

## FMRI repetition suppression for voices is modulated by stimulus expectations

Attila Andics<sup>a,b,\*</sup>, Viktor Gál<sup>a</sup>, Klára Vicsi<sup>c</sup>, Gábor Rudas<sup>a</sup>, Zoltán Vidnyánszky<sup>a,d,\*</sup>

<sup>a</sup> MR Research Center, Szentágotthai János Knowledge Center – Semmelweis University, Budapest, Balassa u. 6., 1083, Hungary

<sup>b</sup> Comparative Ethological Research Group, Hungarian Academy of Sciences – Eötvös Loránd University, Budapest, Pázmány Péter sétány 1/c, 1117, Hungary

<sup>c</sup> Laboratory of Speech Acoustics, Department of Telecommunications and Media-Informatics, Budapest University of Technology and Economics, Budapest,

Magyar tudósok körútja 2., 1117, Hungary

<sup>d</sup> Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Egrý József u. 1., 1111, Hungary

### ARTICLE INFO

#### Article history:

Accepted 17 December 2012

Available online 23 December 2012

#### Keywords:

fMRI repetition suppression

Adaptation

Predictive coding

Amygdala

STG

STS

### ABSTRACT

According to predictive coding models of sensory processing, stimulus expectations have a profound effect on sensory cortical responses. This was supported by experimental results, showing that fMRI repetition suppression (fMRI RS) for face stimuli is strongly modulated by the probability of stimulus repetitions throughout the visual cortical processing hierarchy. To test whether processing of voices is also affected by stimulus expectations, here we investigated the effect of repetition probability on fMRI RS in voice-selective cortical areas. Changing ('alt') and identical ('rep') voice stimulus pairs were presented to the listeners in blocks, with a varying probability of alt and rep trials across blocks. We found auditory fMRI RS in the nonprimary voice-selective cortical regions, including the bilateral posterior STS, the right anterior STG and the right IFC, as well as in the IPL. Importantly, fMRI RS effects in all of these areas were strongly modulated by the probability of stimulus repetition: auditory fMRI RS was reduced or not present in blocks with low repetition probability. Our results revealed that auditory fMRI RS in higher-level voice-selective cortical regions is modulated by repetition probabilities and thus suggest that in audition, similarly to the visual modality, processing of sensory information is shaped by stimulus expectation processes.

© 2013 Elsevier Inc. All rights reserved.

### Introduction

fMRI repetition suppression is one of the most well-known phenomena in cognitive neuroscience (Aguirre, 2007; Epstein et al., 2008; Grill-Spector et al., 2006). It refers to the observation that the immediate repetition of a stimulus leads to a reduced neural response compared to its first appearance. fMRI RS is also robust, it has been found in both visual and auditory modalities, across various brain regions. Visual fMRI RS for face stimuli is present at all stages of the face processing hierarchy (Henson et al., 2000, 2002). Auditory fMRI RS for voice stimuli was found in the nonprimary auditory cortex (Altmann et al., 2007b; Andics et al., 2010; Belin and Zatorre, 2003; Bergerbest et al., 2004; Chandrasekaran et al., 2011; Latinus et al., 2011; Leaver and Rauschecker, 2010; Perrodin et al., 2011; Petkov et al., 2008), and also extratemporally (IFC, Andics et al., 2010; Latinus et al., 2011; IPL, Celsis et al., 1999). fMRI RS is widely explored and used to functionally characterize cortical areas, however, the underlying neural processes are still unclear. Indeed, recent findings (Kovács et al., 2012a; Larsson and Smith, 2012; Summerfield et al., 2008) seriously challenge the classical view that fMRI RS is caused exclusively

by largely automatic, bottom-up processes such as neural adaptation (Grill-Spector et al., 2006), as feature detection models supposed (Hubel and Wiesel, 1965; Riesenhuber and Poggio, 2000).

Visual fMRI RS for face stimuli was shown to be highly dependent on stimulus expectations (Kovács et al., 2012a; Larsson and Smith, 2012; Summerfield et al., 2008). According to Bayesian views of perception, stimulus expectations can be derived from the statistical regularities of the surrounding environment. Stimulus repetition frequency modulation is a simple way to manipulate expectations without explicit instructions on what comes next. Summerfield et al. (2008) found that during face processing, stimulus repetition probability modulates visual fMRI RS in the human brain: when repetitions were frequent (and therefore expected), visual fMRI RS was larger than when repetitions were rare (and therefore unexpected). Visual fMRI RS may thus also be a consequence of top-down perceptual expectations, as predictive coding models (Friston, 2005; Henson, 2003) and stimulus-driven attentional models (Corbetta and Shulman, 2002; Larsson and Smith, 2012) of visual cognition proposed.

Importantly, whether stimulus expectations, and, specifically, repetition probability also has an effect on auditory fMRI RS, similarly to that found for visual fMRI RS, has not been investigated before. One recent MEG experiment, Todorovic et al. (2011) demonstrated probability-based effects in the auditory cortex with tone stimuli. More specifically, they found that repetition probability modulates both early evoked responses and later gamma coherence, suggesting

\* Corresponding authors at: Szentágotthai Knowledge Center, Semmelweis University MR Research Center, H-1083 Budapest, Balassa u. 6, Hungary. Fax: +36 1 459 1580.

E-mail addresses: [attila.andics@gmail.com](mailto:attila.andics@gmail.com) (A. Andics), [vidnyanszky@digitus.itk.ppke.hu](mailto:vidnyanszky@digitus.itk.ppke.hu) (Z. Vidnyánszky).

that modulatory effects of expectation may influence multiple levels of auditory information processing. These results suggest that to better understand the presence or absence of repetition effects to voice stimuli in cortical regions of the auditory, and in specific voice processing hierarchy, it is important to investigate whether and how auditory fMRI RS is influenced by a systematic manipulation of stimulus expectations.

