

# Brain networks underlying mental imagery of auditory and visual information

Mikhail Zvyagintsev,<sup>1,2,3</sup> Benjamin Clemens,<sup>2,4</sup> Natalya Chechko,<sup>1,3</sup> Krystyna A. Mathiak,<sup>3,5</sup> Alexander T. Sack<sup>6</sup> and Klaus Mathiak<sup>1,3,7</sup>

<sup>1</sup>Department of Psychiatry, Psychotherapy and Psychosomatics, Medical School, RWTH Aachen University, Aachen, Germany

<sup>2</sup>IZKF Aachen, RWTH Aachen University, Aachen, Germany

<sup>3</sup>JARA – Translational Brain Medicine, Aachen, Germany

<sup>4</sup>Neurological Clinic, Section Neuropsychology, Medical School, RWTH Aachen University, Aachen, Germany

<sup>5</sup>Department of Child and Adolescent Psychiatry, Medical School, RWTH Aachen University, Aachen, Germany

<sup>6</sup>Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands

<sup>7</sup>Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany

**Keywords:** attention, brain deactivation, default-mode network, fMRI, memory retrieval, mental imagery

## Abstract

Mental imagery is a complex cognitive process that resembles the experience of perceiving an object when this object is not physically present to the senses. It has been shown that, depending on the sensory nature of the object, mental imagery also involves correspondent sensory neural mechanisms. However, it remains unclear which areas of the brain subservise supramodal imagery processes that are independent of the object modality, and which brain areas are involved in modality-specific imagery processes. Here, we conducted a functional magnetic resonance imaging study to reveal supramodal and modality-specific networks of mental imagery for auditory and visual information. A common supramodal brain network independent of imagery modality, two separate modality-specific networks for imagery of auditory and visual information, and a common deactivation network were identified. The supramodal network included brain areas related to attention, memory retrieval, motor preparation and semantic processing, as well as areas considered to be part of the default-mode network and multisensory integration areas. The modality-specific networks comprised brain areas involved in processing of respective modality-specific sensory information. Interestingly, we found that imagery of auditory information led to a relative deactivation within the modality-specific areas for visual imagery, and vice versa. In addition, mental imagery of both auditory and visual information widely suppressed the activity of primary sensory and motor areas, for example deactivation network. These findings have important implications for understanding the mechanisms that are involved in generation of mental imagery.

## Introduction

Mental imagery is a complex cognitive process that resembles the experience of perceiving an object when this object is not physically present to the senses. Depending on the sensory nature of the object, mental imagery is characterized by a vivid re-experience of previously viewed visual material, heard auditory content or perceived other types of sensory information. For example, when we imagine a recent car drive, we re-experience the view from inside the car, the landscape outside the car, and music or voices from the radio. It is possible to voluntarily concentrate on information of a particular sensory modality that we intend to re-experience, and this process leads to mental imagery within the corresponding modality. Behavioural as well as neuroimaging studies provide strong evidence that such a vivid re-experience is very similar to the actual

perception of the same information (e.g. Kosslyn *et al.*, 1997; Ishai *et al.*, 2000; O'Craven & Kanwisher, 2000; Sack *et al.*, 2002, 2005; Ganis *et al.*, 2004; Mechelli *et al.*, 2004; also see Kosslyn *et al.*, 2001; Ishai, 2010a; for reviews). Neuroimaging studies showed that correspondent sensory association cortices are involved in sensory mental imagery. A few studies have also attempted to distinguish a supramodal network of mental imagery, i.e. regardless of its sensory modality. Recently, Daselaar *et al.* (2010) addressed the issue of modality-specific and modality-independent networks of mental imagery for auditory and visual stimuli in a functional magnetic resonance (fMRI) study. Based on a comparison between activation maps obtained during auditory and visual imagery, the authors concluded that the default-mode network (DMN) constitutes a modality-independent 'core' of the imagery network. However, the experimental design employed in Daselaar *et al.*, who used a fast stimulus presentation and a very short imagery phase (3 s), could have obscured some relevant changes in brain activity, and indeed the authors did not observe brain activations usually associated with

Correspondence: Dr M. Zvyagintsev, as above.  
E-mail: mzvyagintsev@ukaachen.de

Received 13 August 2012, revised 18 December 2012, accepted 21 December 2012

information retrieval (see Svoboda *et al.*, 2006; Binder *et al.*, 2009 for reviews). Moreover, previous studies showed that mental imagery is a cognitive process that consists of several stages, and the duration of these stages may vary substantially between participants (Sternberg, 1969; Formisano *et al.*, 2002; Sack *et al.*, 2002, 2005, 2008). Such an intersubject variability could obscure relevant brain activations in case of a short imagery phase. Studying supramodal and modality-specific networks of auditory and visual mental imagery, Daselaar *et al.* (2010) suggested that: (i) the supramodal network for mental imagery is not shared with perception of the same stimuli; and that (ii) supramodal activity may be attributed to the DMN. However, due to the experimental design employed in their study, these hypotheses need to be tested in other experimental settings. Interestingly, involvement of the core areas of DMN in mental imagery was often observed in fMRI and positron emission tomography (PET) studies of memory retrieval and mental imagery (see Ishai *et al.*, 2000; Svoboda *et al.*, 2006; Binder *et al.*, 2009), but their specific role in these processes was not clarified before.

In another fMRI study, Belardinelli *et al.* (2004) investigated patterns of activation corresponding to different imagery conditions and modalities, including mental imagination of shapes, sounds, touches, odours, flavours, self-perceived movements and internal sensations. The authors used block design with sufficiently long imagery intervals. They observed activation in the posterior part of the middle-inferior temporal cortex as a common neural substrate for all imagery conditions. The authors also emphasized the predominant role of the left hemisphere for all imagery conditions, and reported some common activity in the prefrontal and parietal areas, showing a more complex pattern across different imagery conditions. Belardinelli *et al.* (2004) attributed their findings to working memory and attentional processes, and suggested that semantic processes can strongly affect generation of mental images. Interestingly, for some imagery conditions they observed activation in brain areas constituting to the DMN (see Table 2). Their findings are complementary to the findings obtained by Daselaar *et al.* (2010). Unfortunately, Belardinelli *et al.* (2004) did not report whether relative deactivation of any brain region was associated with imagery or not. Previous fMRI studies showed that apart from brain activation, mental imagery also involves brain deactivation, and the level of deactivation has been shown to correlate to behavioural measures (Amedi *et al.*, 2005; Daselaar *et al.*, 2010). In sum, the studies performed by Daselaar *et al.* (2010) and Belardinelli *et al.* (2004) provide important but not complete and partly contradicting insights into the supramodal networks involved in mental imagery.

While the supramodal nature of mental imagery is still a matter of debate, visual (e.g. Ishai *et al.*, 2000; Sack *et al.*, 2002, 2005, 2008; Amedi *et al.*, 2005; Cui *et al.*, 2007) and auditory (Zatorre *et al.*, 1994; Halpern & Zatorre, 1999; Platel *et al.*, 2003; Kraemer *et al.*, 2005; Groussard *et al.*, 2010) imagery studies consistently report the involvement of visual and auditory association cortices, respectively. In particular, the ventral temporal area seems to play a key role in visual imagery (e.g. see Ishai, 2010a,b for a review). Moreover, there is evidence that activation of this area is context-dependent, and that the exact localization of activity varies depending on the type of visual information: imagery of objects activates object-related areas, whereas imagery of faces activates face-related areas in the fusiform gyrus (FG; Ishai *et al.*, 2000). However, the exact role of primary visual areas during imagery of the visual information still remains unclear. Brain lesion studies indicate that visual imagery is possible even without participation of these areas (Chatterjee & Southwood, 1995), but the results of neuroimaging studies are controversial. A PET study by Kosslyn *et al.* (1995) revealed

activation of the primary visual cortex during 'high-resolution' mental imagery. Another PET study from Mellet *et al.* (2000) challenged this hypothesis, by revealing relative deactivation within primary visual areas. Furthermore, Cui *et al.* (2007) did not find involvement of primary visual areas during visual imagery at the group level. At the same time, the authors discovered a positive correlation between the imagery score, measured with the visual imagery questionnaire, and the level of activation in the primary visual cortex at the individual level. The time courses in Cui *et al.* (2007) suggest slight deactivation of primary visual areas during imagery. Amedi *et al.* (2005) revealed deactivation of primary and association auditory cortices, which was negatively correlated to activation in the lateral occipital complex and the imagery score measured with the visual imagery questionnaire. Daselaar *et al.* (2010) found relative deactivation of primary visual areas during visual imagery as compared with auditory imagery. As concerns other brain areas that are involved in visual mental imagery, the study by Mechelli *et al.* (2004) stated that both the frontal eye fields (FEFs) and the superior parietal lobule (SPL) play a crucial role in generation of visual mental images. The authors noted that contribution of these areas is non-selective, context-unspecific and aids in the maintenance of the image in the 'mind's eye'. To conclude, the modality-specific network for visual imagery seems to be represented by context-specific activations (lateral occipital complex), and context-unspecific activations (FEF and SPL). Primary visual as well as auditory areas are slightly deactivated during visual imagery. However, such a summary is purely cumulative – so far no single study has provided clear evidence of such modality-specific network for visual imagery.

With regard to mental imagery of auditory information, several studies have revealed neural correlates of music imagery. The common finding of these studies is activation of the auditory association cortex within the superior temporal gyrus (STG) during re-experience of music (Halpern & Zatorre, 1999; Kraemer *et al.*, 2005; Groussard *et al.*, 2010). However, it is still unclear whether this region is involved in the re-experience of music exclusively, and whether the left and right hemispheres are involved differentially or not. On the one hand, Halpern & Zatorre (1999) described involvement of the right, but not the left, superior temporal cortex in imagery of music without lyrics. They did not report any significant activity within the left temporal lobe. On the other hand, Kraemer *et al.* (2005) found activity only in the left auditory association cortex regardless of whether songs contained lyrics or not. They also found that if the imagined songs contained no lyrics, cortical activity was extended into the left primary auditory cortex. Groussard *et al.* (2010) reported that both left and right auditory association cortices are involved in vivid music retrieval, irrespective of the semantic task used. It should be noted that in both fMRI studies performed by Groussard *et al.* (2010) and Kraemer *et al.* (2005), an event-related design was employed, and some temporally unsynchronized activity may not be detected with such a design (Sternberg, 1969; Bradburn *et al.*, 1987; Formisano *et al.*, 2002).

