Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems

César F. Lima[1,2*], Nadine Lavan[1,3*], Samuel Evans[1], Zarinah Agnew[1,4], Andrea R. Halpern[5], Pradheep Shanmugalingam[1], Sophie Meekings[1], Dana Boebinger[1], Markus Ostarek[1], Carolyn McGettigan[1,3], Jane E. Warren[6], Sophie K. Scott[1]

[1]Institute of Cognitive Neuroscience, University College London, UK

[2]Center for Psychology, University of Porto, Portugal

[3]Department of Psychology, Royal Holloway University of London, UK

[4]Department of Otolaryngology, University of California, San Francisco, USA

[5]Department of Psychology, Bucknell University, USA

[6]Faculty of Brain Sciences, University College London, UK

* These authors have made equal contributions to the work

Manuscript accepted for publication in Cerebral Cortex

May 2015

Address correspondence to César Lima, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR. E-mail: c.lima@ucl.ac.uk or cesarflima@gmail.com

Running title: Auditory imagery and brain structure

Conflict of interest: None declared

Abstract

Humans can generate mental auditory images of voices or songs, sometimes perceiving them almost as vividly as perceptual experiences. The functional networks supporting auditory imagery have been described, but less is known about the systems associated with inter-individual differences in auditory imagery. Combining voxel-based morphometry and fMRI, we examined the structural basis of inter-individual differences in how auditory images are subjectively perceived, and explored associations between auditory imagery, sensory-based processing and visual imagery. Vividness of auditory imagery correlated with grey matter volume in the supplementary motor area (SMA), parietal cortex, medial superior frontal gyrus, and middle frontal gyrus. An analysis of functional responses to different types of vocalizations revealed that the SMA and parietal sites that predict imagery are also modulated by sound type. Using representational similarity analysis, we found that higher representational specificity of sounds in SMA predicts vividness of imagery, indicating a mechanistic link between sensory- and imagery-based processing in sensorimotor cortex. Vividness of imagery in the visual domain also correlated with SMA structure, and with auditory imagery scores. Altogether, these findings provide evidence for a signature of imagery in brain structure, and highlight a common role of perceptual-motor interactions for processing heard and internally generated auditory information.

Keywords: auditory imagery; auditory processing; supplementary motor area; voxel-based morphometry; fMRI

Introduction

Imagine the voice of a close friend when you laugh together, or a piano playing your favorite song. Auditory imagery is a complex process by which an individual generates and processes mental images in the absence of sound perception – "hearing with the mind's ear". Auditory mental images can be so vivid that they resemble the real experience of hearing, and they can be as accurate as representations arising directly from sensory input (Janata 2012). They facilitate several cognitive and motor processes. In music performance, for instance, imagery supports action planning, formation of expectations about upcoming events, and interpersonal coordination (Keller 2012; Novembre et al. 2014). Functional neuroimaging studies have shown that the network of brain regions engaged during auditory imagery minimally includes the superior temporal gyri (STG), parietal, motor and premotor cortices, the inferior frontal gyrus, and the SMA (Shergill et al. 2001; Herholz et al. 2012; Zvyagintsev et al. 2013; for a meta-analysis, McNorgan 2012).

The involvement of STG in auditory imagery has been suggested to reflect the reconstruction of sound-like representations via higher-order cortical mechanisms, contributing to the subjective experience of "hearing" (Kraemer et al. 2005; Zatorre and Halpern 2005). The superior parietal cortex is associated with the manipulation of imagined auditory events, for example when the task requires participants to mentally reverse the notes of a melody (Zatorre et al. 2010). Frontal regions are assumed to underlie general control, working memory, retrieval and semantic processes (Zvyagintsev et al. 2013). The SMA and premotor cortices seem to be directly involved in generating auditory images (Halpern and Zatorre 1999; Herholz et al. 2012), implicating an intimate link between sensorimotor and imagery processes. Consistent with the idea that auditory-motor interactions may be involved in auditory imagery, in a functional magnetic resonance imaging (fMRI) study, Kleber et al. (2007) showed that the premotor cortex and SMA are active both when professional singers overtly sing an Italian aria and when they are asked to imagine the act of singing as vividly as possible without performing any movements. Functional imaging work has additionally revealed that auditory imagery recruits brain networks that also respond to heard auditory information (Zatorre et al. 1996; Kosslyn et al. 2001; Zatorre and Halpern 2005; Herholz et al. 2012). For instance, Zatorre et al. (1996) asked participants to make pitch judgments about words taken from familiar tunes in an imagery condition, in which there was no auditory input, and in a perceptual condition, in which participants could actually hear the song. Common activations were found across conditions despite the differences of input, including the temporal and frontal lobes, the supramarginal gyrus, midbrain, and SMA.

We have a good picture of the functional networks that are active during auditory imagery tasks, but a common aspect to many of the available studies is that findings are based on group averages – similarities across individuals are privileged over inter-individual differences so that general processes may be inferred. Less is known about the predictors of individual differences in how people experience auditory images, or about which neural systems account for these differences. These questions matter, as behavioral data reveal considerable variability in how well individuals perform on tasks that engage imagery abilities, e.g., judging whether or not a final probe note of a scale is mistuned when the initial notes were played but the remaining ones had to be imagined (Janata 2012). People also vary widely in how vividly they experience auditory mental images, as measured by self-report on the Bucknell Auditory Imagery Questionnaire (BAIS; Pfordresher and Halpern 2013). In that study, higher vividness of imagery predicted better accuracy in a pitch imitation task in which participants reproduced sequences of pitches, suggesting that the sensorimotor components of imagery play a role in planning and guiding vocal imitation. In two fMRI studies, individual differences in the BAIS correlated with blood oxygen level-dependent (BOLD) responses in the right superior parietal cortex during a task involving mental reversal of melodies (Zatorre et al. 2010), and in the right STG, right dorsolateral prefrontal cortex and left frontal pole during imagery of familiar tunes (Herholz et al. 2012).

Crucial to the understanding of inter-individual differences in imagery is the question of whether they are determined by the local structure of grey matter. A growing number of studies indicates that individual differences in a range of basic and higher-order cognitive functions are reflected in brain structure, as measured using techniques such as voxel-based morphometry (VBM) and diffusion tensor imaging (for a review, Kanai and Rees 2011). Differences in brain structure have been reported among groups of experts, such as musicians (Gaser and Schlaug 2003), taxi drivers (Woollett and Maguire 2011) and phoneticians (Golestani et al. 2011), as well as in samples from the general population. For instance, among people with no particular expertise, increased grey matter volume in the left thalamus predicts enhanced ability to adjust to degraded speech (Erb et al. 2012), and in the right anterior prefrontal cortex it predicts the ability to introspect about self-performance during perceptual decisions (Fleming et al. 2010).

In the present study, we examine for the first time whether differences in brain structure predict differences in how auditory images are subjectively experienced. Grey matter volume was measured using VBM, and auditory imagery was evaluated in terms of perceived vividness, as well as in terms of perceived control over mental representations, i.e., the ease with which people can change or manipulate representations (Pfordresher and Halpern 2013; Halpern in press). Two additional novel questions were addressed. *First*, we combined VBM and fMRI approaches to investigate whether the structural predictors of imagery co-localize with systems that also play a role in the processing of heard auditory information. Importantly, in addition to looking at co-localization, we examined possible co-variation between inter-individual differences in auditory imagery and in the patterns of online functional responses to auditory input. Electrophysiological studies have shown similar modulations of the N100 component by imagery and sensory-based auditory processes (Navarro-Cebrian and Janata 2010a), and imaging studies have reported common activations during imagery and auditory processing (Zatorre et al. 1996; Kosslyn et al. 2001;

Zatorre and Halpern 2005; Herholz et al. 2012), a result suggestive of converging mechanisms. However, because co-localization does not necessitate shared function (e.g., Woo et al. 2014), more direct evidence for links between the processing of heard and internally generated auditory information is needed. *Second,* in an additional VBM study we aimed to determine the extent to which the structural predictors of auditory imagery reflect the operation of mechanisms that are specialized to auditory information. To that end, links with visual imagery were investigated. Research on imagery is typically confined to a single modality, but some fMRI studies suggest that whereas the STG may play an auditory-specific role, the SMA, premotor, parietal, and prefrontal regions may be involved in imagery within and beyond the auditory domain, forming a modality-independent "core" imagery network (Daselaar et al. 2010; McNorgan 2012; Burianová et al. 2013). Therefore, the current study takes advantage of combining behavioral with structural and functional measures to shed new light on the neural underpinnings of inter-individual differences in auditory imagery, and on how these differences may reflect mechanisms shared with sensory-based processing and the operation of supra-modal processes.