Here we investigated repetition probability effects on auditory fMRI RS for voices. We presented voice pairs that consisted of either two identical stimuli (repetitions) or two different stimuli (alternations). Stimulus expectations were modulated by manipulating repetition probability between blocks: in repetition blocks, a high proportion of voice pairs were repetitions, but in alternation blocks, a high proportion of voice pairs were alternations.

In short, we found auditory fMRI RS in an extended network of cortical areas, involving multiple voice-selective temporal regions, the inferior frontal cortex and the inferior parietal lobule; but no primary auditory regions. We also showed that fMRI RS to voice stimuli, where present, is modulated by repetition probability. Finally, our results showed that both repetition effects and repetition probability effects differ across cortical regions and indicated the possible involvement of both predictive coding and attentional mechanisms.

## Material and methods

### Participants

18 Hungarian listeners (9 females, 9 males, 20–39 years) with no reported hearing disorders were paid to complete the experiment. Written informed consent was obtained from all participants.

### Stimuli

60 Hungarian speakers (30 females, 30 males), saying the Hungarian word “négy” [ne:j] (four) were selected from the Babel–Hungarian Speech Database (Roach et al., 1996). Stimuli were digitized at a 16 bit/20 kHz sampling rate and were equalized for maximum amplitude using CoolEdit software (Version 2.0, 2010). A reduced-volume variant of all stimuli was also included: average volume was reduced to approximately 60% of the original volume. Average syllable duration was 459 ms.

### Design and procedure

Test trials consisted of a pair of words with an onset delay of 850 ms. The second stimulus in the pair was either a repetition of the first one (rep trials) or was spoken by a different person from the same gender (alt trials). Participants were instructed to perform a volume change detection task: in 20% of trials, the volume of the second stimulus in the pair was reduced (target trials). Participants had to signal the detection of the volume change with a button press. In half of the target trials the two stimuli were from the same talker, in the other half the two stimuli were from different talkers. Trials were grouped by talker–gender and these grouped trials (sub-blocks) were presented interleaved: 30 male/female talker trials were followed by 30 female/male talker trials and so on. Stimulus content was identical across sub-blocks: each sub-block contained each voice stimulus from a given gender exactly twice, either in a single stimulus pair, or as one of the stimuli in two different trials. In the latter case, distance between identical stimuli was greater than three trials. Each block consisted of five 30-trial sub-blocks, i.e. 150 trials in total. The main part of the experiment consisted of four 150-trial blocks, with gender order counterbalanced across sub-blocks. In two of the blocks, the proportion of rep trials was 75% (REP blocks); while in the other two blocks the proportion of alt trials was 75% (ALT blocks). REP and ALT blocks were presented in separate functional runs, interleaved for all participants, with block

order counterbalanced across participants. Stimulus ordering was pseudorandom, using a random search algorithm to maximize predictor distinctiveness based on a-optimality criterion (Birn et al., 2002).

Voice selective regions were defined in a separate localizer run (Belin et al., 2000) with 40 s long trials corresponding to (1) vocal sounds (verbal and nonverbal), (2) non-vocal sounds (animals, sounds from the environment, music) matched for number of sources, in duration, and overall energy and (3) silence. Participants were instructed to passively listen to the stimuli. Stimuli were presented at a standard, comfortable volume (but note volume modulation for target trials). Stimuli were controlled using Matlab (version 7.9.0.529) Psychophysics Toolbox 3 (Kleiner et al., 2007). During imaging, stimulus presentation was synchronized by a TTL trigger pulse with the data acquisition. Stimuli were delivered binaurally through MRI-compatible headphones (MR Confon, Magdeburg, Germany).

### Data acquisition

MRI measurements were performed on a Philips Achieva 3 T whole body MR unit (Philips Medical Systems, Best, The Netherlands) equipped with an eight-channel Philips SENSE head coil. For the main tests EPI-BOLD fMRI time series were obtained from 29 transverse slices covering temporal lobes and the inferior part of the frontal lobes with a spatial resolution of  $3.5 \times 3.5 \times 3.5$  mm, including a 0.5 mm slice gap, using a single-shot gradient-echo planar sequence (ascending slice order; acquisition matrix  $64 \times 64$ ; TR = 3600 ms, including 1893 ms acquisition and 1707 ms silent gap; TE = 32.3 ms; and flip angle =  $90^\circ$ ). That is, the acquisition of each volume was followed by a short 1707 ms gap where the scanner was silent. Compared to standard sparse sampling methods, this close-to-continuous sampling method not only increased statistical power by increasing the number of data points, but also made it possible to haemodynamically model each individual stimulus. At the same time it was possible to present all auditory stimuli in silence. The onset of the first stimulus in the pair was at 2050 ms from trial onset, i.e. 157 ms after silent gap onset. The offset of the second stimulus was followed by an on-average 241 ms silent gap before the next volume acquisition. Each of the four runs included 160 volumes.

For the voice localizer there was a longer silent gap between acquisitions (TR = 10000 ms, including 2000 ms acquisition and 8000 ms silent gap; TE = 36 ms). The voice localizer run included 63 volumes. All other parameters were identical to the main test settings.

In addition to the functional time series, a standard T1-weighted three-dimensional scan using a turbo-field echo (TFE) sequence with 180 slices covering the whole brain was collected for anatomical reference at the end of the second scanning session, with  $1 \times 1 \times 1$  mm spatial resolution.