In the present study, we used fMRI to examine supramodal and modality-specific brain networks involved in mental imagery of auditory and visual information. We employed a standard block design with relatively long imagery periods (28 s) to reduce the influence of inter-individual differences in duration of different imagery phases on the brain activity as much as possible. We asked our participants to engage in mental imagery of either auditory or visual information, choosing music and visual objects, respectively. Music and visual objects can be considered equally complex stimuli for humans as they both represent a combination of essential features within the corresponding sensory modality, such as elementary

shapes, colours, etc. for the visual modality; and timbre, pitch, etc. for the auditory modality. Based on the previous findings summarized above, we hypothesized that imagery of both types of information should result in common activation of a supramodal network consisting of brain areas related to memory retrieval, attention, semantic processing, motor preparation, imagery and the DMN (Greenberg & Rubin, 2003; Svoboda *et al.*, 2006; Binder *et al.*, 2009; Daselaar *et al.*, 2010). Further, we expected that imagery of a correspondent sensory information involves modality-specific sensory areas (e.g. Ishai, 2010a,b). Contrasting activation maps for imagery of auditory and visual information and baseline, we revealed brain areas subserving mental imagery for each of the sensory modalities as well as those that subserved both modalities. We further reconstructed the time courses of the revealed brain areas and correlated their activity with behavioural measures such as vividness and difficulty of imagery. Finally, we separated the revealed brain areas into areas subserving mental imagery for both sensory modalities (supramodal network), and the areas subserving imagery for each modality separately (modality-specific networks). Importantly, the present study not only analysed task-related elevations of the blood oxygen level-dependent (BOLD) signal (positive BOLD response), but also focused on its relative deactivation during imagery. Such a negative BOLD signal indicates less neural processing during one task as compared with another task (Amedi *et al.*, 2005; Shmuel *et al.*, 2006).

## Materials and methods

### Participants

Fifteen right-handed participants (seven females, mean age = 25.1 years, SD = 5.7 years) participated in the present study. All participants had normal or corrected-to-normal vision, normal hearing, no contra-indications against MR investigations, no history of neurological or psychiatric illness, and no history of psychopharmacological therapy. The experiment was conducted in accordance with the Declaration of Helsinki, and the experimental protocol was approved by the Ethics Committee of the RWTH Aachen University. Written informed consent was obtained from all participants, following a complete description of the study and all experimental procedures.

### Procedure

The experiment consisted of two fMRI sessions. During a single session, participants were either engaged in mental imagery of visual (Visual) or auditory (Auditory) information, respectively. The order of the sessions was randomized across participants, and all participants were instructed to keep their eyes open during both sessions.

During the Auditory session, participants were instructed to recall familiar melodies that they encounter regularly, imagine listening to these melodies acoustically and to maintain this auditory imagination of one concrete melody in their 'mind's ear' within one fMRI block of 28 s. Participants were instructed to imagine one melody within one imagery block so that the melodies did not repeat within a session. Such a strategy (to imagine only one object within a single block) was chosen as previous studies (i.e. Sternberg, 1969; Bradburn *et al.*, 1987; Formisano *et al.*, 2002) had shown that imagery consists of several phases, such as semantic-related process, memory recall and working memory load. Choosing a design with imagery of one object per block, we attempted to avoid influence of these phases on imagery-related activation. Because all participants

were familiar with mp3 player devices, they were told to imagine melodies similar to listening to a song on the mp3 player. Participants were instructed to concentrate only on the auditory input and the auditory properties of the melody, such as pitch, rhythm and progression.

During the Visual session, participants were instructed to recall objects that they use daily, imagine those objects visually and maintain this visual imagination of one concrete object in their 'mind's eye' within one fMRI block of 28 s (compare Ishai *et al.*, 2000). The instruction was to imagine one particular visual object within one imagery block and not to imagine the same object twice within a session. Participants were explicitly instructed to 'reconstruct' with their 'mind's eye' the details of this object, such as colour, shape, size, etc. They were instructed to focus solely on the visual properties of the object itself, instead of focusing on visual properties more related to context or environment in which the object normally appears. With regard to the specific content of imagery, however, participants were free to choose any object. Free choice of the imagery object was previously applied by Schifferstein (2008). Based on this study we relied on the fact that such free choice provides sufficient and similar vividness for imagery of auditory and visual objects.

Before fMRI sessions, participants were trained in a brief practice session outside the MR scanner until they clearly understood the task and the strategy for either visual or auditory imagery. Immediately after each session, participants were asked to name objects of their choice. They also rated the vividness of each recall, its difficulty and emotional feelings associated with the recall. The vividness and difficulty were rated from 1 (small) to 4 (strong). Each session lasted 8 min and 30 s, consisting of eight 28-s recall blocks and nine 28-s baseline blocks. During the experimental run, participants were presented with visual instructions using MR-compatible goggles (VisuaStimDigital, Resonance Technology, Northridge, CA, USA). The instruction of the current task was displayed in front of a black background: 'auditory', 'visual', 'count back' for the Auditory, Visual and baseline task, respectively. Between the blocks, a hash mark (#) was shown for 2 s, to indicate a switch between the tasks. Each session started and finished with a baseline block (i.e. counting back task). With this task we attempted to achieve the following: (i) a 'wash-out' effect – so that participants can immediately switch between the tasks; (ii) to make auditory and visual conditions comparable; and (iii) to control for inner speech. The counting back task was also chosen because it can be assumed that such a task does not involve memory retrieval or imagery-related activity (e.g. Prado *et al.*, 2011). All participants were explicitly told not to imagine the numbers during calculation. In the brief practice session they were also instructed to develop a strategy to count back and not to imagine numbers herewith. After each session they were explicitly asked whether they imagined numbers during the counting back task. During the counting back task, participants were instructed to subtract 7 from 500 sequentially while keeping their eyes opened.

### Acquisition

Imaging data were collected using 3-Tesla Siemens Trio Scanner (Siemens, Erlangen, Germany) with a standard 12-channel head coil. Anatomical data were acquired using a T1-weighted MPRAGE sequence (TE = 2.98 ms; TR = 2300 ms; TI = 900 ms; flip angle = 9°; FOV = 256 × 256 mm<sup>2</sup>; voxel size = 1 × 1 × 1 mm<sup>3</sup>; 176 sagittal slices). Functional data were acquired using an EPI sequence (TE = 28 ms; TR = 2000 ms; flip angle = 77°; voxel size = 3 × 3 × 3 mm<sup>3</sup>; gap = 0.75 mm; FOV = 192 × 192 mm<sup>2</sup>; matrix size = 64 × 64; 34 transverse slices).

### Data analysis

fMRI data were analysed with the BrainVoyager QX 2.3 software package (Brain Innovation, Maastricht, the Netherlands). The first five images of each functional run were discarded to avoid T1 saturation effects. Preprocessing of functional data included slice-time correction using 'sinc' interpolation, motion correction using 'sinc' detection and interpolation, temporal filtering and spatial smoothing. Drift removal was achieved using a high-pass temporal filter (3 cycles/run, equivalent to 0.006 Hz), and high-frequency fluctuations were removed with a 4-s full-width at half-maximum Gaussian kernel. In the spatial domain, the data were smoothed with 8-mm-width Gaussian kernel. After preprocessing, functional data were co-registered to the individual high-resolution anatomical images. In an initial alignment step, the functional and anatomical data sets were co-registered based on the spatial position information recorded by the MR scanner. Subsequently, a more fine-grained alignment was achieved by applying the intensity-driven, multi-scale alignment procedure as implemented in BrainVoyager QX 2.3. The results of the alignment process were verified visually for each participant separately. Talairach transformation (Talairach & Tournoux, 1988) of the anatomical data sets was performed manually by aligning the sagittal data set with stereotactic axes (anterior and posterior commissure) and defining the extreme points of the cerebrum. The resulting Talairach transformation matrix was applied to both anatomical and functional images (including re-sampling of voxels to  $3 \times 3 \times 3 \text{ mm}^3$ ). The transformed anatomical data sets were averaged across all 15 participants. This average brain was further used as an anatomical mask for the general linear model (GLM) calculations. The functional volumetric time course data were further subjected to GLM analysis. For each participant, we defined predictors of interest corresponding to the Auditory and Visual conditions lasting 28 s each. Baseline predictors were defined implicitly. In addition, we also added six predictors representing the individual motion correction parameters (three rotational and three translational parameters) and five discrete cosine functions as confound predictors. The transition between the baseline and the imagery conditions was modelled using a separate predictor with a duration of 2 s for each transition. Time courses of the main and the latter predictors were derived by convolving an appropriate box-car waveform with a double-gamma haemodynamic response function (Friston *et al.*, 1998), in order to account for the shape, temporal delay and dispersion of the haemodynamic response. All predictors were *z*-transformed, and the single-subject GLMs of the experiment were computed from the *z*-normalized volumetric time course data. Further, individual GLMs were subjected to second-level random effects GLM analysis (RFX-GLM), including calculation and removal of serial correlations (AR 1) and subsequent refitting of the RFX-GLM to the data.

The obtained second-level GLM data were subjected to one-factor ANOVA analysis. The resulting statistical maps were corrected for multiple comparisons using the false discovery rate (FDR) approach, as described by Genovese *et al.* (2002). Within the framework of the ANOVA, the following contrasts were applied: Auditory vs. Baseline; Visual vs. Baseline; Auditory vs. Visual. The results were corrected using  $FDR(q) < 0.01$  and cluster size threshold  $> 20$  voxels equal to  $t > 3.50$  and  $P < 0.001$  (Fig. 1). We also calculated a conjunction map representing (Auditory vs. Baseline) and (Visual vs. Baseline) contrast based on second-level GLM data. In addition to that, an initial uncorrected threshold was applied ( $P = 0.01$ ) to all voxels, and afterwards a minimum cluster size was calculated that protected against false-positive clusters at 5% after 1000 Monte Carlo simulations (Forman *et al.*, 1995; Goebel *et al.*, 2006). All coor-

dinates of activation are reported in the coordinate system of Talairach & Tournoux (1988).