Materials and Methods

Participants

Seventy-four participants were included in the study looking at the structural correlates of auditory imagery (M_{age} = 42.61, *SD* = 17.11; range = 20-81; 40 female). None reported a diagnosis of neurological or psychiatric disorders. Written informed consent was collected and ethical approval was obtained from the UCL Research Ethics Committee. All structural scans were reviewed by a neurologist to identify anatomical abnormalities that could affect their suitability for VBM; this led to the exclusion of 2 participants of the 76 initially included. No participants had significant cognitive impairment (all participants aged ≥ 50 years completed the Montreal Cognitive Assessment, M_{score} = 28, max 30; *SD* = 1.68;

range = 25-30; www.mocatest.org). The participants' age range was wide because these data were collected as part of a larger project on neurocognitive ageing. All participants completed the forward condition of the digit span test of the Wechsler Adult Intelligence Scale (WAIS-III, Wechsler 1997; average number of digits correctly recalled = 7.08; *SD* = 1.21; range = 4-9). Short-term memory is highly correlated with working memory and intelligence (Colom et al. 2008), and therefore it was used as a proxy for general cognitive abilities. Thirty participants had some degree of musical training ($M_{\text{years of training}}$ = 6.03, *SD* = 4.47; range = 1-20).

From the 74 participants, 56 completed the fMRI study examining brain responses during auditory processing (M_{age} = 47.05, *SD* = 17.23; range = 20-81; 31 female).

Forty-six participants took part in the follow-up VBM study looking at the links between auditory and visual imagery (44 of them also participated in the first VBM study; M_{age} = 47.13, *SD* = 17.83; range = 20-81; 24 female).

Materials

Individual differences in imagery

To assess auditory imagery, we used the BAIS (Pfordresher and Halpern 2013; Halpern in press), a self-report measure that includes two 14-item subscales. The first subscale focuses on *vividness* of imagery: participants are asked to generate a mental image of the sound described in each item, and to rate its subjective clarity in a 7-point scale (1 = no image present at all; 7 = as vivid as actual sound), e.g., "consider ordering something over the phone; the voice of an elderly clerk assisting you"; "consider attending classes; the slow-paced voice of your English teacher". The second subscale focuses on *control* of imagery: participants are asked to generate mental images corresponding to pairs of items, and to consider how easily they can change the first image to the second image (1 = no image present at all; 7 = extremely easy to change the item), e.g., "consider ordering something over the phone; image a – the voice of an elderly clerk assisting you; image b – the elderly clerk leaves and the voice of a younger clerk is now on the line". Most of the items cover vocal and musical sounds, with only a minority of them focusing exclusively on environmental sounds (3 items in each subscale; e.g., the sound of gentle rain). The BAIS has appropriate psychometric properties, including high internal reliability, a coherent factor structure, and no association with social desirability (Halpern in press). It has been used in behavioral (Pfordresher and Halpern 2013; Gelding et al. 2015) and fMRI studies (Zatorre et al. 2010; Herholz et al. 2012).

To assess visual imagery, we used the Vividness of Visual Imagery Questionnaire (VVIQ; Marks 1973). In this task, participants are given four hypothetical scenarios and generate four mental images corresponding to different aspects of each scenario, forming 16 items in total (e.g., contour of faces; color and shape of trees; attitudes of body of a friend or relative). Responses are provided on a scale from 1 (perfectly clear and vivid as normal vision) to 5 (no image at all), i.e., lower scores correspond to higher vividness, unlike the BAIS in which the direction of the scale is reversed. For ease of interpretation, scores were inverted so that higher scores correspond to higher vividness both in the auditory (BAIS) and visual domains (VVIQ). The VVIQ is the most frequently used self-report measure of vividness of visual imagery. It has appropriate internal reliability (Kozhevnikov et al. 2005; Campos and Pérez-Fabello 2009) and correlates with brain responses during visual perception and imagery (Cui et al. 2007).

Auditory stimuli

The auditory stimuli used in the fMRI study consisted of five types of human vocal sounds. These included vowels spoken with a neutral intonation (e.g., prolonged "a"), laughter, screams, and sounds of pleasure and disgust (retching sounds). Similarly to imagery processes, these vocal communicative signals are known to engage auditory systems, as well as sensorimotor and control systems involved in higher-order mechanisms and social behavior (Warren et al. 2006; McGettigan et al. 2015). The five sound types were matched for duration (M_{duration} = 1018 ms; *SD* = 326), and 20 different examples of each were included in the experiment (they were generated by 8 different speakers, 4 women; for further details about the stimuli, Sauter et al. 2010; Lima et al. 2013). A sixth condition, intended as an unintelligible distractor set, consisted of sounds created by spectral rotation of a selection of the original vocal sounds. Rotated sounds were generated by inverting the frequency spectrum around 2 kHz, using a digital version of the simple modulation technique described by Blesser (1972). The acoustic signal was first equalised with a filter (essentially high-pass) that gave the rotated signal approximately the same long-term spectrum as the original. This equalizing filter (33-point finite impulse response) was constructed based on measurements of the long-term average spectrum of speech (Byrne et al. 1994), although the roll-off below 120Hz was ignored, and a flat spectrum below 420Hz was assumed (Scott, Rosen et al. 2009; Green et al. 2013). The equalized signal was then amplitude modulated by a sinusoid at 4kHz, followed by low pass filtering at 3.8kHz. Spectral rotation retains the acoustic complexity of human sounds while rendering them unintelligible. Rotated sounds are used in numerous imaging studies of vocalizations and speech perception (Scott et al. 2000; Narain et al. 2003; Warren et al. 2006; Okada et al. 2010; Evans and Kyong et al. 2014).

MRI acquisition and data processing

MRI data were acquired using a 32-channel birdcage headcoil on a Siemens 1.5T Sonata MRI scanner (Siemens Medical, Erlangen, Germany). High-resolution anatomical images were acquired using a T1-weighted MPRAGE sequence (repetition time = 2730 ms, echo time = 3.57 ms, flip angle = 7º, slice thickness = 1 mm, 160 sagittal slices, acquisition matrix = 256 x 224 x 160 mm, voxel size = 1 mm^3). Echo planar fMRI images were acquired with repetition time = 9 s, TA = 3 s, echo time = 50 ms, flip angle = 90º, 35 axial slices, 3

mm^3 in-plane resolution, using a sparse-sampling routine in which sounds were presented in the silent gap between brain acquisitions (Hall et al. 1999).

Voxel-based morphometry

The structural images were subjected to VBM, as implemented in SPM8 (Wellcome Trust Centre for Neuroimaging, UK). SPM8 provides an integrated routine that combines segmentation into different tissue classes, bias correction, and spatial normalization in the same model (New Segment). After being re-oriented into a standard space (via manual alignment along the anterior-posterior commissure), each participant's T1-weighted image was segmented into grey matter, white matter, and cerebro-spinal fluid. Diffeomorphic Anatomical Registration was performed through exponentiated lie algebra (DARTEL) for non-linear inter-subject registration of the grey and white matter images (Ashburner 2007). This involves iteratively matching the images to a template generated from their own mean, i.e., sample-specific grey and white matter templates were generated.

Because we were interested in differences across subjects in the absolute *amount* (volume) of grey matter, the spatial normalization step was implemented with modulation in order to preserve the total amount of grey matter signal in the normalized partitions. This is necessary as the process of normalizing images introduces volumetric changes in brain regions; for the structural images to be aligned and matched across subjects, expansions or contractions may be needed due to individual differences in brain structure. To account for the amount of expansion and contraction, the modulation step adjusts the normalized grey matter values by multiplying by its relative volume before and after spatial normalization (e.g., if a participant's temporal lobe doubles in volume during normalization, the correction will halve the intensity of the signal in this region; Mechelli et al. 2005). The resulting values at each voxel thus denote the absolute amount of tissue that is grey matter at that location, after having adjusted for the confounding effects of nonlinear warping. While an analysis based on modulated data (implemented in the current study) tests for variability in the amount of grey matter, an analysis without modulation tests for variability in *concentration* of grey matter (Ashburner and Friston 2000; Mechelli et al. 2005). Finally, the images were transformed to Montreal Neurological Institute (MNI) stereotactic space (voxel size = 1.5 mm^3), and smoothed using a 10 mm full-width half-maximum (FWHM) isotropic Gaussian kernel. VBM provides a mixed measure of cortical surface (or cortical folding) as well as cortical thickness, unlike surface-based approaches, that emphasize measures of thickness derived from geometric models of the cortical surface (e.g., Huton et al. 2009). Further work is needed to specify the exact cellular basis of local differences in the amount of grey matter as measured by VBM. However, these are assumed to potentially reflect variability in the number and size of neurons or glia, or in axonal architecture (May and Gaser 2006; Kanai and Rees 2011).