### Data analysis

Image preprocessing and statistical analysis were performed using SPM8 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). The functional EPI-BOLD images were realigned, co-registered to the subject's T1-weighted MR images, spatially normalized, and transformed into a common anatomical space, as defined by the SPM Montreal Neurological Institute (MNI) T1 template. Next, the functional EPI-BOLD images were spatially filtered by convolving the functional images with an isotropic 3-D Gaussian kernel (10 mm FWHM). The fMRI data were then statistically analyzed using a general linear model and statistical parametric mapping (Friston et al., 2007). For the main analyses, every token pair was modeled as a separate event. Condition regressors for each run of the main analysis were constructed for three trial types: alt, rep and target trials. Repetition suppression and its modulation by stimulus repetition probability was measured in a test contrasting alt and rep trials in ALT and REP blocks. The regressor for target trials

was not included in the reported contrasts. Across-block effects on across-condition contrasts may be quantified as condition by blocks interactions in a repeated-measures ANOVA with the factors block (ALT, REP) and condition (alt, rep). We report this ANOVA, but, as direct comparisons of conditions from different functional runs may be problematic, as the activity level of the participant and the noise level of the magnet might change across runs, we consider this as an additional analysis only, and base our conclusions on the direct, block-specific tests instead. Conditions in the voice localizer run were modeled as 8 s long blocks.

Realignment regressors for each run were also included to model potential movement artifacts. A high-pass filter with a cycle-cutoff of 128 s was implemented in the design to remove low-frequency signals. Single-subject fixed effect analyses were followed by whole-volume random effects analyses on the group level. Clusters of voxels from the main test (alt vs rep trials) surviving a height threshold  $p < 0.001$  ( $T(17) > 3.646$ , uncorrected) and an extent threshold  $p < 0.05$  (69 voxels for alt > rep (REP), 57 voxels for rep > alt (ALT), uncorrected) are reported. Clusters surviving a cluster-level family-wise error (FWE) correction ( $p < 0.05$ ) were considered significant (Table 1).

Clusters of voxels from the functional localizer's vocal vs non-vocal contrast surviving a rigorous height threshold  $p < 0.00005$  ( $T(17) > 5.044$ , uncorrected), an extent threshold  $p < 0.05$  (25 voxels, uncorrected) and a cluster-level FWE-correction ( $p < 0.05$ ) were used to define group ROIs (Table 2). We applied this more conservative threshold for the voice localizer test to localize sub-units and distant local maxima of the extensive bilateral STS activity. Additionally, a single cluster was defined per hemisphere within the anatomically specified primary auditory cortex (PAC; Heschl's gyrus, BA 41, WFU Pickatlas, Version3.0), by the functional localizer's vocal vs silence contrast (same thresholds). Non-overlapping group ROIs were constructed around each cluster peak from the functional localizer as a sphere with a radius of 8 mm. We used group ROIs, because individual local maxima could not be reliably determined for these ROIs. An acknowledged disadvantage of using group ROIs instead of individual ROIs is the inability to accommodate subtle anatomical variations across individuals (e.g., Swallow et al., 2003). Nevertheless, to increase statistical power, functional localizer-based group ROIs are often used to define regions for individual data extraction (e.g., Egner et al., 2010).

**Results**

*Behavioural results*

Response accuracies during the fMRI experiment were at ceiling (mean = 97%, SD = 3%). Response times (RT) recorded for target trials

**Table 1**  
List of brain regions showing repetition effects in high repetition probability (REP) and high alternation probability (ALT) blocks.

Region	T(17)	p(corr)	x	y	z	Cluster size
<b>alt &gt; rep (REP blocks)</b>						
R posterior MTG	5.47	<b>&lt;0.001</b>	54	-44	-22	465
R IFC	5.31	<b>0.001</b>	48	20	8	361
R IPL	5.33	<b>0.035</b>	38	-60	46	177
L posterior MTG	4.43	0.075	-62	-44	-6	141
L IPL	5.02	0.172	-40	-52	46	104
rep > alt (REP blocks)	No suprathreshold activity					
alt > rep (ALT blocks)	No suprathreshold activity					
<b>rep &gt; alt (ALT blocks)</b>						
L cingulate gyrus	6.98	<b>0.001</b>	-12	-18	30	324
R cingulate gyrus	5.53	0.265	12	-24	28	76
Thalamus	5.17	0.088	8	-14	-4	117
Anterior cingulate cortex	4.42	0.225	0	54	-4	82

Clusters surviving a height threshold  $p < 0.001$  (uncorrected) and an extent threshold  $p < 0.05$  (uncorrected) are listed. Significant probability values [cluster-level family-wise error (FWE) correction ( $p < 0.05$ )] are highlighted in bold.

**Table 2**  
Regions of interest as defined by the voice localizer test.

Region	T(17)	p(corr)	x	y	z	Cluster size
<i>Vocal &gt; silence (within BA 41)</i>						
R PAC	10.33	<b>&lt;0.001</b>	42	-34	6	186
L PAC	7.36	<b>&lt;0.001</b>	-42	-34	8	95
<i>Vocal &gt; non-vocal (whole volume)</i>						
R pSTS	10.56	<b>&lt;0.001</b>	54	-30	-10	1260
R aSTG	6.65	0.037	60	4	6	32
R TP	6.1	0.007	42	6	-22	54
R amygdala	6.4	0.028	28	-4	-14	37
L PAC	7.37	<b>&lt;0.001</b>	-54	-28	12	284
L pSTS	5.37	0.023	-58	-48	4	31
L aSTG	8.73	<b>&lt;0.001</b>	-56	-6	2	178
L TP	7.62	<b>&lt;0.001</b>	-42	-2	-24	275

Clusters surviving a height threshold  $p < 0.00005$  (uncorrected) and an extent threshold  $p < 0.05$  (FWE-corrected) are listed.

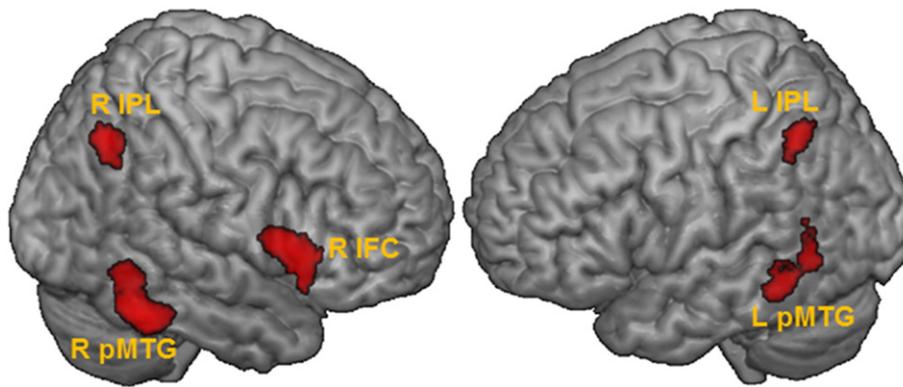
were tested for condition- and block-related effects (mean = 579 ms, SD = 150 ms). No significant differences were found in RTs either across blocks, or across conditions, or across conditions within either block type (all  $F_s < 1$ ).