For the visualization purposes, statistical maps were projected both on the average brain and on the 3D cortical surface reconstruction representing the cortically aligned group average of the brains of all participants. In addition to that, we derived 3D cortical surfaces from the segmented brain of each individual participant. Such a surface-based, cortically driven inter-subject alignment of individual brains was recommended for multi-subject averaging in fMRI experiments investigating cortical structures (Fischl *et al.*, 1999). This procedure was performed as follows: first, we used largely automatic segmentation routines (Kriegeskorte & Goebel, 2001), to segment the grey/white matter boundary of each individual brain transformed into Talairach space. If necessary, additional manual corrections were applied to improve the results of the segmentation and to ensure that topologically correct mesh representations of each individual brain were created. Subsequently, the individual mesh representations of the 15 brains were aligned to one another using the cortex-based alignment procedures as implemented in BrainVoyager 2.3. The resulting 3D cortical surface representation was used for visualization of statistical results.

## Results

### Behavioural data

Participants' reports of vividness and difficulty of imagery for each object/melody were averaged for each session. The average vividness of visual imagery was  $2.7 \pm 0.6$  (mean  $\pm$  SD; scale from 1 to 4) and the average vividness of auditory imagery was  $2.6 \pm 0.6$ . The difficulty of the Visual condition was rated  $2.5 \pm 0.4$  and difficulty of the Auditory condition was rated  $3.0 \pm 0.5$ . These results were submitted to paired *t*-test statistics, revealing no significant difference between vividness of auditory and visual imagery ( $t_{14} = -0.13$ ,  $P = 0.9$ ). However, auditory imagery was subjectively more difficult than visual imagery ( $t_{14} = 4.8$ ,  $P < 0.001$ ). There was no significant correlation between vividness of auditory and visual imagery ( $r = 0.26$ ,  $P > 0.1$ ), or between vividness and difficulty of each type of imagery ( $r = 0.21$ ,  $P > 0.1$  and  $r = 0.4$ ,  $P > 0.1$  for visual and auditory imagery, respectively). However, we found a positive correlation between difficulty of auditory and visual imagery ( $r = 0.6$ ,  $P < 0.05$ ).

### fMRI data

#### Auditory imagery

The contrast Auditory vs. Baseline revealed brain activity that was significantly stronger in the Auditory condition as compared with the counting back task and vice versa. All clusters showing a significant difference in BOLD activity in this contrast [ $FDR(q) < 0.01$ ,  $P < 0.001$ ,  $t > 3.5$ , cluster size  $> 20$  voxels] are depicted in Fig. 1. Complete results for this contrast, including Talairach coordinates, *t*-statistics and *P*-values for peak voxel of all activated clusters are further summarized in Table 1. Localization of the peak voxel for each cluster was further specified using Talairach Client 2.4.2 software. During auditory imagery, we found increased activity in the inferior frontal gyrus (IFG) bilaterally (BA 11, 13, 45, 46, 47), left medial frontal gyrus (MeFG; BA 8, 10), left middle frontal gyrus (MFG; BA 6), left supplementary motor area (SMA; BA 6), left posterior cingulate cortex (PCC; BA 30), left angular gyrus (AG; BA 39), STG bilaterally (BA 22), left middle temporal gyrus (MTG; BA 21), parahippocampal gyrus (PHG) bilaterally (BA 36), cerebel-

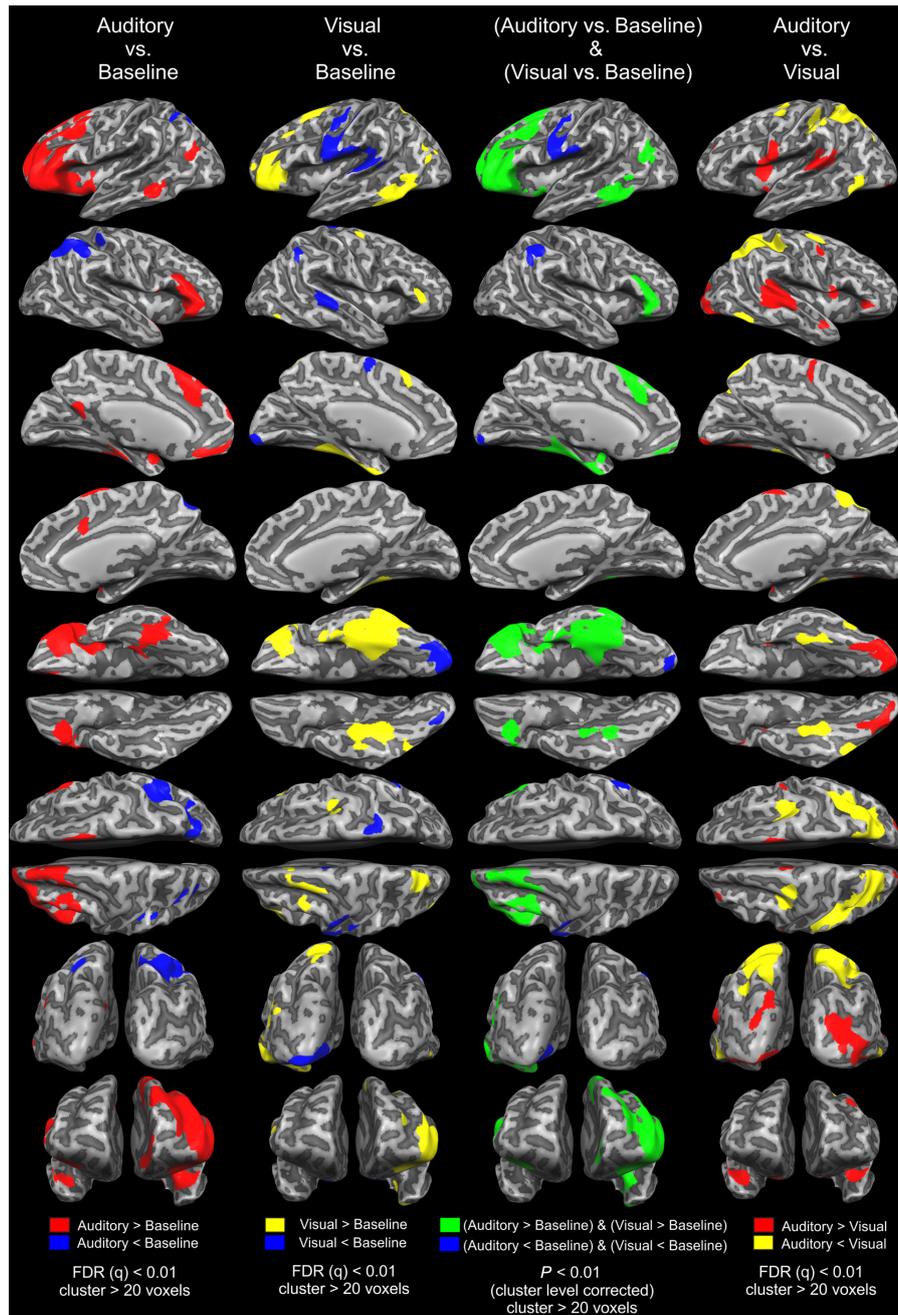


FIG. 1. Brain activation observed in the correspondent contrasts. Distributed areas of relative activation and deactivation can be observed in all the contrasts. The conjunction contrast (Auditory vs. Baseline) and (Visual vs. Baseline) reveals the areas that putatively represent modality-unspecific activation/deactivation, whereas the contrast Auditory vs. Visual displays the areas that activate differentially in the Auditory and Visual imagery conditions.

lum bilaterally, right thalamus and lentiform nucleus. Additionally, relative deactivation was observed in SPL and inferior parietal lobule (IPL) bilaterally (BA 7, 40). This activation pattern is consistent with previous neuroimaging studies investigating the neural correlates of music imagery (Halpern & Zatorre, 1999; Platel *et al.*, 2003; Groussard *et al.*, 2010) and memory retrieval (Svoboda *et al.*, 2006; Binder *et al.*, 2009).

#### Visual imagery

The contrast Visual vs. Baseline revealed brain activity that was stronger in the Visual conditions than in the counting back task and

vice versa. In accordance with our hypothesis and the results from previous neuroimaging studies (Ishai *et al.*, 2000; Sack *et al.*, 2002, 2005, 2008; Mechelli *et al.*, 2004; Amedi *et al.*, 2005; Cui *et al.*, 2007), we found increased activity for the Visual condition in IFG bilaterally (BA 46, 47), left MeFG (BA 8, 9), left MFG (BA 6), left SMA (BA 6), right FEF (BA 6), left precuneus and SPL (BA 7), PHG bilaterally (BA 36), FG bilaterally (BA 37), AG (BA 39), left superior occipital gyrus (BA 19), left MTG (BA 21) and right cerebellum as compared with the Baseline (Fig. 1; Table 2). For the contrast Visual vs. Baseline, relative deactivation was observed in STG bilaterally (BA 22, 41, 42), precentral gyrus (PG) bilaterally (BA 6), right IPL (BA 40), lingual gyrus (LG) bilaterally (BA 18),

TABLE 1. Clusters of activity observed in Auditory vs. Baseline contrast [FDR ( $q$ ) < 0.01, cluster size > 20 voxels]

<i>N</i>	Brain area	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i> -value	<i>P</i> -value	Size	BA
Auditory > Baseline								
1	IFG, STG (R)	47	31	-3	7.080292	0.000000	> 9.999	46, 47, 22
2	Cerebellum (R)	32	-74	-36	5.618052	0.000001	9.191	-
3	PHG, PCC, MTG, cerebellum (L)	-25	-29	-18	6.238183	0.000000	> 9.999	36, 30, 21
4	Thalamus, lentiform nucleus (R)	20	-14	6	4.746045	0.000024	757	-
5	IFG, MeFG, SMA, MFG, SFG, STG (L)	-49	40	-4	8.705457	0.000000	> 9.999	11, 13, 45, 46, 47, 8, 9, 6, 22
6	MeFG (L)	-4	52	-3	4.582251	0.000041	3.384	10
7	AG (L)	-40	-65	21	4.144580	0.000161	1.522	39
Baseline > Auditory								
1	SPL (L)	-22	-59	45	-5.039203	0.000009	1.756	7
2	IPL (L)	-40	-38	39	-4.426653	0.000067	953	40
3	SPL, IPL (R)	29	-65	45	-5.301348	0.000004	> 9.999	7,40

AG, angular gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; MeFG, medial frontal gyrus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; PCC, posterior cingulate cortex; PHG, parahippocampal gyrus; SFG, superior frontal gyrus; SMA, supplementary motor area; SPL, superior parietal lobule; STG, superior temporal gyrus.