Multiple regressions were conducted on the smoothed grey matter images. At the whole brain level, per-participant auditory imagery scores were entered into a general linear model, including age, gender, total grey matter volume (Peelle et al. 2012), short-term memory and years of musical training as nuisance variables in the design matrix to regress out any potential confounding effects related to them. Musical training was included because this has been shown to correlate with vividness of auditory imagery (Pfordresher and Halpern 2013), with the acuity of mental auditory images in performance-based tasks (Janata and Paroo 2006; Navaro-Cebrian and Janata 2010a; Navarro-Cebrian and Janata 2010b), as well as with differences in brain structure (Gaser and Schlaug 2003). Regressing out variability in short-term memory is important to ensure that correlations between imagery and grey matter cannot be attributed to nonspecific factors linked to general cognitive functioning. While a memory component may be involved in imagery (e.g., Navarro-Cebrian and Janata 2010b), the need to control for the general cognitive demands of the tasks has been highlighted (Halpern et al. 2004; Zatorre and Halpern 2005), and this is of special relevance in the context of an off-line self-report measure as the one used here. Any voxels showing grey matter intensity < .05 were excluded using an absolute masking threshold to avoid possible edge effects around the border between grey matter and white matter. Statistical maps were thresholded at $p < .005$ peak level uncorrected, cluster corrected with a Family Wise Error (FWE) correction at $p < .05$, whilst accounting for non-stationary correction (Ridgway et al. 2008). In addition to whole-brain analysis, more sensitive region of interest (ROI) analyses were conducted within regions for which we had *a priori* hypotheses, based on a recent Activation Likelihood Estimation meta-analysis of fMRI studies of imagery across modalities (McNorgan 2012). We covered two networks identified by this meta-analysis, one derived from auditory imagery studies only (8 studies), and the other one from studies involving imagery across multiple modalities (65 studies). When a region was reported in both networks, we choose the coordinate of the auditory-specific one. Table 1 presents the list of ROIs and corresponding MNI coordinates. Statistical significance within these ROIs was assessed using small volume correction (Worsley et al. 1996) at a threshold of $p < .05$ (FWE corrected), within spheres with 12 mm radius centered at each of the coordinates.

fMRI procedure and analyses

Functional and structural data were acquired on the same day. Participants were told that they would hear different kinds of sounds, and that they should listen attentively to them. They listened passively to the sounds and were asked to perform a vigilance task consisting of pressing a button every time a “beep” was presented. The sounds were presented in 2 runs of 140 echo-planar whole-brain volumes; each run lasted 21 minutes. The first three volumes from each run were discarded to allow longitudinal magnetization to reach equilibrium. Auditory onsets occurred 5.5 s (± 0.5 s jitter) before the beginning of the following whole-brain volume acquisition. On each trial, participants listened to two randomly selected sounds of the same type. The sounds were presented in a pseudo-randomized order for each participant, and we ensured that no more than three trials of the same type were consecutively presented. All 120 sounds were presented twice per run (plus 9 vigilance and 8 rest/silence trials per run). Sounds were presented using Psychtoolbox (Brainard 1997) via a Sony STR-DH510 digital AV control center (Sony, Basingstoke, UK) and MRI-compatible insert earphones (Sensimetrics Corporation, Malden, MA, EUA).

Data were analyzed using SPM8. Functional images were realigned to the first image, unwarped, coregistered to the structural image, and spatially normalized to MNI space using the parameters acquired from segmentation (Ashburner and Friston 2005); they were resampled to 2 mm^3 voxels and smoothed with a 10 mm FWHM Gaussian kernel. The hemodynamic response was modeled using a first-order finite impulse response (FIR) filter with a window length equal to the time taken to acquire a single volume. At the first level, the 5 types of vocal sounds, the unintelligible rotated sounds, and the vigilance trials (and 6 movement regressors of no interest) were entered into a general linear model. The rest/silence trials were used as an implicit baseline. At the second level, a one-way repeated-measures ANOVA was conducted using contrast images from the first level to identify brain regions in which the magnitude of responses varied as a function of the type of human vocalization; separate contrast images for each of the 5 types of intelligible sounds versus rest baseline were entered in this model (for a similar approach, Warren et al. 2006). The results are presented at an uncorrected threshold of $p < .005$ peak level, with non-stationary correction of $p < .05$ at cluster level for the whole-brain analysis.

To examine whether the neural systems involved in imagery co-localize with those involved in auditory processing, ROI analyses were conducted focusing on the regions shown to predict auditory imagery in the VBM study (at whole-brain and ROI levels); small volume correction was used at a threshold of $p_{\mathrm{FWE}} < .05$, within spheres with 12 mm radius, centered at the peak of the clusters. Among these ROIs, when the one-way repeated-measures ANOVA revealed an effect, a more sensitive multivariate Representational Similarity Analysis was also conducted (Kriegeskorte et al. 2008). This analysis was conducted to directly explore whether there is an association between inter-individual differences in the specificity of neural representations of heard vocal sounds and variation in self-report auditory imagery ratings. This was achieved by extracting data from the whole brain t-statistic maps of each of the 5 types of intelligible vocal sounds relative to the resting baseline, and Pearson product-moment correlating these maps with each other. We used T-maps because, as they combine the effect size weighted by error variance for a modeled response, they provide higher classification accuracy in multivariate analyses; results are not unduly influenced by large, but highly variable response estimates (Misaki et al. 2010). In each participant, the correlation coefficients reflecting the relationship between neural responses to each of the 5 conditions with every other condition were converted to a *z* value using a Fisher transformation so as to conform to statistical assumptions (normality) required for parametric statistical tests. These values were averaged to provide a summary statistic for each participant, a higher value reflecting higher similarity between neural responses, i.e., lower discrimination between conditions; and a lower value reflecting lower similarity between neural responses, i.e., high discrimination between conditions or more distinct representations. These values were then Pearson product-moment correlated with ratings of auditory imagery.

Results

Neuroanatomical predictors of individual differences in auditory imagery

There were large individual differences in auditory imagery ratings: For the total imagery scale, ratings ranged between 2.5 and 7 ($M = 5.12$; $SD = 0.87$); on the Vividness subscale, they ranged between 2.86 and 7 ($M = 4.96$; $SD = 0.95$); and on the Control subscale, they ranged between 2 and 7 ($M = 5.28$; $SD = 0.95$). Consistent with previous evidence (Pfordresher and Halpern 2013), vividness and control of imagery were highly correlated with each other ($r = .68$, $p < .001$). No significant associations were found between imagery and age (total imagery scale, $r = -.18$, $p = .13$; vividness subscale, $r = -.14$, $p = .25$;

control subscale, $r = -.19$, $p = .11$), suggesting that these processes are relatively stable across the adult life span. Shapiro-Wilk tests confirmed that the ratings were normally distributed (*p*s > .13).

The goal of this experiment was to determine whether individual differences in how people perceive auditory images can be predicted from differences in brain morphology. A first whole-brain analysis focusing on the total imagery ratings (average of the two scales) revealed that higher ratings correlated with larger grey matter volume in a cluster with a peak voxel in the left paracentral lobule, extending to the right paracentral lobule, left precuneus, and left superior frontal gyrus (cluster size = 3369 voxels, $p_{\mathrm{FWE}} = .03$; MNI coordinate for peak voxel: $x = -8$, $y = -12$, $z = 69$, $t_{(1,67)} = 3.63$, Z = 3.45, $p < .001$ uncorrected). No associations were found between higher imagery ratings and decreased grey matter (for the negative contrast, lowest $p_{\mathrm{FWE}} = .43$). To directly investigate the structural predictors of each of the two auditory imagery components, whole-brain analyses were also conducted on vividness and control ratings separately (we refrained from including the two subscales in the same design matrix because they were very highly correlated with each other). For individual differences in control of imagery, no clusters survived correction, either for positive or for negative correlations (lowest $p_{\mathrm{FWE}} = .26$). For vividness of imagery, on the other hand, a positive correlation was found with regional grey matter volume in a cluster with a peak voxel situated within the left SMA, extending to the left and right paracentral lobules (cluster size = 3531 voxels, $p_{\mathrm{FWE}} = .03$; MNI coordinate for peak voxel: $x = -6$, $y = -13$, $z = 67$, $t_{(1,67)} = 3.57$, Z = 3.40, $p < .001$). This cluster is shown in Figure 1, along with a scatterplot between grey matter residuals and vividness scores ($r = .46$, $p < .001$). No results were found for negative correlations (lowest $p_{\mathrm{FWE}} = .84$). We extracted grey matter residuals within this SMA cluster and observed that the correlation with higher vividness of imagery remained significant after regressing out variability accounted for by the other subscale, control of imagery (partial correlation, $r = .34$, $p = .003$). This indicates that the role of this structure for vividness of imagery cannot be reduced to nonspecific factors (e.g., confidence of participants in their judgments or temporal processing), as these would be similarly engaged across subscales.