*Whole-volume analyses*

In REP blocks, we found fMRI RS effects (alt > rep) in the right posterior middle temporal gyrus (MTG), extending both superiorly to the STS (superior temporal sulcus) and inferiorly, the right inferior parietal lobule (IPL), including the angular and supramarginal gyri, and a region in the right inferior frontal cortex (IFC). Left hemisphere counterparts of the posterior MTG and IPL clusters were also identified, but these regions did not survive cluster-level correction. In REP blocks, no region showed the reverse (i.e. rep > alt) effect. In ALT blocks, in contrast, no region showed a repetition suppression effect. The reverse effect, that is, a decrease in neural response for alt compared to rep trials (repetition enhancement) was identified in ALT blocks in anterior and middle regions of the cingulate cortex and in the thalamus, but only a single region, the left cingulate gyrus survived cluster-level correction (Table 1, Fig. 1). A repeated-measures ANOVA on response estimates, within spheric ROIs around each cluster peak, with the factors condition (alt vs rep) and block (ALT vs REP) confirmed a condition by block interaction for all four regions that survived cluster-level correction (R MTG/R IPL/R IFC/L cingulate gyrus;  $F(1,17) = 4.899/7.204/22.674/20.572$ ;  $p = 0.041/0.016/<0.001/<0.001$ ).

*ROI analyses*

Next we tested how voice-selective brain regions responded to voice repetitions (rep trials) and voice alternations (alt trials) in blocks with high probability of repetitions (REP blocks) and alternations (ALT blocks), respectively. Voice-selective regions were defined by an independent functional localizer comparing vocal to non-vocal sounds (Andics et al., 2010; Belin et al., 2000). Temporal and paralimbic regions were identified and used to construct ROIs (see Methods and Table 2, Fig. 2), including the left primary auditory cortex (L PAC), the bilateral posterior superior temporal sulcus (R/L pSTS), the bilateral anterior superior temporal gyrus (R/L aSTG), the bilateral temporal pole (R/L TP) and the right amygdala. Additionally, bilateral primary auditory cortex regions (R/L PAC) were defined within the anatomically specified BA 41 by the functional localizer's vocal versus silence contrast. The anatomically and functionally (i.e. vocal > silence) informed left PAC cluster showed an extensive overlap with the voice-selective (i.e. vocal > non-vocal) L PAC cluster, and the former one was therefore not considered for the ROI analyses. All other clusters were non-overlapping, with any two cluster peaks