TABLE 2. Clusters of activity observed in Visual vs. Baseline contrast [FDR ( $q$ ) < 0.01, cluster size > 20 voxels]

<i>N</i>	Brain area	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i> -value	<i>P</i> -value	Size	BA
Visual > Baseline								
1	IFG (R)	50	31	12	5.294889	0.000004	1.244	46, 47
2	PHG, FG (R)	29	-29	-18	6.756016	0.000000	8.909	36, 37
3	Cerebellum (R)	38	-65	-42	4.223292	0.000126	803	-
4	Cerebellum (R)	17	-32	-43	5.088282	0.000008	1.392	-
5	FEF (R)	32	-8	48	5.117187	0.000007	785	6
6	MFG, MeFG, SMA (L)	-19	10	57	5.257388	0.000005	5.669	6, 8, 9
7	Precuneus, SPL (L)	-16	-59	51	4.850259	0.000017	2.769	7
8	PHG, FG, MTG (L)	-31	-29	-18	9.867262	0.000000	> 9.999	36, 37, 21
9	IFG, MFG (L)	-46	37	6	6.748977	0.000000	> 9.999	46, 47
10	SOG (L)	-37	-80	27	4.619222	0.000036	641	19
11	AG (L)	-40	-68	21	4.500707	0.000053	619	39
Baseline > Visual								
1	STG (L)	50	-29	3	-4.931558	0.000013	2.353	22, 41, 42
2	IPL (R)	47	-50	45	-5.002246	0.000011	1.239	40
3	LG (R)	19	-80	-10	-4.056370	0.000212	564	18
4	Cerebellum (R)	20	-59	-21	-5.361018	0.000003	869	-
5	PG (R)	17	-32	69	-4.489368	0.000055	1.056	4
6	MOG, LG (L)	-22	-86	-9	-6.666066	0.000000	9.819	18
7	MeFG (L)	-7	-5	57	-5.616691	0.000001	603	6
8	PG, STG (L)	-61	-2	31	-7.954671	0.000000	> 9.999	6, 22, 41, 42

AG, angular gyrus; FEF, frontal eye field; FG, frontal gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; LG, lingual gyrus; MeFG, medial frontal gyrus; MFG, middle frontal gyrus; MOG, middle occipital gyrus; MTG, middle temporal gyrus; PG, precentral gyrus; PHG, parahippocampal gyrus; SMA, supplementary motor area; SOG, superior occipital gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus.

left middle occipital gyrus (MOG; BA 18), cerebellum and left MeFG (BA 6).

#### Auditory and visual imagery

The conjunction contrast (Auditory vs. Baseline) and (Visual vs. Baseline) allowed us to distinguish brain regions that are involved in imagery of both auditory and visual information. This contrast revealed brain activity that was stronger in both conditions as compared with the counting back task and vice versa. The set of brain areas commonly activated for both Auditory and Visual conditions included IFG bilaterally (BA 46, 47), PHG bilaterally (BA 36), left MFG (BA 6), left superior frontal gyrus (SFG; BA 8), left MTG (BA 21), left AG (BA 39), left SMA (BA 6), left MeFG (BA 10) and right cerebellum (Fig. 1; Table 3). Relative deactivation for both auditory and visual imagery was observed in the left LG (BA

18), left PG (BA 6), left STG (BA 42) and right IPL (BAs 7, 40). Overall, this activation pattern is consistent with the findings reported by Daselaar *et al.* (2011). However, in addition to their findings we also observed common activation in brain areas previously described as memory retrieval network in the meta-analysis performed by Svoboda *et al.* (2006).

#### Auditory vs. visual imagery

The contrast Auditory vs. Visual further allowed us to specify the differences between brain networks underlying imagery of the auditory and visual information. According to our hypothesis, we found that in the Visual condition the activity in FEF (BA 6), SPL/IPL (BA 7, 40), PHG (BA 36) and FG bilaterally (BA 37) was stronger compared with the Auditory condition. On the other hand, the activity in STG bilaterally (BA 22, 42), left MeFG (BA 6), left insula

TABLE 3. Clusters of activity observed in (Auditory vs. Baseline) and (Visual vs. Baseline) contrast ( $P < 0.01$ , cluster level corrected, cluster size  $> 20$  voxels)

N	Brain area	x	y	z	t-value	P-value	Size	BA
(Auditory > Baseline) and (Visual > Baseline)								
1	IFG (R)	50	31	12	6.900780	0.000000	7.183	9, 45, 46, 47
2	Cerebellum (R)	35	-68	-36	5.354670	0.000003	8.181	-
3	PHG (R)	26	-29	-18	5.681752	0.000001	5.390	36
4	IFG, MFG, SFG, MeFG, SMA, CG, PHG, AG, MTG (L)	-49	37	6	8.819283	0.000000	> 9.999	11, 45, 46, 47, 6, 8, 32, 36, 39, 21
5	MeFG (L)	-6	49	-6	4.235001	0.000122	1.148	10
(Auditory < Baseline) and (Visual < Baseline)								
1	LG (L)	-10	-86	-3	-4.617059	0.000037	1.739	18
2	PG, STG (L)	-58	-2	33	-6.293737	0.000000	7.409	6, 42
3	IPL (R)	47	-50	45	-5.600952	0.000001	3.127	7, 40

AG, angular gyrus; CG, central gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; LG, lingual gyrus; PG, precentral gyrus; PHG, parahippocampal gyrus; MeFG, medial frontal gyrus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; SFG, superior frontal gyrus; SMA, supplementary motor area; STG, superior temporal gyrus.

(BA 13), right IFG (BA 46), right MOG (BA 18), right PG (BA 4), right cuneus (BA 19), left MFG (BA 10), and left thalamus and lentiform nucleus was stronger in the Auditory condition as compared with the Visual one (Fig. 1; Table 4).

#### Regions of interest (ROIs) selection and average time course plots

Based on the fMRI results, which are summarized above and depicted in Fig. 1, we derived all ROIs from one of the above-mentioned contrasts (Table 5) at the correspondent threshold for further exploratory analysis (Poldrack, 2007). In general, ROIs were defined based on the brain areas activated in the (Auditory vs. Baseline) and (Visual vs. Baseline) contrast. However, because some regions

showed differential activity between the Auditory and Visual conditions, we defined those areas based on the results obtained from the Auditory vs. Visual contrast. Brain areas within the motor system and the visual cortex showing relative deactivation were defined based on the Visual vs. Baseline contrast. Furthermore, the left PCC was defined based on the Auditory vs. Baseline contrast. The large cluster encompassing the left IFG, MFG and SFG was separated into three smaller clusters solely in order to obtain more specific results for the left frontal regions. The separation process was done manually using standard BrainVoyager 2.3 procedures. It was based on an anatomical definition representing the three gyri, and it resulted in three separate clusters representing the IFG, MFG and SFG. In total, 26 ROIs were obtained. Then, the time courses of activity for these ROIs were created and averaged across blocks and

TABLE 4. Clusters of activity observed in Auditory vs. Visual contrast [FDR ( $q$ ) < 0.01, cluster size  $> 20$  voxels]

N	Brain area	x	y	z	t-value	P-value	Size	BA
Auditory > Visual								
1	STG (R)	41	-41	9	7.048966	0.000000	> 9.999	22, 41, 42
2	STG (R)	38	4	-18	6.155190	0.000000	5.448	22, 38
3	PG (R)	53	-5	45	5.124508	0.000007	927	4
4	IFG (R)	41	28	0	4.355637	0.000084	668	47
5	MOG (R)	24	-82	-3	4.941570	0.000013	> 9.999	18
6	Lentiform nucleus, thalamus, putamen, insula (R)	23	1	9	6.177456	0.000000	4.612	-
7	PG (R)	20	-26	63	4.611839	0.000037	711	4
8	MOG (L)	-19	-86	-6	6.043367	0.000000	> 9.999	18
9	Cuneus (R)	14	-89	34	4.150004	0.000159	590	19
10	Thalamus (L)	-7	-2	9	4.584326	0.000041	1.035	-
11	Cuneus (L)	-10	-92	31	5.164182	0.000006	1783	19
12	MFG (L)	-34	56	18	4.471950	0.000058	1.728	10
13	Insula, STG (L)	-33	2	12	6.436804	0.000000	> 9.999	13, 22, 38
14	STG (L)	-68	-29	12	6.186934	0.000000	4.722	22, 42
15	MeFG (L)	-1	-5	63	6.407889	0.000000	4440	6
Visual > Auditory								
1	FG, ITG (R)	50	-53	-6	-5.266391	0.000004	2.726	37
2	IPL, SPL, precuneus (R)	23	-59	45	-6.999419	0.000000	> 9.999	7, 40
3	PHG, FG (R)	32	-35	-12	-4.778229	0.000022	2.008	36, 37
4	FEF, MFG (R)	26	-8	51	-6.428694	0.000000	3.227	6
5	Precuneus, IPL, SPL (L)	-19	-59	48	-8.461289	0.000000	> 9.999	7, 40
6	FEF, MFG (L)	-28	-8	45	-5.984469	0.000000	3.280	6
7	PHG (L)	-34	-23	-21	-5.493242	0.000002	2.781	36
8	MOG, FG, ITG (L)	-52	-62	-6	-6.405902	0.000000	3.287	19, 37

FEF, frontal eye field; FG, fusiform gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; MeFG, medial frontal gyrus; MFG, middle frontal gyrus; MOG, middle occipital gyrus; PG, precentral gyrus; PHG, parahippocampal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus.

TABLE 5. ROIs for the ROI RFX-GLM and correlation analysis were selected based on the maps obtained from different contrasts

Contrast	Area
(Auditory vs. Baseline) and (Visual vs. Baseline) $P < 0.01$ , cluster level corrected cluster size $> 20$ voxels	(1–3) IFG L, MFG L, SFG L (4) IFG R (5) PC L (6) AG L (7) SMA L (8) IPL R (9) Cerebellum R
Visual vs. Auditory FDR ( $q$ ) $< 0.01$ cluster size $> 20$ voxels	(10–11) FG L&R (12–13) SPL L&R (14–15) FEF L&R (16–17) STG L&R (BA 22) (18–19) PHG L&R (20–21) STG L&R (BA 42)
Auditory vs. Baseline FDR ( $q$ ) $< 0.01$ cluster size $> 20$ voxels	(22) PCC L (23) MTL L
Visual vs. Baseline FDR ( $q$ ) $< 0.01$ cluster size $> 20$ voxels	(24–25) LG L&R (26) PG L

AG, angular gyrus; FDR, false discovery rate; FEF, frontal eye field; FG, fusiform gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; LG, lingual gyrus; MFG, middle frontal gyrus; MTL, medial temporal lobe; PC, postcentral gyrus; PCC, posterior cingulate cortex; PG, precentral gyrus; PHG, parahippocampal gyrus; SFG, superior frontal gyrus; SMA, supplementary motor area; SPL, superior parietal lobule; STG, superior temporal gyrus.

participants for the Auditory and Visual conditions. Averaged time course plots were calculated for all the voxels within the entire ROI. We further performed ROI-based RFX-GLM analysis to statistically validate the differences in activities of the selected ROIs between the conditions of the experiment.