ROI analyses, using small volume correction, were also conducted within regions hypothesized to be involved in auditory and domain general imagery generation, as identified by a recent meta-analysis of fMRI studies of imagery (McNorgan 2012). We found positive correlations between grey matter volume and vividness of auditory imagery within five ROIs. Two of them are part of the auditory imagery network, and they partly overlap with the SMA cluster revealed by the more conservative whole-brain analysis (see Table 1 for full anatomical and statistical details). The other three are part of the general imagery network: one in left inferior parietal lobule, one in right medial superior frontal gyrus, and one in left middle frontal gyrus. Additionally, a negative correlation was found between vividness of auditory imagery and the amount of grey matter in the right superior parietal lobule. Similar analyses focusing on control of imagery ratings revealed a marginally significant association between higher control and increased grey matter volume within the left SMA ROI (MNI coordinate for peak voxel within ROI: x = -11, y = -9, z = 72, $t_{(1,67)}$ = 3.11, Z = 3, p_{FWE} = .05), and a negative association in the right medial superior frontal gyrus ROI (MNI coordinate for peak voxel within ROI: x = 6, y = 15, z = 33, $t_{(1,67)}$ = 3.12, Z = 3.01, p_{FWE} = .05).

Functional responses to heard auditory information

In the whole-brain analysis, significant modulations of neural responses as a function of sound type were found in a number of brain regions, shown in Figure 2 and listed in Table 2. Consistent with earlier work using similar stimuli (e.g., Warren et al., 2006; McGettigan et al. 2015), activations were largely bilateral and included the STG, precentral and prefrontal cortices, parietal regions, cuneus and precuneus, insula and thalamus.

To assess whether regions involved in auditory imagery co-localized with those involved in the processing of heard auditory information, analyses were conducted looking at hemodynamic responses within the clusters in which grey matter volume correlated with vividness imagery ratings in the main VBM study. Using small volume correction, we found that the left SMA (cluster presented in Figure 1) shows significant modulation of the neural response as a function of sound type (MNI coordinate for peak voxel: $x = -8$, $y = -2$, $z = 62$, $F_{(4,220)} = 5.51$, $Z = 3.43$, $p_{\text{FWE}} = .03$), suggesting that this region plays a role in imagery and in the processing of heard information. Crucially, we additionally conducted a representational similarity analysis (see Materials and Methods) to examine whether this co-localization in SMA reflects the operation of converging mechanisms. Activity patterns associated with each pair of intelligible vocal sound types were compared (linear correlations, $n = 10$), the pairs were assembled, and an average similarity was computed for each participant ($M = .83$; $SD = .1$; range = .47 – .97); this analysis was conducted within a sphere with 12 mm radius (925 voxels). In keeping with the hypothesis that mechanisms are shared, lower neural similarity between vocal sounds correlated with higher vividness of auditory imagery, i.e., participants with higher specificity of neural representations during the processing of heard auditory information also reported experiencing more vivid mental auditory images ($r = -.34$, $p = .01$; after regressing out demographic and cognitive variables, as in the main VBM study, $r = -.42$, $p = .001$). This association is shown in Figure 3. A further model was conducted to examine whether the magnitude of the distinction between intelligible vocal sounds and the condition of unintelligible sounds was also associated with imagery. We computed an average of similarity of neural responses between each type of vocal sound and rotated sounds for each participant (linear correlations, $n = 5$; neutral sounds vs. rotations, laughter vs. rotations, etc.), and found a significant correlation between lower similarity and higher vividness of auditory imagery ($r = -.42$, $p = .001$; after regressing out demographic and cognitive variables, $r = -.50$, $p < .001$). This finding suggests that participants reporting higher vividness of mental auditory images not only show higher representational specificity of different intelligible vocal sounds, as they also appear to show sharper distinctions between vocal and unintelligible sounds within SMA.

Perceptual-functional modulations as a function of sound type were also found in three of the clusters selected from the imagery meta-analysis (and in which the amount of grey matter predicted vividness ratings in the current study; see Table 1): one in left SMA as well (MNI coordinate for peak voxel: $x = -8$, $y = 0$, $z = 60$, $F_{(4,220)} = 5.52$, Z = 3.43, $p_{\mathrm{FWE}} = .03$), one in the left inferior parietal lobule (MNI coordinate for peak voxel: $x = -32$, $y = -48$, $z = 44$, $F_{(4,220)} = 8$, Z = 4.42, $p_{\mathrm{FWE}} < .001$), and one in the right superior parietal lobule (MNI coordinate for peak voxel: $x = 16$, $y = -50$, $z = 50$, $F_{(4,220)} = 7.03$, Z = 4.07, $p_{\mathrm{FWE}} < .004$). Representational similarity analyses were also conducted for these clusters. Correlations between representational similarity and vividness of imagery approached significance for the left SMA cluster ($r = -.23$, $p = .09$; after regressing out demographic and cognitive variables, $r = -.33$, $p = .01$), but they were non-significant for the left inferior parietal ($r = -.10$, $p = .48$; after regressing out demographic and cognitive variables, $r = -.10$, $p = .45$) and right superior parietal ones ($r = -.12$, $p = .39$; after regressing out demographic and cognitive variables, $r = -.09$, $p = .5$).

These results suggest that brain regions whose structure predicts individual differences in auditory imagery, notably the SMA and parietal systems, are also engaged by processing of auditory information. A direct association between imagery and sensory-based processing could however be established for the SMA only.

Links between auditory and visual imagery

From the results described so far, it cannot be determined whether the underlying mechanisms are specialized for auditory information or whether they are supra-modal in nature to some extent. To shed light on this question, we investigated behavioral and neural correlations between auditory and visual imagery. Considerable individual differences were obtained in visual imagery ratings (VVIQ): ratings ranged between 1.19 and 5 (5 = maximally vivid; $M = 3.63$; $SD = 0.81$). A strong behavioral correlation was found between reported vividness of auditory and visual imagery ($r = .57$, $p < .001$; see Figure 4), a correlation that remains significant after regressing out demographic and cognitive variables ($r = .53$, $p < .001$). This indicates that participants who report generating highly vivid auditory images also report generating highly vivid visual images. Additionally, higher vividness of visual imagery correlated with grey matter volume within the SMA cluster previously shown to correlate with vividness of auditory imagery (in the whole-brain VBM analysis, Figure 1; MNI coordinate for peak voxel: $x = 4$, $y = -12$, $z = 72$, $t_{(1,39)} = 3.25$, Z = 3.04, $p_{\mathrm{FWE}} = .048$). To investigate whether this association reflects unique variance associated with visual imagery (i.e., independent of auditory imagery), we correlated grey matter residuals with visual imagery while regressing out variability in vividness of auditory imagery; the partial correlation coefficient was not significant ($r = .03$, $p = .82$). No other associations between grey matter and visual imagery were found, both in whole-brain analysis and after small volume corrections within other regions implicated in imagery (Table 1).

Discussion

The present study examined the structural basis of inter-individual differences in auditory imagery, and how these differences reflect commonalities in sensory-based processing and mechanisms that are involved in imagery across modalities. We present four novel findings. First, using VBM, we established that differences among individuals in the reported vividness of auditory imagery are predicted by the amount of grey matter in the SMA, inferior and superior parietal lobules, medial superior frontal gyrus and middle frontal gyrus. Second, in an fMRI experiment, these SMA, inferior and superior parietal sites were also modulated as a function of vocalization type during the processing of heard auditory information. Third, a representational similarity analysis revealed that higher representational specificity of different types of vocal sounds within SMA predicts higher vividness of mental auditory images, a result that directly links sensory- and imagery-based processing. Fourth, cross-modal interactions were found at behavioral and structural levels: self-report behavioral measures of auditory and visual imagery were correlated, and individual differences in visual imagery were also predicted by the amount of grey matter in SMA. These findings are discussed in the next paragraphs.