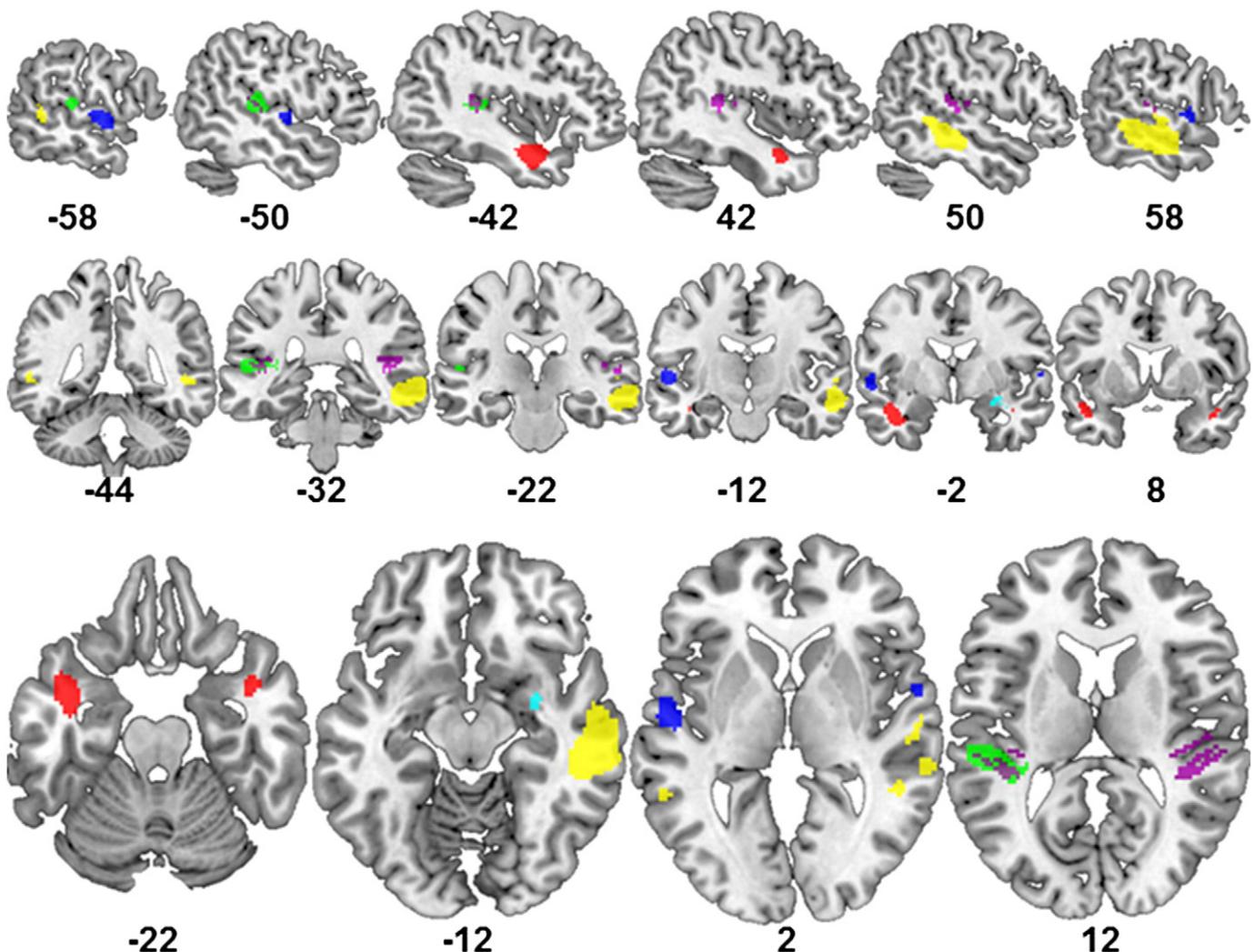


**Fig. 1.** Cortical regions showing RS for voices, displayed on a volume rendered brain. Alt versus rep contrast in high repetition probability (REP) blocks, height threshold  $p < 0.001$  (uncorrected), extent threshold  $p < 0.05$  (uncorrected). R/L pMTG: right/left posterior middle temporal gyrus; R IFC: right inferior frontal cortex; R/L IPL: right/left inferior parietal lobule.

more than 16 mm apart (minimal distance: 19 mm between R amygdala and R TP, Table 2).

Response estimates were then assessed for each ROI, in a within-participant  $2 \times 2$  factorial design, with the factors condition (alt vs rep) and block (ALT vs REP). Main effects of condition and

block, a condition by block interaction, and follow-up direct comparisons are shown in Table 3. No main effect of block was found. A main effect of condition (alt > rep) was found in the bilateral posterior STS (R pSTS/L pSTS;  $F(1,17) = 8.756/7.367$ ,  $p = 0.009/0.015$ ). A condition by block interaction was found in the right anterior STG ( $F(1,17) = 4.706$ ,



**Fig. 2.** Cortical regions revealed by the voice localizer test, displayed on sagittal, coronal and axial brain slices. Vocal versus non-vocal contrast, height threshold  $p < 0.00005$ , extent threshold  $p < 0.05$  (FWE-corrected); R/L pSTS (yellow); R/L aSTG (blue); R/L TP (red); R amygdala (cyan); L PAC (green). Vocal versus silence contrast within the anatomically defined Brodmann area 41 (PAC, same threshold; violet).

**Table 3**  
Effects of condition (alt, rep) and block (ALT, REP) in a repeated-measures ANOVA and in specific t-tests.

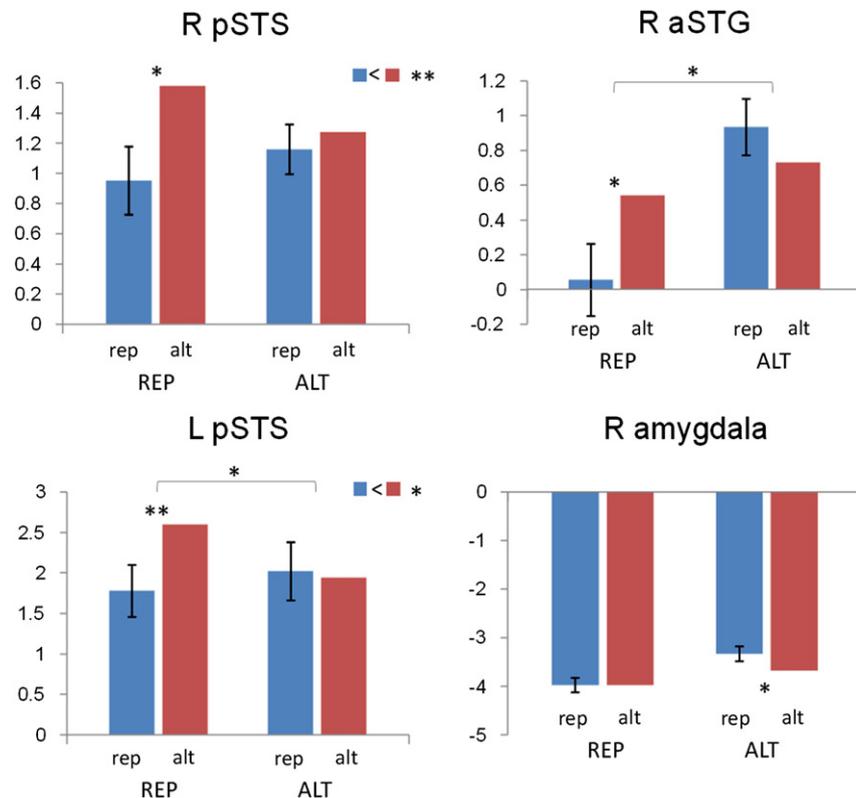
Region of interest (peak coordinate)	Condition F(1,17)	Block F(1,17)	Block x condition F(1,17)	REP: alt>rep T(17)	ALT: alt>rep T(17)	ALTrep>REPrep T(17)	ALTalt>REPalt T(17)
R PAC	0.029	0.204	<0.001	0.074	-0.122	0.407	0.438
R STS	<b>8.756**</b>	0.007	2.845	<b>2.788*</b>	0.692	0.357	-0.455
R aSTG	2.120	1.787	<b>4.706*</b>	<b>2.345*</b>	-1.260	<b>2.214*</b>	0.408
R TP	0.264	1.326	1.293	1.087	-1.006	1.448	0.776
R amygdala	3.800	0.816	1.949	0.016	<b>-2.233*</b>	1.254	0.537
L PAC	1.026	1.331	1.999	0.805	-1.639	1.593	0.587
L pSTS	<b>7.367*</b>	0.751	<b>6.056*</b>	<b>3.299**</b>	-0.389	0.388	-0.925
L aSTG	0.511	3.691	0.899	0.489	-1.098	<b>2.133*</b>	1.439
L TP	0.574	0.494	1.318	0.679	-1.030	0.916	0.473