The event-related average time course plots and the ROI-based RFX-GLMs revealed that the activity of the ROIs can be classified into the following patterns.

1. The activity in the Auditory condition was stronger than in the Baseline condition, and the activity in the Visual condition was equal or weaker than in the Baseline condition. The areas included the left and right STG (BA 22; Fig. 2; Fig. S1).

2. The activity in the Visual condition was stronger than in the Baseline condition, and the activity in the Auditory condition was equal or weaker than in the Baseline condition. These areas included SPL L&R (BA 7, 40), FEF L&R (BA 6), FG L&R (BA 37; Fig. 3; Fig. S2).

3. The activity in both conditions was stronger than in the Baseline condition. The areas that showed such a pattern of activity included IFG L&R (BA 46, 47), MFG L (BA 6), SFG L (BA 6), SMA L (BA 6), MTG L (BA 21), AG L (BA 39), MeFG L (BA 10), PCC L (BA 30) and cerebellum (Fig. 4; Fig. S3).

4. The activity in both conditions was equal or weaker than the activity in the Baseline condition. The areas included PG L&R (BA 4), STG L&R (BA 41, 42), LG L&R (BA18), IPL R (BA 40; Fig. 5; Fig. S4).

#### Exploratory correlation analysis between fMRI results and behavioural measures

In order to reveal whether the strength of brain (de)activation was correlated with behavioural measures, we extracted averaged beta-values for each brain area of interest and each participant from the results of the GLM analysis and correlated them with the behavioural measures. More specifically, we computed Pearson correlation coefficients between the obtained beta-values and the vividness score of each condition. We found that vividness in the Auditory condition was positively correlated with activity in the left SMA ( $r = 0.65$ ,  $P < 0.01$ ), right IFG ( $r = 0.65$ ,  $P < 0.01$ ) and left STG (BA22;  $r = 0.63$ ,  $P = 0.01$ ). Vividness in the Visual condition was positively correlated with the activity in bilateral FG ( $r = 0.84$ ,  $P < 0.01$  and  $r = 0.67$ ,  $P < 0.01$  for the left and right hemisphere, respectively), left PHG ( $r = 0.61$ ,  $P < 0.01$ ), left MTG ( $r = 0.63$ ,  $P < 0.01$ ) and left IFG ( $r = 0.71$ ,  $P < 0.01$ ). Then we correlated the beta-values with the difficulty score. The difficulty score of the

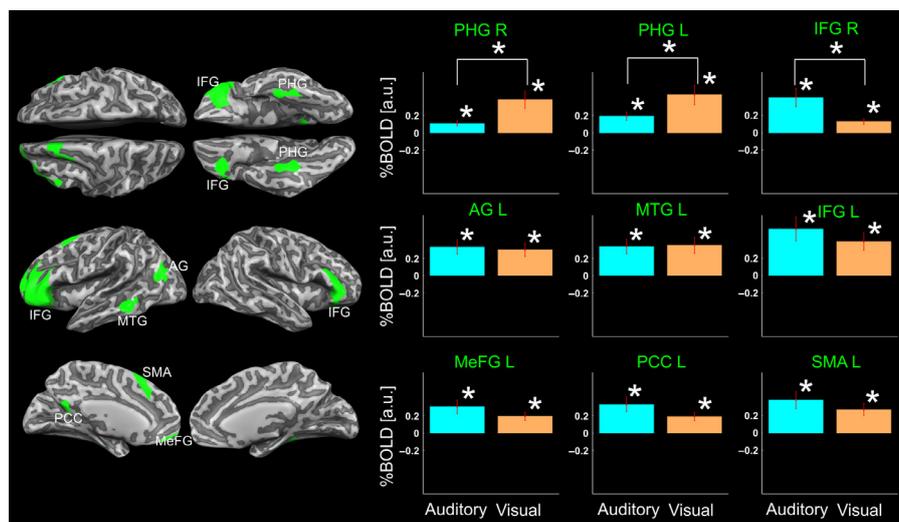


FIG. 2. The regions of interest (ROIs) and their relative change of activity ( $\pm$  SE) for the areas that exhibited positive deflection in both Visual and Auditory conditions (see also Fig. S1). Left-lateralized topography of activation can be noted. The statistically significant differences of the levels of activation between the Auditory, Visual conditions and the Baseline are marked ( $P < 0.05$ ). Stronger activation of the right inferior frontal gyrus (IFG) for the Auditory condition, and bilateral parahippocampal gyrus (PHG) for the Visual one can be noted.

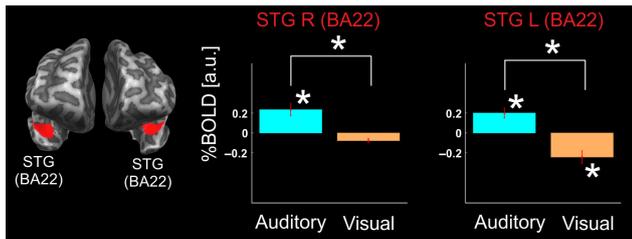


FIG. 3. The regions of interest (ROIs) and their relative change of activity ( $\pm$  SE) for the areas that exhibited positive deflection in the Auditory condition and negative deflection in the Visual condition (see also Fig. S2). Bilateral location within the auditory association cortex can be noted. The statistically significant differences of the levels of activation between the Auditory, Visual conditions and the Baseline are marked ( $P < 0.05$ ).

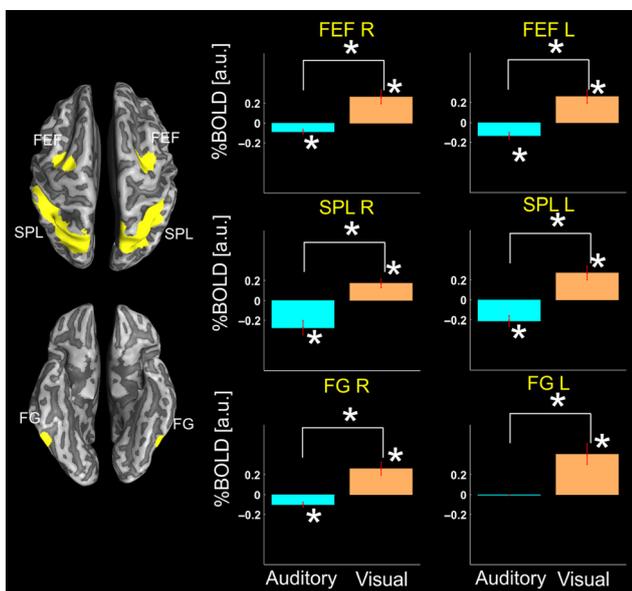


FIG. 4. The regions of interest (ROIs) and their relative change of activity ( $\pm$  SE) for the areas that exhibited positive activity deflection in the Visual condition and negative deflection in the Auditory condition (see also Fig. S3). The bilateral areas within both the ventral and dorsal visual pathways can be observed. The statistically significant differences of the levels of activation between the Auditory, Visual conditions and the Baseline are marked ( $P < 0.05$ ).

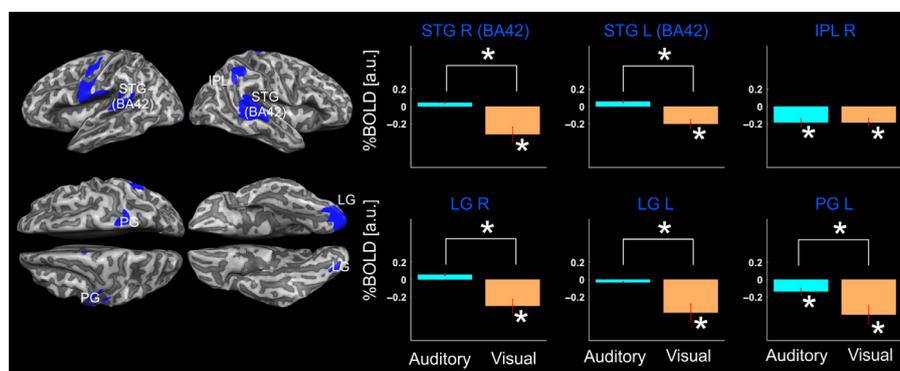


FIG. 5. The regions of interest (ROIs) and their relative change of activity ( $\pm$  SE) for the areas that exhibited negative deflection in either both Visual and Auditory conditions, or only in the Visual condition. The bilateral topography can be noted. The statistically significant differences of the levels of activation between the Auditory, Visual conditions and the Baseline are marked ( $P < 0.05$ ). A distributed deactivation network in the Visual condition can be noted. Auditory imagery is accompanied with deactivation in the right inferior parietal lobule (IPL) and left precentral gyrus (PG).

Auditory and Visual conditions did not correlate significantly with activity in any brain area. We also correlated the difference between beta-values obtained for the Auditory and the Visual predictors with the difference of the difficulty scores between the conditions. We always used difficulty score for the Auditory minus difficulty score for the Visual condition. This difference was positively correlated with the difference between beta-values in the right IFG ( $r = 0.61$ ,  $P < 0.01$ ), left MeFG ( $r = 0.61$ ,  $P = 0.01$ ), left PCC ( $r = 0.69$ ,  $P < 0.01$ ), left AG ( $r = 0.66$ ,  $P < 0.01$ ), left MFG ( $r = 0.61$ ,  $P = 0.01$ ) and left SFG ( $r = 0.65$ ,  $P < 0.01$ ; Fig. 6).