Although a number of studies have shown that temporal, parietal, motor and prefrontal regions are typically active during auditory imagery tasks (e.g., Shergill et al. 2001; Herholz et al. 2012; Zvyagintsev et al. 2013), relatively little was known about which of these systems (and how) predict variability in behavioral correlates of imagery (Daselaar et al. 2010; Zatorre et al. 2010; Herholz et al. 2012). Consistent with previous performance-based (Janata 2012) and self-report evidence (Pfordresher and Halpern 2013; Gelding et al. 2015), our behavioral measure revealed that auditory imagery varies considerably across individuals. Crucially, here we show for the first time that this variability relates to differences in the local structure of grey matter. The association between higher perceived vividness of auditory images and increased grey matter volume in SMA adds to functional research reporting activity in this region during auditory imagery tasks requiring the imagination of tunes (Herholz et al. 2012; Zvyagintsev et al. 2013), timbre of musical instruments (Halpern et al. 2004), verbal information (Shergill et al. 2001; Linden et al. 2011), and anticipating sound sequences (Leaver et al. 2009). It is also in accord with exploratory results showing a correlation between the magnitude of BOLD responses in SMA and higher vividness ratings during a task involving imagery of familiar melodies (Zvyagintsev et al. 2013). The other regions in which the amount of grey matter predicted vividness of imagery, namely left inferior and right superior parietal cortices, right medial superior frontal gyrus, and left middle frontal gyrus, were recently highlighted by a meta-analysis as part of a core imagery network (McNorgan 2012), and they have been shown to be engaged across different kinds of auditory imagery tasks (Shergill et al. 2001; Zatorre et al. 2010; Linden et al. 2011; Zvyagintsev et al. 2013).

Extending previous findings, the present study demonstrates not only that these systems are functionally implicated in imagery, but also that their structural features are diagnostic of behavioral outcomes. Our results were obtained using an off-line self-report measure that covers ecologically valid and diverse scenarios, which was completed in comfortable conditions, i.e., not constrained by being inside an MRI scanner. Importantly, this measure has been shown to index mechanisms that are also involved in active, performance-based, imagery tasks. It correlates with brain activity during active imagery tasks (reversal of melodies, Zatorre et al. 2010; imagery of familiar tunes, Herholz et al. 2012), and with performance levels in behavioral tasks: pitch discrimination (Pfordresher and Halpern 2013), and detection of mismatches between a probe note and the last note of an imagined sequence (Gelding et al. 2015). This adds to the mounting evidence that self-report measures provide rich information about individual differences in an array of cognitive processes, and can significantly relate to brain structure (Kanai et al. 2011; Banissy et al. 2012). For instance, Kanai et al. (2011) observed that a self-report measure of everyday distractibility correlates with grey matter volume in the left superior parietal cortex, as well as with a performance-based measure of attention capture. Because of the characteristics of these measures, however, one concern regards the potential confounding effects of participants' abilities to report on their own experience (metacognition), or of their general cognitive ability (e.g., working memory; attention). Our results are unlikely to be reducible to such processes: we controlled for performance on a short-term memory task that correlates with working memory and intelligence (Colom et al. 2008), and we showed that associations between vividness and brain structure remain significant after accounting for responses on the other BAIS subscale focusing on control of imagery, which would load on the same nonspecific metacognitive factors. Moreover, the ability to introspect about self-performance correlates with grey matter volume in the right anterior prefrontal cortex (Fleming et al. 2010), a region involved in high-level control of cognition and in the integration of perceptual information with decision output. This region does not overlap with those identified here.

It was unexpected that we did not find an association between auditory imagery and the structure of STG, even after small volume correction. Auditory association areas were previously found to be more strongly activated during auditory versus others forms of imagery (Zvyagintsev et al. 2013), and they have been assumed to support the reconstruction of auditory-like representations (Janata 2001; Kraemer et al. 2005; Lange 2009; Navarro-Cebrian and Janata 2010a). It was further reported that the magnitude of BOLD responses within these areas predicts vividness ratings during imagery (Daselaar et al. 2010; Herholz et al. 2012; Zvyagintsev et al. 2013), even though this finding is not always replicated (Leaver et al. 2009). Our null result does not weaken the well-established idea that STG plays a functional role for auditory imagery, but it suggests that macroscopic grey matter differences in this region are not a source of inter-individual variability in the behavioral measure used here. This may indicate that anterior control and sensorimotor systems have a more prominent role than posterior auditory ones for individual differences in imagery, or that the structural predictors partly depend on the specific task demands. Indeed, there is fMRI and electrophysiological evidence that activity in auditory association areas is preceded and controlled by more anterior regions during imagery. Herholz et al. (2012) found increased connectivity between STG and prefrontal areas for imagery versus perception of tunes. Linden et al. (2011) showed that activity in SMA precedes that of auditory areas during voluntary imagery, and that this timing is impaired during hallucinations (lack of voluntary control). In the visual domain, Borst et al. (2011) showed that activity in frontal regions precedes that of more posterior regions, namely of occipital cortex, in a scene imagery task. In addition to being activated first, responses in frontal regions also predicted reaction times on the scene imagery task (consisting of judging whether a visually presented fragment of the scene was mirrored or not), while other regions did not. Concerning possible task effects, the self-report measure used here focuses on perceived vividness and on the sense of control over auditory images; it remains to be seen whether individual differences in performance-based imagery tasks requiring a fine-grained analysis of sound representations would reveal a structural role of STG (e.g., judging whether a probe note is mistuned or not, Janata & Paroo 2006; Navarro-Cebrian and Janata 2010a; Navarro-Cebrian and Janata 2010b).

The amount of grey matter in SMA was the most robust predictor of vividness of auditory imagery, an effect found both in whole-brain analysis and in the ROI analyses based on the meta-analysis of functional studies on imagery (McNorgan, 2012). Supporting the hypothesis that imagery partly engages the network that responds to heard auditory information, we also observed that this region was modulated by vocal sound category in the fMRI study, along with other regions that are typically engaged by intelligible vocal information, such as bilateral STG (e.g., Warren et al. 2006; Okada et al. 2010; Evans and Kyong et al. 2014; McGettigan et al. 2015). Our functional results are consistent with previous work reporting the engagement of motor systems during the processing of vocal information (Warren et al. 2006; McGettigan et al. 2015). We focus on vocalizations only, but these systems seem to be recruited by complex sounds more generally (Scott, McGettigan and Eisner 2009), such as music (Zatorre et al. 2007; Herholz et al. 2012), degraded speech (Mattys et al. 2012), and sounds derived from human actions like kissing or opening a zipper (Gazzola et al. 2006). Regarding the links between imagined and heard information, although previous studies observed common activations in SMA using linguistic and musical stimuli (Zatorre et al. 1996; Herholz et al. 2012), here we went a step further: we show co-localization across structural and functional levels and, crucially, we provide the first evidence for co-variation between vividness of auditory imagery and specificity of neural representations of heard auditory information within this region. Such an association is central to the argument that co-localization reflects the operation of similar mechanisms.

The SMA provides a crucial link between perception and action, and its functional attributes facilitate many cognitive and motor processes. It is involved in aspects of action including planning, initiation and inhibition, in learning new associations between stimuli and motor responses, in cognitive control processes such as switching between motor plans, and in the passive observation of grasping actions and emotional expressions (Warren et al. 2006; Kleber et al. 2007; Nachev et al. 2008; Mukamel et al. 2010). Mukamel et al. (2010) recorded single-neuron responses in humans during the observation and execution of grasping actions and facial gestures, and found that a significant number of neurons in SMA responded to both conditions, revealing sensorimotor properties. As for the structure of SMA, previous studies demonstrated that it may vary across individuals as a function of motor learning and expertise: there is longitudinal evidence of increments in the volume of grey matter during six weeks of learning of a complex motor task (Taubert et al. 2010), as well as cross-sectional evidence of expertise-related structural differences in gymnasts (Huang et al. 2013) and ballet dancers (Hänggi et al. 2010). That sensorimotor systems respond to different types of complex auditory information, even when participants are not required to perform or plan any movements, may reflect the automatic engagement of some level of sensorimotor simulation. Processing and evaluating complex sounds – human vocalizations, in the case of the current study – would involve the activation of motor representations that link sensory information to actions related to the production of those sounds (Gazzola et al. 2006; Warren et al. 2006; Scott, McGettigan and Eisner 2009; Scott et al. 2014; McGettigan et al. 2015). We argue that the same mechanism of covert simulation may support auditory imagery – an imagery-to-action pathway. Accessing auditory-motor representations may be central for the generation of different types of mental auditory images, such as vocal and musical ones (Halpern and

Zatorre 1999; Meteyard et al. 2012; Zvyagintsev et al. 2013), and the structure of sensorimotor systems may be a determinant of the efficiency of this mechanism. The perceived vividness of mental images and the representational specificity of heard information would both be shaped by how efficiently relevant sensorimotor information is retrieved.