\* p<0.05.  
\*\* p<0.01.

p = 0.045) and the left posterior STS (F(1,17) = 6.056, p = 0.025). In REP blocks, repetition suppression effects were found in the bilateral posterior STS (R pSTS/L pSTS; T(17) = 2.788/3.299, p = 0.013/0.004) and in the right anterior STG (T(17) = 2.345, p = 0.031), with rep trials eliciting a decrease in neural response compared to alt trials. In contrast, no significant repetition suppression was found in ALT blocks in any of the ROIs. The reverse effect (rep > alt) was found for no clusters in the REP blocks, and for a single cluster, the right amygdala in the ALT blocks (T(17) = 2.233, p = 0.039). A further direct comparison showed that in the bilateral anterior STG, rep trials elicited less activity in REP than in ALT blocks (R aSTG/L aSTG; T(17) = 2.214/2.133; p = 0.041/0.048). No region showed more activity in REP than in ALT blocks for rep trials. For alt trials, no between-block significant difference was found in any region, in either direction. These data clearly show that repetition effects in voice-sensitive regions in the bilateral posterior STS, the right anterior STG, and the amygdala were modulated by repetition probability (see Fig. 3). Note that compared to the baseline, parameter estimates of amygdala activity are positive in the voice localizer for vocal stimuli,

and are around zero for the non-vocal stimuli, but they are negative for the voice stimuli of the main test. Although there exist various explanations for negative BOLD responses, it has recently been demonstrated that a large amount of the negative response originates in neural activity decreases (Shmuel et al., 2006), suggesting that a more negative BOLD response can be interpreted analogously to a less positive BOLD response. The decreased activity level in the amygdala for main test stimuli is not surprising, since localizer stimuli were emotionally loaded, and main test stimuli were neutral, and the amygdala is known to be sensitive to the emotional intensity of voice stimuli (Fecteau et al., 2007).

The functional localizer also confirmed the voice-sensitivity of some regions found in the whole-volume analyses, namely the posterior MTG and the right IFC. The posterior MTG regions from the whole-volume analysis overlapped with the posterior STS regions as identified by the functional localizer. Right IFC's voice-sensitivity was also confirmed by the present data (IFC cluster's peak coordinate was [48, 28, -8], T = 4.23, p < 0.0005 uncorrected), but it failed to survive the rigorous height



**Fig. 3.** Auditory repetition effects and repetition probability effects. Bars indicate average parameter estimates for selected ROIs, for rep (blue) and alt (red) trials in REP and ALT blocks. Error bars indicate within-subject s.e.m., derived from between-condition differences, per block. Asterisks indicate within-block and overall condition effects, and condition by block interactions where significant, \* p < 0.05, \*\* p < 0.01.

threshold  $p < 0.00005$  that we applied in order to localize sub-units of the extensive STS activity (see [Methods](#)).

## Discussion

### *Repetition effects*

In the present auditory fMRI study, we investigated stimulus expectation effects on the fMRI RS in case of voice stimuli. We presented voice pairs that consisted of either two identical stimuli (repetitions) or two different voice identity stimuli (alternations). We modulated voice stimulus expectations by manipulating between blocks the repetition probability within a voice pair. Auditory fMRI RS was found in higher-level voice-selective cortical regions, but not in the primary auditory cortex. Furthermore, our results also revealed that auditory fMRI RS is strongly modulated by repetition probability and that repetition probability effects differ in the different voice-selective cortical regions.

In line with recent reports of short-term RS for voice stimuli, RS was found in higher-level voice-selective cortical regions, including the bilateral posterior STS ([Andics et al., 2010](#); [Latinus et al., 2011](#); [Leaver and Rauschecker, 2010](#)), the right anterior STG ([Belin and Zatorre, 2003](#); [Latinus et al., 2011](#)) and the right IFC ([Andics et al., 2010](#); [Latinus et al., 2011](#)). But not all stages of the voice processing stream showed fMRI RS. First, we found no repetition effects in the primary auditory cortex. This is consistent with earlier fMRI studies ([Altmann et al., 2007b](#); [Andics et al., 2010](#); [Belin and Zatorre, 2003](#); [Bergerbest et al., 2004](#); [Chandrasekaran et al., 2011](#); [Latinus et al., 2011](#)) but see [De Lucia et al. \(2010\)](#) and might indicate that fMRI RS in low-level auditory areas might be less pronounced than in the more downstream, higher-level areas. A similar explanation has been challenged in vision, however, where a lack of low-level visual fMRI RS was also reported initially ([Boynton and Finney, 2003](#)), but later studies have demonstrated fMRI RS in V1 using lower-contrast test stimuli ([Larsson et al., 2006](#)). Another explanation for why fMRI studies fail to replicate electrophysiological RS findings in early auditory cortex ([Altmann et al., 2007a](#); [Haenschel et al., 2000](#); [Todorovic et al., 2011](#); [Valentini et al., 2011](#)) can be that anticipation-related vascular responses might not directly reflect local neural responses, as demonstrated in the primate brain ([Sirotnin and Das, 2009](#)).

Furthermore, no fMRI RS was found in the voice-selective bilateral temporal poles and the amygdala. The lack of RS for vocal stimuli in these voice-selective regions may indicate that neither the temporal poles nor the amygdala are core voice regions in the sense that their activity is not based upon the sensitivity to short-term voice changes. This notion is in line with a recent proposal that there is a paralimbic network including both the amygdala and the temporal pole which is specialized for person identification, across different modalities ([Olson et al., 2007](#)). The amygdala is also suggested to be tuned to emotional stimuli more than to neutral stimuli, and to faces more than to voices, although recent results indicate that it nevertheless participates in the representation of person identity given neutral voice stimuli ([Andics et al., 2010](#)).