## Discussion

The present fMRI study investigated the supramodal and modality-specific neural activity underlying mental imagery of visual and auditory information. Overall, the results showed that imagery of both auditory and visual information is associated with changes of activity within widely distributed brain areas (Fig. 1). ROI analysis of activity of these brain areas and correlation analysis of their activity and the behavioural measures such as vividness and difficulty of imagery in the Auditory and Visual conditions allowed us to combine these areas into brain networks for mental imagery. In particular, four patterns of the time course activity change were detected. Each pattern was observed within a distributed brain network. The first pattern represented an increase of activity in both Auditory and Visual conditions (Fig. 2), and was observed in the network including the frontal, parietal, temporal areas and cerebellum. The difference in the activity of some of these areas between the Auditory and Visual conditions was positively correlated with the difference of the difficulty score between these conditions. The second pattern represented an increase of activity in the Auditory condition, and a decrease or no change of activity in the Visual one (Fig. 3), and was observed in the bilateral STG (BA22). The activity in the left STG was positively correlated with the vividness score in the Auditory condition. The third pattern represented an increase of activity in the Visual condition, and a decrease or no change of activity in the Auditory one (Fig. 4), and was observed in the FG, FEF and SPL bilaterally. Moreover, the activity in both FG positively correlated with the vividness score for the Visual condition. The fourth pattern represented a decrease of activity during either Auditory and Visual or at least Visual conditions (Fig. 5), and was observed in the primary sensory and motor areas.

Finally, we composed four putative networks that play different roles in mental imagery of auditory and visual information (Fig. 7):

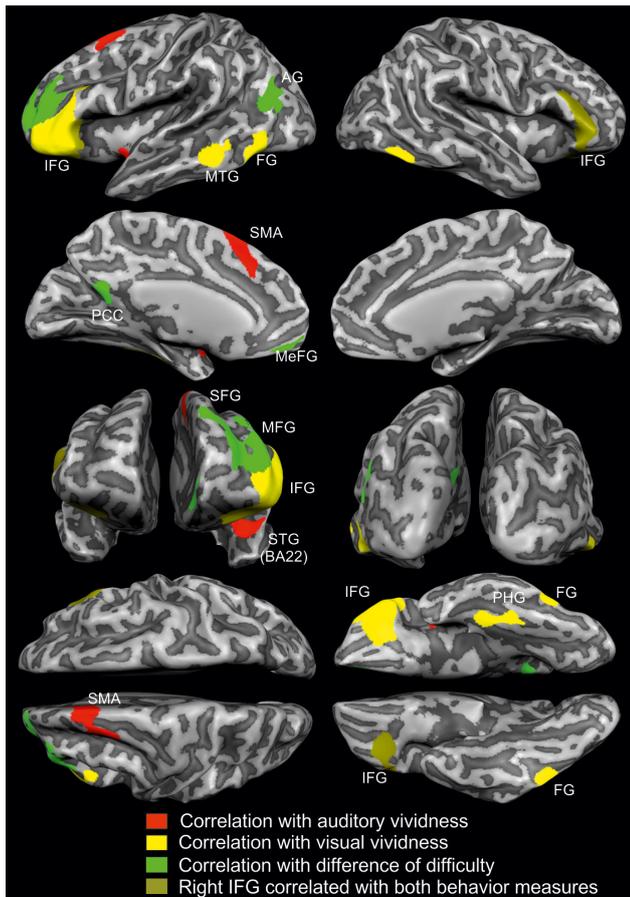


FIG. 6. The regions of interest (ROIs) where the beta-values of activity correlated with the subjective scores of vividness and difficulty for the Auditory/Visual conditions. The activity in both putatively auditory modality-specific and some of the supramodal areas for mental imagery positively correlated with the vividness score in the Auditory condition. Also, a positive correlation was observed between the activity in visual modality-specific as well as some other supramodal areas and the vividness score in the Visual condition. Although we did not find any correlation between the difficulty score and the brain activity, the difference of the difficulty scores between the conditions positively correlated with the difference between the levels of activity in these conditions. The latter correlations were mainly observed in the areas representing the default-mode network (DMN).

1. Supramodal network;
2. Auditory network;
3. Visual network;
4. Deactivation network.

#### Supramodal network of auditory and visual imagery

The supramodal network was represented by the brain areas that showed positive deflection in the time course activity in both Auditory and Visual conditions (Fig. 2). It included IFG L&R, MFG L, SFG L, SMA L, PHG L&R, AG L, MTG L, MeFG L, PCC L and cerebellum (Fig. 7). These areas can also be observed in the conjunction contrast (Auditory vs. Baseline) and (Visual vs. Baseline; Fig. 1). The supramodal network seems to be responsible for both types of mental imagery regardless of the modality. Brain areas of the supramodal network corroborate results from studies on mental imagery as well as memory retrieval (Belardinelli *et al.*, 2004; Svoboda *et al.*, 2006; Binder *et al.*, 2009; Daselaar *et al.*, 2010). Previ-

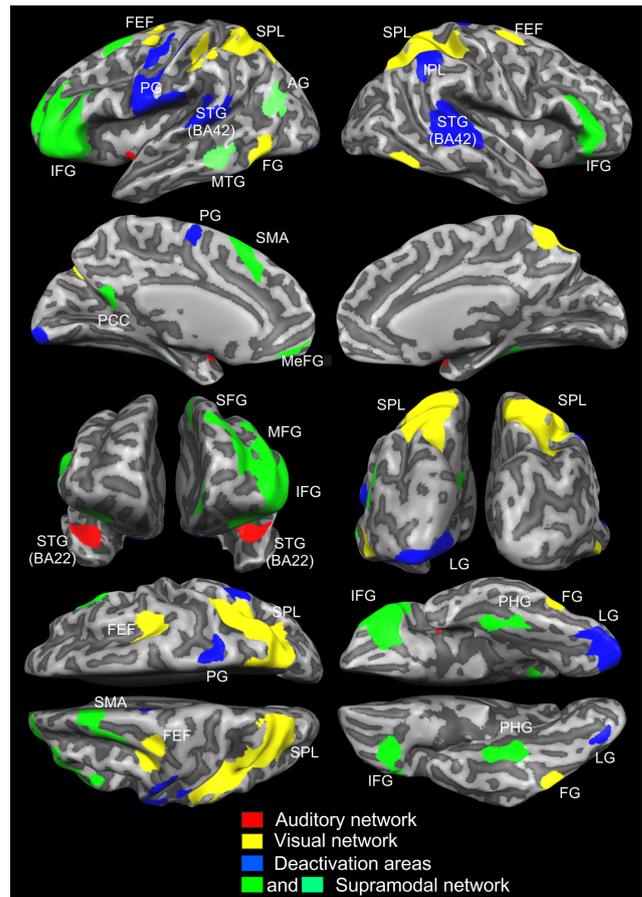


FIG. 7. Summary of the brain networks for mental imagery (see text).

ous findings (Greenberg & Rubin, 2003; Buckner *et al.*, 2008; Andrews-Hanna *et al.*, 2010; Daselaar *et al.*, 2010) suggest that the revealed supramodal network is composed of six functionally separated sub-components: memory retrieval-related areas, such as the PHG and IFG bilaterally (Halpern & Zatorre, 1999; Svoboda *et al.*, 2006; see Binder *et al.*, 2009 for review); attention-related areas, such as the left SFG and left MFG (Corbetta & Shulman, 2002; Clemens *et al.*, 2011); semantic processing areas, such as left IFG (Svoboda *et al.*, 2006); motor preparation areas, such as left SMA (Picard & Stick, 2001); the DMN areas, such as left MeFG and left PCC (Daselaar *et al.*, 2010); and multisensory integration areas, that includes the left MTG and left AG. The latter areas are known to be involved in multisensory integration mechanisms, including visual and tactile stimulation (i.e. Ionta *et al.*, 2011), or visual and auditory stimulation (i.e. Kamke *et al.*, 2012). Animal studies also showed that neurons in AG discharge in response to at least three different stimulation modalities, including visual, tactile and vestibular (Grusser *et al.*, 1990; Bremmer *et al.*, 2002). This body of evidence on the multimodal nature of AG and MTG supports their inclusion within the supramodal imagery network. Moreover, in this view the present study extends the multisensory role of AG and MTG from sensory stimulation to mental imagery on the supramodal level.

The contrast Auditory vs. Visual as well as the time courses of activity within the areas of the supramodal network further specified the differences between the Auditory and Visual conditions within this network. Particularly, it was found that auditory imagery involves stronger activation of the right IFG. Moreover, similar to the findings of Herholz *et al.* (2012), the activity of this area was

also correlated with the vividness score in the Auditory condition. Thus, our study further supports the hypothesis that the right IFG is involved in processing of auditory information as it was proposed by previous studies (Zatorre *et al.*, 1994; Mathiak *et al.*, 2004), which suggested its involvement in auditory working memory. Taken together, IFG is responsible for the mnemonic component of mental imagery, which is more prominent in auditory imagery (Herholz *et al.*, 2012), which explains its correlation with vividness in the Auditory condition. Further, it was found that imagery of auditory information involves stronger activation within the left PCC and left MeFG. These areas represent the core areas of the DMN (Daselaar *et al.*, 2010; see Andrews-Hanna *et al.*, 2010 for a review). In the present study, the difference in the level of activation between the Auditory and the Visual conditions within the DMN correlated with the difference of the difficulty score for imagery of auditory and visual information. In this sense, imagery of auditory information probably implies more efforts in internal mental activity than imagery of visual information (Andrews-Hanna *et al.*, 2010). Interestingly, this specific finding is in accordance with the results presented by Daselaar *et al.* (2010; Fig. 2). However, in their study the difference between auditory and visual imagery conditions was not significant, probably due to the study design. Using a block design with relatively long intervals for mental imagery, we were able to specify differences between imagery of auditory and visual information in activation patterns within the 'core' parts of the DMN. Our results also show that imagery of auditory information involves stronger activation within the left MFG, bilateral insula, left thalamus and right lentiform nucleus. We suggest that this activation may reflect increased cognitive load during the Auditory condition, as also indicated by the subjectively higher level of the task difficulty during this condition. During imagery of visual information, bilateral PHG showed stronger activation as compared with imagery of auditory information. This may suggest that PHG plays a similar role in imagery of visual information as the IFG does in imagery of auditory information. In other words, PHG is responsible for visual memory.

With regard to the role of SMA, previous studies suggested that it is involved in sub-vocalization and mental singing (Zatorre *et al.*, 1996; Lotze *et al.*, 2003; Kleber *et al.*, 2007). Unless we did not find a significant difference in this area between the Auditory and Visual conditions, the activation cluster for the Auditory condition was much larger (Fig. 1) and its activity correlated to the vividness score in the same condition (Fig. 6). This further suggests that SMA is responsible for sub-vocalization during both imagery conditions.