Such an imagery-to-action pathway is unlikely to be specialized to auditory information, as other forms of imagery (e.g., visual, motor) may also have associated action components and engage sensorimotor processes to some extent. Indeed, activity in SMA is observed in functional studies conducted on non-auditory modalities of imagery (Guillot et al. 2009; Borst et al. 2012; McNorgan 2012; Hétu et al. 2013; Zvyagintsev et al. 2013). Furthermore, SMA is similarly active during motor imagery and execution, suggesting that movement sensations are simulated during motor imagery (Naito et al. 2002; Ehrsson et al. 2003; Hanakawa et al. 2003). The same was suggested in the visual domain (Grèzes and Decety 2002; Solodkin et al. 2004; Zacks 2008; Mukamel et al. 2010). However, despite the suggestive evidence of cross-modal commonalities in the mechanisms supporting imagery, only rarely have different modalities been directly compared (Halpern et al. 2004; Solodkin et al. 2004; Daselaar et al. 2010). We established that participants reporting highly vivid auditory images also report experiencing highly vivid visual images. That vividness of visual imagery is reflected in differences in grey matter volume in SMA, paralleling the findings for auditory imagery, suggests that converging sensorimotor simulation processes may operate across modalities. These commonalities may further reflect the fact that everyday imagery often involves multisensory components, i.e., mental images are frequently not confined to one single modality (Hubbard 2013). Even in an experimental setting in which the task requires participants to focus on a particular modality, components from other modalities may be spontaneously retrieved. When asked to generate an image of an auditory scene, for instance, concurrent visual and kinesthetic images might spontaneously appear (e.g., when imagining the cheer of the crowd as a player hits the ball – one of the BAIS items – individuals may also generate a visual image of the crowd in a stadium). In future studies it would be interesting to examine whether the diversity of components retrieved for an auditory or visual scene may actually contribute to enhance the impression of vividness.

To conclude, the present study forms the first demonstration that inter-individual differences in auditory imagery have a signature in brain structure, adding to the growing body of evidence that individual differences can be an invaluable source of information to link behavior and cognition to brain anatomy. Building upon prior functional neuroimaging studies, our results establish a role for the structure of parietal, prefrontal and sensorimotor systems (in particular SMA) in supporting auditory imagery. In SMA, we further established links between auditory imagery, processing of heard vocal information, and visual imagery. We argue for sensorimotor simulation as a candidate mechanism for such commonalities. Future investigations could extend this work to refine the exploration of converging computations between imagery and auditory processing, e.g., by including different types of perceived and imagined sounds that afford a wider range of variability in terms of the accessibility of relevant sensorimotor representations. Our focus was on links between heard human vocal information and auditory imagery mostly for voices and music (the main domains covered by the BAIS). Further work will also need to specify the microstructural basis of the large-scale anatomical differences reported here, and to determine how they are shaped by environmental and genetic factors.

Funding: This work was supported by a Wellcome Trust Senior Research Fellowship (WT090961MA) awarded to Sophie Scott. During the planning and execution of this project, César Lima was funded by a postdoctoral fellowship from the Portuguese Foundation for Science and Technology (SFRH/BPD/77189/2011), and Andrea Halpern was partly funded by a Leverhulme Visiting Professorship.

Acknowledgements: We thank the staff at the Birkbeck-UCL Centre for NeuroImaging (BUCNI) for technical advice.

References

Ashburner J. 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38:95-113.

Ashburner J, Friston KJ. 2000. Voxel-based morphometry – The methods. NeuroImage 11:805-821.

Ashburner J, Friston KJ. 2005. Unified segmentation. NeuroImage 26:839-851.

Banissy MJ, Kanai R, Walsh V, Rees G. 2012. Inter-individual differences in empathy are reflected in human brain structure. NeuroImage 62:2034-2039.

Blesser B. 1972. Speech perception under conditions of spectral transformation. J Speech Hear Res 15:5-41.

Borst AW, Sack AT, Jansma BM, Esposito F, Martino F, Valente G, Roebroeck A, Salle F, Goebel R, Formisano E. 2012. Integration of "what" and "where" in frontal cortex during visual imagery of scenes. Neuroimage 60:47-58.

Brainard DH. 1997. The Psychophysics Toolbox. Spat Vis 10:433-436.

Burianová H, Marstaller L, Sowman P, Tesan G, Rich AN, Williams M, Savage G, Johnson BW. 2013. Multimodal functional imaging of motor imagery using a novel paradigm. NeuroImage 71:50-58.

Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox RHB, Hetu RKJ, Liu C, Kiessling J, Kotby MN, Nasser NHA, Elkholy WAH, Nakanishi YOH, Powell R, Stephens D, Meredith R, Sirimanna T, Tavartkiladze GF, Westerman S, Ludvigsen C. 1994. An international comparison of long-term average speech spectra. J Acoust Soc Am 96:2108-2120.

Campos A, Pérez-Fabello MJ. 2009. Psychometric quality of a revised version of the Vividness of Visual Imagery Questionnaire. Percept Mot Skills 108:798-802.

Colom R, Abad FJ, Quiroga MA, Shih PC, Flores-Mendoza C. 2008. Working memory and intelligence are highly related constructs, but why? Intelligence 36:584-606.

Cui X, Jeter CB, Yang D, Montague PR, Eagleman DM. 2007. Vividness of mental imagery: Individual variability can be measured objectively. Vision Res 47:474-478.

Daselaar SM, Porat Y, Huijberts W, Pennartz CMA. 2010. Modality-specific and modality-independent components of the human imagery system. NeuroImage 52:677-685.

Erb J, Henry MJ, Eisner F, Obleser J. 2012. Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. Neuropsychologia 50:2154-2164.

Evans S, Kyong JS, Rosen S, Golestani N, Warren JE, McGettigan C, Mourão-Miranda J, Wise RJS, Scott SK. 2014. The pathways for intelligible speech: Multivariate and univariate perspectives. Cereb Cortex 24:2350-2361.

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. 2010. Relating introspective accuracy to individual differences in brain structure. Science 329:1541-1543.

Green T, Rosen S, Faulkner A, Paterson R. 2013. Adaptation to spectrally-rotated speech. J Acoust Soc Am 134:1369-1377.

Guillot A, Collet C, Nguyen VA, Malouin F, Richards C, Doyon J. 2009. Brain activity during visual versus kinesthetic imagery: An fMRI study. Hum Brain Mapp 30:2157-2172.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. 1999. "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213-223.

Halpern AR. In press. Differences in auditory imagery self-report predict neural and behavioral outcomes. Psychomusicology: Music, Mind, and Brain.

Halpern AR, Zatorre RJ. 1999. When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. Cereb Cortex 9:697-704.

Halpern AR, Zatorre RJ, Bouffard M, Johnson JA. 2004. Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia 42:1281-1292.

Hanakawa T, Immisch I, Toma K, Dimyan MA, Gelderen P, Hallett M. 2003. Functional properties of brain areas associated with motor execution and imagery. J Neurophysiol 89:989-1002.

Hänggi J, Koeneke S, Bezzola L, Jäncke L. 2010. Structural neuroplasticity in the sensorimotor network of professional female ballet dancers. Hum Brain Mapp 31:1196-1206.

Herholz SC, Halpern AR, Zatorre RJ. 2012. Neuronal correlates of perception, imagery, and memory for familiar tunes. J Cogn Neurosci 24:1382-1397.

Ehrsson HH, Geyer S, Naito E. 2003. Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. J Neurophysiol 90:3304-3316.

Gaser C, Schlaug G. 2003. Brain structures differ between musicians and non-musicians. J Neurosci 23:9240-9245.

Gazzola V, Ziz-Zadeh L, Keysers C. 2006. Empathy and the somatotopic auditory mirror system in humans. Curr Biol 16:1824-1829.

Gelding RW, Thompson WF, Johnson BW. 2015. The pitch imagery arrow task: Effects of musical training, vividness, and mental control. PLoS One [accepted 23rd February 2015].

Golestani N, Price C, Scott SK. 2011. Born with an ear for dialects? Structural plasticity in the expert phonetician brain. J Neurosci 31:4213-4220.

Grèzes J, Decety J. 2002. Does visual perception of object afford action? Evidence from a neuroimaging study. Neuropsychologia 40:212-222.

Hétu S, Grégoire M, Saimpont A, Coll M, Eugène F, Michon P, Jackson P. 2013. The neural network of motor imagery: An ALE meta-analysis. Neurosci Biobehav Rev 37:930-949.