On the other hand, fMRI RS was not restricted to voice-selective cortical regions. fMRI RS was also found in the bilateral inferior parietal lobule (IPL). This appears to be in agreement with previous findings suggesting that the bilateral IPL (including the supramarginal gyrus) might play an important role in phonological processing ([Hartwigsen et al., 2010](#)), as well as with an earlier study showing repetition effects in the left supramarginal gyrus for syllable and tone stimuli ([Celsis et al., 1999](#)).

### *Repetition probability effects*

Crucially, auditory fMRI RS, where present, was strongly modulated by repetition probabilities in this study. The generality of expectation-based modulation of auditory fMRI RS is similar to what

was found in the visual modality. Indeed, all stages of the face processing hierarchy where visual fMRI RS was found were also consistently found to be influenced by repetition probabilities ([Kovács et al., 2012a](#); [Larsson and Smith, 2012](#); [Summerfield et al., 2008](#)). However, a recent electrophysiological investigation with primates failed to find probability-based modulation on visual repetition effects ([Kaliukhovich and Vogels, 2011](#)). [Kovács et al. \(2012b\)](#) also found no repetition probability effects on visual fMRI RS when stimuli were objects and not faces. These negative results raised questions about the generality of probabilistic coding effects. Our data suggest that probability effects on fMRI RS are present across modalities and across processing stages.

Importantly, our repetition probability manipulation differed from that used by previous studies ([Kovács et al., 2012a](#); [Larsson and Smith, 2012](#); [Summerfield et al., 2008](#)) in an important aspect. In these studies, all stimuli were trial-unique. [Summerfield et al. \(2008\)](#) argued that this was done to ensure that they manipulate stimulus repetition probability *per se*, and not stimulus repetition frequency. However, trial-uniqueness also had the consequence that stimuli of rep trials were presented twice, while stimuli of alt trials were presented only once throughout the experiment. That is, REP blocks and ALT blocks not only consisted of visually different sets of stimuli, but also REP blocks contained much less stimuli in total than ALT blocks, suggesting that visual variation within REP blocks was lower than that within ALT blocks. [Summerfield et al. \(2008\)](#) argued that the greater repetition suppression in REP blocks was caused by a top-down effect, namely an increase in stimulus predictability. But the repetition suppression in REP blocks could also be caused by a bottom-up effect, namely a reduction in stimulus variation. In contrast, the present study applied a design where every voice stimulus, including those in rep and those in alt trials was presented twice in every block: either within the same trial, or in two different, distant trials. Therefore, stimulus content was identical across blocks. Our experimental manipulation of stimulus predictability was thus not confounded by acoustic differences between blocks. This ascertained that any modulation of repetition effects between blocks is caused by top-down, expectation-driven effects rather than by bottom-up, stimulus-driven effects.

Repetition probability effects differed across different cortical regions, indicating that different neural mechanisms may underlie expectation-based modulations. A reduction of activity was found for rep compared to alt trials in REP but not in ALT blocks in the bilateral posterior STS, the right anterior STG, the right IFC and the bilateral IPL. In the anterior STG (but not in the posterior STS regions), rep trials elicited less activity in REP than in ALT blocks, while no difference was found for alt trials across blocks. This pattern of results is fully consistent with a predictive coding account of repetition probability effects. According to predictive coding models, repetition effects are modulated by probability-based expectations through the following mechanism: a mispredicted or unpredicted stimulus, in contrast with a predicted stimulus, is associated with a surprise signal of prediction error processing units and thus with an increased neural response ([Arnal and Giraud, 2012](#); [Baldeweg, 2006](#); [Egner et al., 2010](#); [Friston, 2005](#); [Henson, 2003](#); [Summerfield et al., 2008](#)).

A clearly different pattern of repetition probability effects was found for the amygdala, the thalamus and the cingulate cortex. In these regions, reduced neural responses were found for expected alternations compared to unexpected repetitions (i.e. alt versus rep trials in ALT blocks). Stimuli in alt and rep trials in ALT blocks are equally unpredicted, since the listener does not expect repetition and cannot predict the identity of the alternating second stimulus in the voice pair trial. But alt trials are still less surprising than rep trials in ALT blocks. While the predictive coding model predicts no difference in neural response between expected alternations and unexpected repetitions, the attentional model ([Larsson and Smith, 2012](#)) makes clear predictions for this contrast: an increased neural response is

predicted for unexpected compared to expected stimulus relations (i.e. rep versus alt trials in ALT blocks). These findings for the amygdala, the thalamus and the cingulate cortex are thus compatible with the attentional account of expectation-based repetition effects, but not with the predictive coding account. Nevertheless, further work is needed to distinguish between the two accounts.

To summarize, prior expectations on stimulus repetitions were shown to modulate auditory fMRI RS for voices. This pattern of results clearly supports expectation-based models of voice processing over feature-detection models.

## Conclusions

In conclusion, here we demonstrated for the first time that similarly to the visual modality, fMRI RS can also be used to reveal stimulus expectation effects in the auditory modality. Auditory fMRI RS effects for voice stimuli were found in higher-level voice-selective cortical regions, but not in the primary auditory cortex. We presented evidence that auditory fMRI RS was strongly modulated by the probability of stimulus repetitions and argued that the repetition probability effects are caused by top-down stimulus expectations. Our findings indicated that both effects caused by predictive coding mechanisms (i.e., increased prediction error for unpredicted and mispredicted stimuli) and stimulus-driven attentional effects (i.e., increased attention for unexpected stimulus relations) may contribute to the expectation-based modulations of auditory repetition effects, but the assessment of the relative contribution of these factors will have to be the subject of future studies.

## Acknowledgments

This work was supported by the Hungarian Scientific Research Fund (OTKA CNK 80369 to Z.V.), the National Development Agency of Hungary (TÁMOP-4.2.1.B-09/1/KMR-2010-0001) and the Hungarian Academy of Sciences (MTA 01 031).