Taken together, these findings complement the conclusions drawn by Daselaar *et al.* (2010), who proposed that the DMN represents the core of the modality-independent imagery network. Our results extend this conclusion, specifying that the supramodal network is widely distributed, and includes memory retrieval-related areas, attention-related areas, semantic processing areas, motor preparation areas, areas of the DMN, as well as areas involved in multisensory processing. We hypothesize here that only the latter seem to be involved in supramodal imagery network *per se* unifying neural mechanisms of imagery and perception, whereas other areas are responsible for accompanying imagery processes, such as memory retrieval, sub-vocalization, attention and introspection.

#### *Auditory network: modality-specific network for imagery of auditory information*

The auditory network was represented by the brain areas that showed positive deflection in the Auditory condition and negative

deflection in the Visual one. The auditory network is responsible for modality-specific imagery of auditory information. Our results show that modality-specific areas for auditory imagery are located in the auditory association cortex bilaterally (STG; BA22; Figs 3 and 6), corroborating findings from previous research (Halpern & Zatorre, 1999; Kraemer *et al.*, 2005; Daselaar *et al.*, 2010; Groussard *et al.*, 2010). The present results also confirm the results presented by Amedi *et al.* (2005), who suggested that the STG is deactivated in visual imagery. Furthermore, these results are in agreement with the results obtained by Groussard *et al.* (2010), finding a bilateral involvement of STG in imagery of music. Here we also show that the level of activation in STG correlates with the subjective vividness score of auditory imagery, and that the same area (bilateral STG) is activated in music imagery and suppressed in imagery of visual information.

Concerning the activity of the primary auditory cortex, it is important to note that similar to fMRI studies of auditory/music retrieval/imagery, we did not observe a significant increase of activity in the primary auditory areas, but rather in the auditory association cortex (e.g. anterior/middle part of STG at BA 22). Interestingly, PET studies (Zatorre *et al.*, 1996; Halpern & Zatorre, 1999) on auditory/music imagery found activation in the primary auditory areas such as BA 41, 42, whereas comparable fMRI studies did not report involvement of these areas (compare with Kraemer *et al.*, 2005). The present study revealed the dual pattern of activity of the primary auditory cortex with a long deactivation phase (see below). Nevertheless, the majority of the neuroimaging studies, regardless of the technique used to assess brain activity, seem to agree upon the role of STG (BA 22) in auditory imagery/retrieval tasks (e.g. Groussard *et al.*, 2010). This statement was confirmed in our study.

#### *Visual network: modality-specific network for imagery of visual information*

The visual network was represented by the areas that showed a positive deflection in the Visual condition, and a negative deflection in the Auditory condition (Fig. 4). This network is responsible for modality-specific imagery of visual information, and includes areas within the ventral and the dorsal visual pathways. Our study suggests that the visual network comprises the FG bilaterally (BA37) and a frontoparietal network (FPN) involving the SPL and FEF bilaterally (BA 7 and 6, respectively; Figs 4 and 6). These findings are in agreement with previous research (Ishai *et al.*, 2000; Mechelli *et al.*, 2004) showing activation of the FG and SPL/FEF during visual imagery. Our results extend the existing knowledge by providing two novel aspects: (i) only the activity in bilateral FG was correlated with the subjective vividness score; (ii) the entire modality-specific visual imagery network is activated during visual imagery, but deactivated during auditory imagery. The former observation extends the conclusion drawn by Mechelli *et al.* (2004), who suggested that only FG shows the content-related activity whereas the FPN contributes to generation of mental images regardless of their content. The FG is a part of the visual ventral pathway, and its activation was often documented in neuroimaging studies on mental imagery (e.g. Ishai *et al.*, 2000). Our study confirms the importance of FG in visual mental imagery and provides evidence that FG activity correlates with the subjective vividness score. The activity of the FPN was also increased, indicating a higher level of visual working memory load and increased visuo-spatial attention demands during visual imagery (Corbetta & Shulman, 2002; Périn *et al.*, 2010; Clemens *et al.*, 2011). As such, our findings could also

be taken to imply that visual imagery involves mechanisms that are similar to goal-directed visual search.

### Deactivation network

The deactivation network was represented by the brain areas that showed a decrease of activity in either both Auditory and Visual, or at least Visual conditions (Fig. 5). This network includes PG, LG, STG (BA 42) and IPL. This finding suggests that imagery of both visual and auditory information leads to a decrease of activity within the primary sensory and motor areas. This is the first study to describe such a distributed decrease of the BOLD signal in the primary auditory, visual and motor cortices (c.f. Amedi *et al.*, 2005; Daselaar *et al.*, 2010). It seems, therefore, that attention to imagery of sensory information suppressed currently irrelevant types of sensory/motor activity. In the present study, visual imagery was accompanied with stronger deactivation within sensory/motor areas than it was in auditory imagery. Most likely, such differential suppression here was related to the experimental design – all the instructions were presented visually. As such, visual imagery resulted in strong suppression of the irrelevant perceptual information, segregating two visual flows. Finally, it seems that such top-down suppression is modality-unspecific, and involves both primary sensory and motor areas. It is also important that these results do not support previous findings of Kosslyn *et al.* (1995), where they found activation of the primary visual cortex in high-resolution visual imagery. In fact, one may expect that long visual imagery of a particular object may elicit such high-resolution imagery. Nonetheless, similar to subsequent previous findings on neural correlates of visual imagery (Ishai *et al.*, 2000; Amedi *et al.*, 2005; Daselaar *et al.*, 2010) or high-resolution visual imagery (Mellet *et al.*, 2000), we did not report activation of the primary visual cortex in mental imagery.

In summary, our data show that top-down modulation during both auditory and visual imagery resulted in deactivation of primary sensory/motor cortices as well as deactivation of modality-specific imagery areas of a currently irrelevant imagery modality. The latter finding further suggests a competition between imagery modalities. Most likely, this modality-specific suppression is related to the strategies of imagery: irrelevant imagery activities are actively suppressed by the subject. However, the exact nature of such competition cannot be clarified with the present study, and future studies should concentrate on revealing mechanisms of such an antagonistic relationship between imagery modalities.

Interestingly, the present study revealed a two-stage pattern of the time course activity within primary sensory and motor areas (Fig. 5). During the first stage we observed relative activation, and relative deactivation during the second stage. The exact meaning of this dual behaviour cannot be clarified completely within the present fMRI investigation, but perhaps this pattern represents the brain's reaction to the tasks' change (Dosenbach *et al.*, 2006). Such a dual pattern of activation may help to explain why in some studies primary sensory/motor areas were activated, whereas the same areas were deactivated in other studies. We also did not observe any correlation between deactivation of the primary sensory/motor areas and vividness of imagery.

### Limitations

A potential limitation of the present study might have been the use of the counting back task as a high-level baseline control condition, making it difficult to fully disentangle activities related to imagery/counting. Based on previous investigations concerning brain activity

related to number processing, counting and general arithmetic processes (for a review, see Dehaene *et al.*, 2004), we suggest that activation in the right IPL observed in the present study might have been related to the counting back task, and indeed the right IPL showed relative deactivation in all three contrasts: Auditory vs. Baseline; Visual vs. Baseline; (Auditory vs. Baseline) and (Visual vs. Baseline) (Fig. 1). Moreover, the time course of the right IPL did not differ between the Auditory and Visual conditions (Fig. 2; Fig. S1). Another limitation could be that we did not control the mnemonic component of mental imagery (e.g. Herholz *et al.*, 2012).

### Conclusion

In summary, this study yielded three novel findings. First, we described a distributed supramodal network for mental imagery, two modality-specific networks for auditory and visual imagery, and a deactivation network consisting of the brain areas that are suppressed during imagery. Second, we found that the modality-specific networks for imagery of visual and auditory information showed a dual pattern of activation – they were activated during imagery of the correspondent modality and deactivated during imagery of the other modality. Finally, we showed that vividness of mental imagery was correlated with activity within the modality-specific networks, whereas difficulty of imagery was correlated with activity in the supramodal network and the DMN in particular.

The modality-specific areas for both auditory and visual imagery were located bilaterally, whereas the supramodal network was lateralized to the left hemisphere, corroborating the conclusions of the meta-analysis by Svoboda *et al.* (2006). With regard to the supramodal imagery network, our findings imply that this network includes multisensory integration areas as well as the areas that serve accompanying imagery mental activities. These mental activities include semantic processing, memory retrieval, attention, subvocalization and introspection. This hypothesis further implies that mental imagery shares its network with perception on the supramodal level and the above-mentioned activities. Further, the possible difference between imagery and not-vivid retrieval of information lies at the level of activation within modality-specific areas and supramodal multisensory integration areas. Future studies are needed to rigorously test this hypothesis.

Interestingly, although the supramodal network for mental imagery was the same, its activity resulted in different mental representations – namely, either auditory or visual imagery. Hence, the question appears: what is the exact mechanism for differentiation between imagery modalities? The design of our study did not allow us to distinguish the mechanisms enabling the brain to switch between imagery of auditory/visual information. Nevertheless, based on our findings, we may propose the following scenario for this switching process: storage and retrieval of semantic auditory and visual information is modality-independent and takes place in memory storage areas, such as the IFG and PHG (Svoboda *et al.*, 2006; Binder *et al.*, 2009). Involvement of the modality-specific areas seems to be based on the type of retrieved information, and future studies should concentrate on identifying the 'trigger' area that differentiates the type of information that has to be retrieved/imagined.

### Supporting Information

Additional supporting information can be found in the online version of this article:

Fig. S1. The ROIs and their time courses for the areas which exhibited positive deflection in both Visual and Auditory conditions.

Fig. S2. The ROIs and their time courses for the areas which exhibited positive deflection in the Auditory condition and negative deflection in the Visual condition.

Fig. S3. The ROIs and their time courses for the areas which exhibited positive activity deflection in the Visual condition and negative deflection in the Auditory condition.

Fig. S4. The ROIs and their time courses for the areas which exhibited negative deflection in both Visual and Auditory conditions. The bilateral topography can be noted.

Fig. S5. The ROIs and their time courses for all four types of areas (see text and Fig. S1–S4).

## Acknowledgements

This research was supported by the DFG (IRTG 1328, MA2631/4-1) and the ICCF Aachen (N4-2). We thank Andrey Nikolaev (KU Leuven, Belgium) for valuable discussions. We thank all anonymous reviewers for their thorough evaluation and constructive recommendations for improving this manuscript.