Huang R, Lu M, Song Z, Wang J. 2013. Long-term intensive training induced brain structural changes in world class gymnasts. Brain Struc Func. Dec 3 [Epub ahead of print].

Hubbard TL. 2013. Auditory imagery contains more than audition. In: S Lacey, R Lawson, editors. Multisensory Imagery. New York: Springer. p 221-246.

Hutton C, Draganski B, Ashburner J, Weiskopf N. 2019. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. NeuroImage 48:371-380.

Janata P. 2001. Brain electrical activity evoked by mental formation of auditory expectations and images. Brain Topogr 13:169-193.

Janata P. 2012. Acuity of mental representations of pitch. Ann N Y Acad Sci 1252:214-221.

Janata P, Paroo K. 2006. Acuity of auditory images in pitch and time. Percept Psychophys 68:829-844.

Kanai R, Rees G. 2011. The structural basis of inter-individual differences in human behaviour and cognition. Nat Rev Neurosci 12:231-242.

Kanai R, Dong MY, Bahrami B, Rees G. 2011. Distractibility in daily life is reflected in the structure and function of human parietal cortex. J Neurosci 31:6620-6626.

Keller PE. 2012. Mental imagery in music performance: Underlying mechanisms and potential benefits. Ann N Y Acad Sci 1252:206-2013.

Kleber B, Birbaumer N, Veit R, Trevorrow T, Lotze M. 2007. Overt and imagined singing of an Italian area. NeuroImage 36:889-900.

Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Kosslyn SM, Ganis G, Thompson WL. 2001. Neural foundations of imagery. Nat Rev Neurosci 2:635-642.

Kozhevnikov M, Kosslyn S, Shephard J. 2005. Spatial versus object visualizers: A new characterization of visual cognitive style. Mem Cognition 33:710-726.

Kraemer DJM, Macrae CN, Green AR, Kelley WM. 2005. Sound of silence activates auditory cortex. Nature 434:158.

Lange K. 2009. Brain correlates of early auditory processing are attenuated by expectations for time and pitch. Brain Cogn 69:127-137.

Leaver AM, Van Lare J, Zielinski B, Halpern, AR, Rauschecker JP. 2009. Brain activation during anticipation of sound sequences. J Neurosci 29:2477-2485.

Lima CF, Castro SL, Scott SK. 2013. When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. Behav Res Meth 45:1234-1245.

Linden DEJ, Thornton K, Kuswanto CN, Johnston SJ, van de Ven V, Jackson MC. 2011. The brain's voices: Comparing nonclinical auditory hallucinations and imagery. Cereb Cortex 21:330-337.

Marks DF. 1973. Visual imagery diferences in the recall of pictures. Br J Psychol 64:17-24.

Mattys SL, David MH, Bradlow AR, Scott SK. 2012. Speech recognition in adverse conditions: A review. Lang Cognitive Proc 27:953-978.

May A, Gaser C. 2006. Magnetic resonance-based morphometry: A window into structural plasticity of the brain. Curr Opin Neurol 19:407-411.

Mechelli A, Price CJ, Friston KJ, Ashburner J. 2005. Voxel-based morphometry of the human brain: Methods and applications. Curr Med Imaging Rev 1:105-113.

Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G. 2012. Coming of age: A review of embodiment and the neuroscience of semantics. Cortex 48:788-804.

McGettigan C, Walsh E, Jessop R, Agnew ZK, Sauter DA, Warren JE, Scott SK. 2015. Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. Cereb Cortex 25:246-257.

McNorgan C. 2012. A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. Front Human Neurosci 6:285.

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. NeuroImage 53:103-118.

Mukamel R, Ekstrom AD, Kaplan J, Iacoboni M, Fried I. 2010. Single-neuron responses in humans during execution and observation of actions. Curr Biol 20:750-756.

Naito E, Kochiyama T, Kitada R, Nakamura S, Matsumura M, Yonekura Y, Sadato N. 2002. Internally simulated movement sensations during motor imagery activate cortical motor areas and the cerebellum. J Neurosci 22:3683-3691.

Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen SD, Matthews PM. 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362-1368.

Navarro-Cebrian A, Janata P. 2010a. Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. Brain Res 1342:39-54.

Navarro-Cebrian A, Janata P. 2010b. Influences of multiple memory systems on auditory mental image acuity. J Acoust Soc Am 127:3189-3202.

Novembre G, Ticini L, Schütz-Bosbach S, Keller P. 2014. Motor simulation and the coordination of self and other in real-time joint action. Soc Cogn Affect Neurosci 9:1062-1068.

Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb Cortex 20:2486-2495.

Peelle JE, Cusack R, Henson RNA. 2012. Adjusting for global effects in voxel-based morphometry: Grey matter decline in normal aging. Neuroimage 60:1503-1516.

Pfordresher PQ, Halpern AR. 2013. Auditory imagery and the poor-pitch singer. Psychon B Rev 20:747-753.

Preacher K, Hayes AF. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Meth 40:879-981.

Ridgway GR, Henley SMD, Rohrer JD, Scahill RI, Warren JD, Fox NC. 2008. Ten simple rules for reporting voxel-based morphometry studies. Neuroimage 40:1429-1435.

Sauter DA, Eisner F, Calder AJ, Scott SK. 2010. Perceptual cues in nonverbal vocal expressions of emotion. Q J Exp Psychol 63:2251-2272.

Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400-2406.

Scott SK, Lavan N, Chen S, McGettigan C. 2014. The social life of laughter. Trends Cogn Sci 18:618.620.

Scott SK, McGettigan C, Eisner F. 2009. A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. Nat Rev Neurosci 10:295-302.

Scott SK, Rosen S, Beaman CP, Davis JP, Wise RJ. 2009. The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. J Acoust Soc Am 125:1737-1743.

Shergill SS, Bullmore ET, Brammer MJ, Williams SCR, Murray RM, McGuire PK. 2001. A functional study of auditory verbal imagery. Psychol Med 31:241-253.

Solodkin A, Hlustik P, Chen EE, Small SL. 2004. Fine modulation in network activation during motor execution and motor imagery. Cereb Cortex 14:1246-1255.

Taubert M, Draganski B, Anwander A, Mueller K, Horstmann A, Villringer A, Ragert P. 2010. Dynamic properties of human brain structure: Learning-related changes in cortical areas and associated fiber connections. J Neurosci 30:11670-11677.

Warren JE, Sauter DA, Eisner F, Wiland J, Dresner MA, Wise RJS, Rosen S, Scott SK. 2006. Positive emotions preferentially engage an auditory-motor “mirror” system. J Neurosci 26:13067-13075.

Wechsler D. 1997. Wechsler Adult Intelligence Scale – Third Edition (WAIS-III). San Antonio: Pearson.

Woo C, Koban L, Kross E, Lindquist MA, Banich MT, Ruzic L, Andrews-Hanna JR, Wager TD. 2014. Separate neural representations for physical pain and social rejection. Nat Commun 5:5380.

Woollett K, Maguire EA. 2011. Acquiring “the Knowledge” of London’s layout drives structural brain changes. Curr Biol 21:2109-2114.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58-73.

Zacks JM. 2008. Neuroimaging studies of mental rotation: A meta-analysis and review. J Cogn Neurosci 20:1-19.

Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci 8:547-558.

Zatorre RJ, Halpern AR. 2005. Mental concerts: Musical imagery and auditory cortex. Neuron, 47:9-12.

Zatorre RJ, Halpern AR, Bouffard M. 2010. Mental reversal of imagined melodies: A role for the posterior parietal cortex. J Cogn Neurosci 22:775-789.

Zatorre RJ, Halpern AR, Perry DW, Meyer E, Evans AC. 1996. Hearing in the mind’s ear: A PET investigation of musical imagery and perception. J Cogn Neurosci 8:29-46.

Zvyagintsev M, Clemens B, Chechko N, Mathiak KA, Sack AT, Mathiak K. 2013. Brain networks underlying mental imagery of auditory and visual information. Eur J Neurosci 37: 1421-1434.

Table 1. VBM results for vividness of auditory imagery on regions previously identified to be functionally associated with auditory imagery and general imagery.

