cute, S.L., Ahlstrom, J.B., Dubno, J.R., Eckert, M.A., 2013. The cingulo-opercular network provides word-recognition benefit. *J. Neurosci.* 33, 18979–18986.
- Westbury, C.F., Zatorre, R.J., Evans, A.C., 1999. Quantifying variability in the planum temporale: a probability map. *Cereb. Cortex* 9, 392–405.
- Yantis, S., 1993. Stimulus-driven attentional capture and attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 676–681.
- Yantis, S., 2008. The neural basis of selective attention: cortical sources and targets of attentional modulation. *Curr. Dir. Psychol. Sci.* 17, 86–90.