## Abbreviations

AG, angular gyrus; BOLD, blood oxygen level-dependent; DMN, default-mode network; FDR, false discovery rate; FEF, frontal eye field; FG, fusiform gyrus; fMRI, functional magnetic resonance imaging; FPN, frontoparietal network; GLM, general linear model; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; LG, lingual gyrus; MeFG, medial frontal gyrus; MFG, middle frontal gyrus; MOG, middle occipital gyrus; MTG, middle temporal gyrus; PCC, posterior cingulate cortex; PET, positron emission tomography; PG, precentral gyrus; PHG, parahippocampal gyrus; ROI, region of interest; SFG, superior frontal gyrus; SMA, supplementary motor area; SPL, superior parietal lobule; STG, superior temporal gyrus.

## References

Amedi, A., Malach, R. & Pascual-Leone, A. (2005) Negative BOLD differentiates visual imagery and perception. *Neuron*, **48**, 859–872.

Andrews-Hanna, J.R., Reidler, J.S., Huang, C. & Buckner, R.L. (2010) Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.*, **104**, 322–335.

Belardinelli, M.O., Di Matteo, R., Del Gratta, C., De Nicola, A., Ferretti, A., Tartaro, A., Bonomo, L. & Romani, G.L. (2004) Intermodal sensory image generation: An fMRI analysis. *Eur. J. Clin. Pharmacol.*, **16**, 729–752.

Binder, J.R., Desai, R.H., Graves, W.W. & Conant, L.L. (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex*, **19**, 2767–2796.

Bradburn, N.M., Rips, L.J. & Shevell, S.K. (1987) Answering autobiographical questions: the impact of memory and inference on surveys. *Science*, **236**, 157–161.

Bremmer, F., Klam, F., Duhamel, J.R., Ben Hamed, S. & Graf, W. (2002) Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.*, **16**, 1569–1586.

Buckner, R.L., Andrews-Hanna, J.R. & Schacter, D.L. (2008) The brain's default network: anatomy, function, and relevance to disease. *Ann. NY Acad. Sci.*, **1124**, 1–38.

Chatterjee, A. & Southwood, M.H. (1995) Cortical blindness and visual imagery. *Neurology*, **45**, 2189–2195.

Clemens, B., Zvyagintsev, M., Sack, A., Heinecke, A., Willmes, K. & Sturm, W. (2011) Revealing the Functional Neuroanatomy of Intrinsic Alertness Using fMRI: methodological Peculiarities. *PLoS ONE*, **6**, e25453.

Corbetta, M. & Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, **3**, 201–215.

Cui, X., Jeter, C.B., Yang, D., Montague, P.R. & Eagleman, D.M. (2007) Vividness of mental imagery: individual variability can be measured objectively. *Vision Res.*, **47**, 474–478.

Daselaar, S.M., Porat, Y., Huijbers, W. & Pennartz, C.M. (2010) Modality-specific and modality-independent components of the human imagery system. *NeuroImage*, **52**, 677–685.

Dehaene, S., Molko, N., Cohen, L. & Wilson, A.J. (2004) Arithmetic and the brain. *Curr. Opin. Neurobiol.*, **14**, 218–224.

Dosenbach, N.U., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L. & Petersen, S.E. (2006) A core system for the implementation of task sets. *Neuron*, **50**, 799–812.

Fischl, B., Sereno, M.I., Tootell, R.B. & Dale, A.M. (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.*, **8**, 272–284.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. & Noll, D.C. (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.*, **33**, 636–647.

Formisano, E., Linden, D.E., Di Salle, F., Trojano, L., Esposito, F., Sack, A.T., Grossi, D., Zanella, F.E. & Goebel, R. (2002) Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron*, **35**, 185–194.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A.P., Rugg, M.D. & Turner, R. (1998) Event-related fMRI: characterizing differential responses. *NeuroImage*, **7**, 30–40.

Ganis, G., Thompson, W.L. & Kosslyn, S.M. (2004) Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Brain Res. Cogn. Brain Res.*, **20**, 226–241.

Genovese, C.R., Lazar, N.A. & Nichols, T.E. (2002) Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, **15**, 870–878.

Goebel, R., Esposito, F. & Formisano, E. (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.*, **27**, 392–401.

Greenberg, D.L. & Rubin, D.C. (2003) The neuropsychology of autobiographical memory. *Cortex*, **39**, 687–728.

Groussard, M., Rauchs, G., Landeau, B., Viader, F., Desgranges, B., Eustache, F. & Platel, H. (2010) The neural substrates of musical memory revealed by fMRI and two semantic tasks. *NeuroImage*, **53**, 1301–1309.

Grusser, O.J., Pause, M. & Schreier, U. (1990) Localization and responses of neurones in the parieto-insular vestibular cortex of awake monkeys (*Macaca fascicularis*). *J. Physiol.*, **430**, 537–557.

Halpern, A.R. & Zatorre, R.J. (1999) When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex*, **9**, 697–704.

Herholz, S.C., Halpern, A.R. & Zatorre, R.J. (2012) Neuronal Correlates of Perception, Imagery, and Memory for Familiar Tunes. *J. Neurosci.*, **24**, 1382–1397.

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., Gassert, R. & Blanke, O. (2011) Multisensory mechanisms in tempo-parietal cortex support self-location and first-person perspective. *Neuron*, **70**, 363–374.

Ishai, A. (2010a) Seeing faces and objects with the “mind's eye”. *Arch. Ital. Biol.*, **148**, 1–9.

Ishai, A. (2010b) Seeing with the mind's eye: top-down, bottom-up, and conscious awareness. *F1000 Biol Rep.*, **2**, 34.

Ishai, A., Ungerleider, L.G. & Haxby, J.V. (2000) Distributed neural systems for the generation of visual images. *Neuron*, **28**, 979–990.

Kamke, M.R., Vieth, H.E., Cottrell, D. & Mattingley, J.B. (2012) Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage*, **62**, 1334–1341.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T. & Lotze, M. (2007) Overt and imagined singing of an Italian aria. *NeuroImage*, **36**, 889–900.

Kosslyn, S.M., Thompson, W.L., Kim, I.J. & Alpert, N.M. (1995) Topographical representations of mental images in primary visual cortex. *Nature*, **378**, 496–498.

Kosslyn, S.M., Thompson, W.L. & Alpert, N.M. (1997) Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *NeuroImage*, **6**, 320–334.

Kosslyn, S.M., Ganis, G. & Thompson, W.L. (2001) Neural foundations of imagery. *Nat. Rev. Neurosci.*, **2**, 635–642.

Kraemer, D.J.M., Macrae, C.N., Green, A.E. & Kelley, W.M. (2005) Musical imagery: sound of silence activates auditory cortex. *Nature*, **434**, 158.

Kriegeskorte, N. & Goebel, R. (2001) An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage*, **14**, 329–346.

Lotze, M., Scheler, G., Tan, H.R.M., Braun, C. & Birbaumer, N. (2003) The musician's brain: functional imaging of amateurs and professionals during performance and imagery. *NeuroImage*, **20**, 1817–1829.

- Mathiak, K., Hertrich, I., Grodd, W. & Ackermann, H. (2004) Discrimination of temporal information at the cerebellum: functional magnetic resonance imaging of nonverbal auditory memory. *NeuroImage*, **21**, 154–162.
- Mechelli, A., Price, C.J., Friston, K.J. & Ishai, A. (2004) Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb. Cortex*, **14**, 1256–1265.
- Mellet, E., Tzourio-Mazoyer, N., Bricogne, S., Mazoyer, B., Kosslyn, S.M. & Denis, M. (2000) Functional anatomy of high-resolution visual mental imagery. *J. Cogn. Neurosci.*, **12**, 98–109.
- O'Craven, K.M. & Kanwisher, N. (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.*, **12**, 1013–1023.
- Périn, B., Godefroy, O., Fall, S. & de Marco, G. (2010) Alertness in young healthy subjects: an fMRI study of brain region interactivity enhanced by a warning signal. *Brain Cognition*, **72**, 271–281.
- Picard, N. & Stick, P.L. (2001) Imaging the premotor areas. *Curr. Opin. Neurobiol.*, **11**, 663–672.
- Platel, H., Baron, J.C., Desgranges, B., Bernard, F. & Eustache, F. (2003) Semantic and episodic memory of music are subserved by distinct neural networks. *NeuroImage*, **20**, 244–256.
- Poldrack, R.A. (2007) Region of interest analysis for fMRI. *SCAN*, **2**, 67–70.
- Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A.S., Minas, J.E. & Booth, J.R. (2011) Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Hum. Brain Mapp.*, **32**, 1932–1947.
- Sack, A.T., Sperling, J.M., Prvulovic, D., Formisano, E., Goebel, R., Di Salle, F., Dierks, T. & Linden, D.E. (2002) Tracking the mind's image in the brain II: transcranial magnetic stimulation reveals parietal asymmetry in visuospatial imagery. *Neuron*, **35**, 195–204.
- Sack, A.T., Camprodon, J.A., Pascual-Leone, A. & Goebel, R. (2005) The dynamics of interhemispheric compensatory processes in mental imagery. *Science*, **308**, 702–704.
- Sack, A.T., Jacobs, C., De Martino, F., Staeren, N., Goebel, R. & Formisano, E. (2008) Dynamic premotor-to-parietal interactions during spatial imagery. *J. Neurosci.*, **28**, 8417–8429.
- Schiffstein, H.N.J. (2008) Comparing Mental Imagery Across the Sensory Modalities Imagination. *Cognition and Personality*, **28**, 371–388.
- 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.
- Sternberg, S. (1969) Memory scanning: mental processes revealed by reaction-time experiments. *Am. Sci.*, **57**, 421–457.
- Svoboda, E., McKinnon, M.C. & Levine, B. (2006) The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia*, **44**, 2189–2208.
- Talairach, J. & Tournoux, P. (1988) *Co-planar stereotaxic atlas of the human brain: three-dimensional proportional system*. Georg Thieme, Stuttgart.
- Zatorre, R.J., Evans, A.C. & Meyer, E. (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.*, **14**, 1908–1919.
- Zatorre, R.J., Halpern, A.R., Perry, D.W., Meyer, E. & Evans, A.C. (1996) Hearing in the mind's ear: a PET investigation of musical imagery and perception. *J. Cognitive Neurosci.*, **8**, 29–46.