## References

- Aguirre, G.K., 2007. Continuous carry-over designs for fMRI. *Neuroimage* 35, 1480–1494.
- Altmann, C.F., Bledowski, C., Wibral, M., Kaiser, J., 2007a. Processing of location and pattern changes of natural sounds in the human auditory cortex. *Neuroimage* 35, 1192–1200.
- Altmann, C.F., Doehrmann, O., Kaiser, J., 2007b. Selectivity for animal vocalizations in the human auditory cortex. *Cereb. Cortex* 17, 2601–2608.
- Andics, A., McQueen, J.M., Petersson, K.M., Gál, V., Rudas, G., Vidnyánszky, Z., 2010. Neural mechanisms for voice recognition. *Neuroimage* 52, 1528–1540.
- Arnal, L.H., Giraud, A.L., 2012. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398.
- Baldeweg, T., 2006. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn. Sci.* 10, 93–94.
- Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14, 2105–2109.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
- Bergerbest, D., Ghahremani, D.G., Gabrieli, J.D.E., 2004. Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *J. Cogn. Neurosci.* 16, 966–977.
- Birn, R.M., Cox, R.W., Bandettini, P.A., 2002. Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. *Neuroimage* 15, 252–264.
- Boynton, G.M., Finney, E.M., 2003. Orientation-specific adaptation in human visual cortex. *J. Neurosci.* 23, 8781–8787.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Nespoulous, J.L., Chollet, F., 1999. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9, 135–144.
- Chandrasekaran, B., Chan, A.H.D., Wong, P.C.M., 2011. Neural processing of what and who information in speech. *J. Cogn. Neurosci.* 23, 2690–2700.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.
- De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R.A., Clarke, S., Murray, M.M., 2010. Perceptual and semantic contributions to repetition priming of environmental sounds. *Cereb. Cortex* 20, 1676–1684.
- Egner, T., Monti, J.M., Summerfield, C., 2010. Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* 30, 16601–16608.
- Epstein, R.A., Parker, W.E., Feiler, A.M., 2008. Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *J. Neurophysiol.* 99, 2877–2886.
- Fecteau, S., Belin, P., Joanette, Y., Armony, J.L., 2007. Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36, 480–487.
- Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836.
- Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), 2007. *Statistical parametric mapping: the analysis of functional brain images*. Academic, London.
- Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23.
- Haenschel, C., Baldeweg, T., Croft, R.J., Whittington, M., Gruzelier, J., 2000. Gamma and beta frequency oscillations in response to novel auditory stimuli: a comparison of human electroencephalogram (EEG) data with in vitro models. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7645–7650.
- Hartwigsen, G., Baumgaertner, A., Price, C.J., Koehnke, M., Ulmer, S., Siebner, H.R., 2010. Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16494–16499.
- Henson, R.N., 2003. Neuroimaging studies of priming. *Prog. Neurobiol.* 70, 53–81.
- Henson, R., Shallice, T., Dolan, R., 2000. Neuroimaging evidence for dissociable forms of repetition priming. *Science* 287, 1269–1272.
- Henson, R.N.A., Shallice, T., Gorno-Tempini, M.L., Dolan, R.J., 2002. Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cereb. Cortex* 12, 178–186.
- Hubel, D.H., Wiesel, T.N., 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* 28, 229–289.
- Kaliukhovich, D.A., Vogels, R., 2011. Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb. Cortex* 21, 1547–1558.
- Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in psychtoolbox-3? *Perception* 36, 14 (EVP 2007 Abstract Supplement).
- Kovács, G., Iffland, L., Vidnyánszky, Z., Greenlee, M.W., 2012a. Stimulus repetition probability effects on repetition suppression are position invariant for faces. *Neuroimage* 60, 2128–2135.
- Kovács, G., Kaiser, D., Kaliukhovich, D., Vogels, R., 2012b. Stimulus repetition probability does not affect repetition suppression for non-face stimuli in the human lateral occipital cortex. 2012 Neuroscience Meeting Planner. Society for Neuroscience, New Orleans, LA (Online).
- Larsson, J., Smith, A.T., 2012. fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex* 22, 567–576.
- Larsson, J., Landy, M.S., Heeger, D.J., 2006. Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *J. Neurophysiol.* 95, 862–881.
- Latinus, M., Crabbe, F., Belin, P., 2011. Learning-induced changes in the cerebral processing of voice identity. *Cereb. Cortex* 21, 2820–2828.
- Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.
- Olson, I.R., Plotzker, A., Ezzyat, Y., 2007. The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 130, 1718–1731.
- Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I., 2011. Voice cells in the primate temporal lobe. *Curr. Biol.* 21, 1408–1415.
- Petkov, C.I., Kayser, C., Studel, T., Whittingstall, K., Augath, M., Logothetis, N.K., 2008. A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374.
- Riesenhuber, M., Poggio, T., 2000. Models of object recognition. *Nat. Neurosci.* (3 Suppl.), 1199–1204.
- Roach, P., Arnfield, S., Barry, W.J., Baltova, J., Boldea, M., Fourcin, A., Gonet, W., Gubrynowicz, R., Hallum, E., Lamel, L., Marasek, K., Marchal, A., Meister, E., Vicsi, K., 1996. BABEL: an Eastern European multi-language database. International Conference on Speech and Language Processing (ICSLP), Philadelphia, pp. 1892–1893.
- Shmuel, A., Augath, M., Oeltermann, A., Logothetis, N.K., 2006. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat. Neurosci.* 9, 569–577.
- Sirotnin, Y.B., Das, A., 2009. Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* 457, 475–479.
- Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006.
- Swallow, K.M., Braver, T.S., Snyder, A.Z., Speer, N.K., Zacks, J.M., 2003. Reliability of functional localization using fMRI. *Neuroimage* 20, 1561–1577.
- Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* 31, 9118–9123.
- Valentini, E., Torta, D.M.E., Mouraux, A., Iannetti, G.D., 2011. Dishabituation of laser-evoked EEG responses: dissecting the effect of certain and uncertain changes in stimulus modality. *J. Cogn. Neurosci.* 23, 2822–2837.