Region of interest	VBM results								
Area	MNI Coordinates			Peak Coordinates			Z score	$t_{(1,67)}$	p
	x	y	z	x	y	z			
Auditory Imagery Network									
R Superior Temporal Gyrus	64	-30	9						n.s.
L Inferior Frontal Gyrus	-48	24	-5						n.s.
	-51	17	9						n.s.
L Putamen	-21	-1	4						n.s.
L Superior Temporal Gyrus	-60	-38	15						n.s.
L Precentral Gyrus	-52	1	47						n.s.
L Supramarginal Gyrus	-58	-38	28						n.s.
R Inferior Frontal Gyrus	56	38	2						n.s.
L Supplementary Motor Area	-1	-14	53	-4	-24	52	3.22	3.36	.03
	-8	1	69	9	-9	73	3.26	3.40	.03
General Imagery Network									
L Inferior Parietal Lobule	-30	-56	52	-28	-55	43	3.2	3.34	.03
	-38	-38	46						
L Superior Parietal Lobule	-16	-62	54						n.s.
R Superior Parietal Lobule	20	-66	54	21	-61	51	3.27	3.41	.02
R Medial Superior Frontal Gyrus	6	20	44	14	17	48	3.47	3.65	.01
L Middle Frontal Gyrus	-30	0	56	-35	-7	63	2.98	3.10	.05

Note. The column “MNI Coordinates” shows the coordinates of ROIs, taken from a meta-analysis of imagery studies (McNorgan 2012); anatomical labels for each ROI were determined based on these coordinates, using the SPM Anatomy Toolbox v1.8. Small volume correction was used within 12-mm spheres centered at each of the coordinates. *p* values are FWE corrected (*p* < .05) and the obtained peak locations within each sphere are presented (column “Peak Coordinates”). R = right; L = left; n.s.: no local maxima exceeded the specified threshold.

Table 2. Brain regions showing significant modulations of BOLD responses as a function vocalization type during auditory processing.

	fMRI results						
Region	# Voxels	MNI Coordinates			Z score	$F_{(4,220)}$	*p*
		x	*y*	*z*			
R Superior Temporal Gyrus	10842	60	-24	8	> 8	72.85	< .001
R Superior Temporal Gyrus		62	-14	2	> 8	63.28	
R Primary Auditory Cortex		40	-26	12	> 8	55.64	
R Insula Lobe		34	24	4	5.96	13.16	
R Inferior Frontal Gyrus		44	16	28	5.72	12.25	
R Inferior Parietal Cortex		46	-36	48	3.77	6.31	
R Inferior Parietal Cortex		64	-32	42	3.67	6.05	
R Postcentral Gyrus		38	-36	50	3.65	6.01	
R Inferior Temporal Gyrus		52	-50	-8	3.49	5.64	
R SupraMarginal Gyrus		68	-30	34	3.48	5.62	
R Postcentral Gyrus		52	-22	48	3.45	5.56	
R Insula Lobe		42	14	-14	3.35	5.33	
R SupraMarginal Gyrus		32	-38	44	3.32	5.27	
R Postcentral Gyrus		38	-28	40	3.09	4.79	
R Precentral Gyrus		46	-14	56	2.77	4.18	
L Superior Temporal Gyrus	10449	-40	-32	12	Inf	71.04	< .001
L Insula Lobe		-32	26	6	6.62	15.93	
L Superior Temporal Gyrus		-52	2	-2	5.62	11.86	
L Inferior Frontal Gyrus		-34	6	26	4.59	8.49	

L Inferior Frontal Gyrus		-44	16	22	4.30	7.66	
L Inferior Frontal Gyrus		-48	10	16	4.01	6.89	
L Inferior Frontal Gyrus		-56	28	18	3.97	6.79	
L Inferior Frontal Gyrus		-40	8	16	3.91	6.64	
L Precentral Gyrus		-48	-4	48	3.86	6.50	
L Inferior Frontal Gyrus		-36	38	12	3.85	6.50	
L Precentral Gyrus		-46	4	32	3.62	5.93	
L Inferior Frontal Gyrus		-48	34	6	3.48	5.63	
L Precentral Gyrus		-48	0	40	3.29	5.21	
L Inferior Frontal Gyrus		-48	34	16	3.25	5.13	
L Middle Frontal Gyrus		-36	34	28	3.25	5.11	
L Cuneus	6227	-16	-56	22	4.79	9.08	< .001
L Precuneus		-14	-58	30	4.70	8.81	
L Middle Occipital Gyrus		-36	-74	30	4.70	8.81	
L Inferior Parietal Lobule		-30	-48	42	4.60	8.50	
L Superior Parietal Lobule		-22	-64	44	4.53	8.31	
L Middle Occipital Gyrus		-22	-62	34	4.29	7.64	
R Middle Occipital Gyrus		40	-70	30	4.29	7.64	
R Precuneus		6	-56	20	4.28	7.60	
R Angular Gyrus		50	-60	26	4.16	7.28	
L Inferior Parietal Lobule		-36	-40	40	4.11	7.16	
R Superior Parietal Lobule		16	-60	50	4.07	7.03	
L Inferior Parietal Lobule		-44	-40	42	4.06	7.02	

L Precuneus		-4	-60	20	4.00	6.88	
R Superior Parietal Lobule		26	-56	46	3.65	6.02	
L Cuneus		-8	-72	30	3.34	5.31	
Cerebellar Vermis	579	2	-38	-6	4.45	8.09	.01
R Thalamus		22	-18	-8	3.78	6.31	
R Thalamus		12	-26	-6	3.46	5.58	
R Thalamus		10	-10	2	3.34	5.32	
R Hippocampus		30	-18	-16	3.34	5.31	
L Posterior Cingulate Cortex		-8	-42	12	3.15	4.92	

Note. The results listed in the table (*F* contrast, one-way repeated-measures ANOVA) are presented at an uncorrected threshold of $p < .005$ peak level, corrected with non-stationary correction of $p < .05$ at cluster level. R = right; L = left. We report a maximum of 15 grey matter local maxima (that are more than 8 mm apart) per cluster.

Figure Captions

Figure 1. Association between grey matter volume and vividness of auditory imagery. A, cluster with peak in left SMA showing a significant positive correlation with vividness of auditory imagery in whole-brain analysis. Statistical maps were thresholded at *p* < .005 peak level uncorrected, cluster corrected with a Family Wise Error (FWE) correction at *p* < .05. B, scatterplot showing the association between vividness ratings and adjusted grey matter volume within the cluster depicted in A.

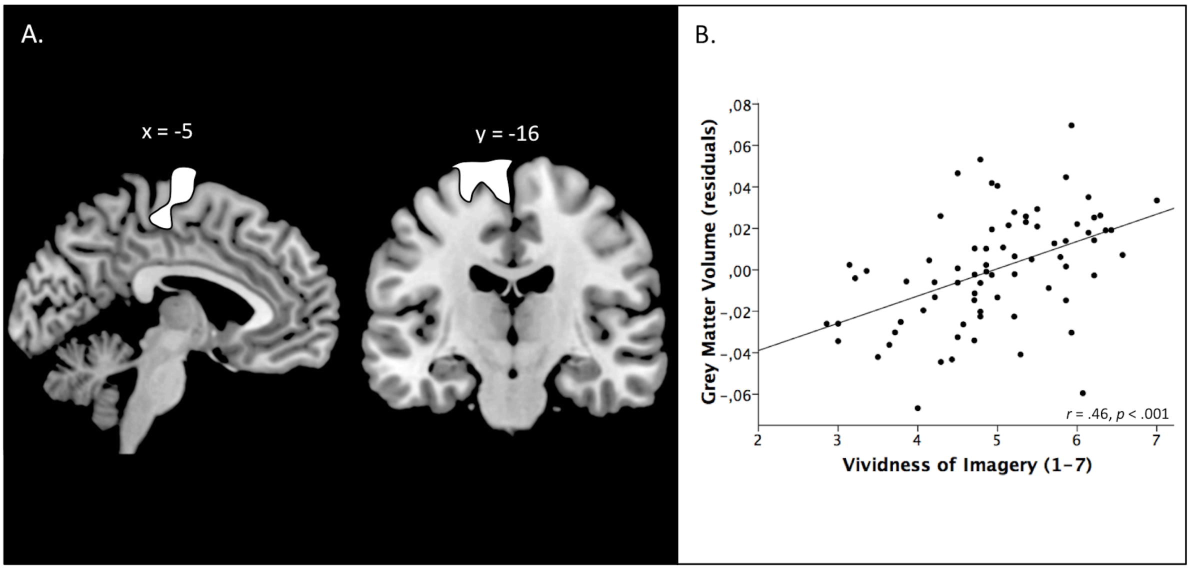


Figure 2. Brain regions in which BOLD responses were modulated by sound type during the processing of heard auditory information. The dotted dark red circle denotes a 12-mm sphere centered at the peak of the SMA cluster where the amount of grey matter was shown to correlate with auditory imagery (VBM study); this sphere was used for the representational similarity analysis looking at the links between representational specificity of heard sounds and vividness of imagery. For visualization purposes, activation maps were thresholded at *p* < .005 peak level uncorrected (full details of activated sites are presented in Table 2).

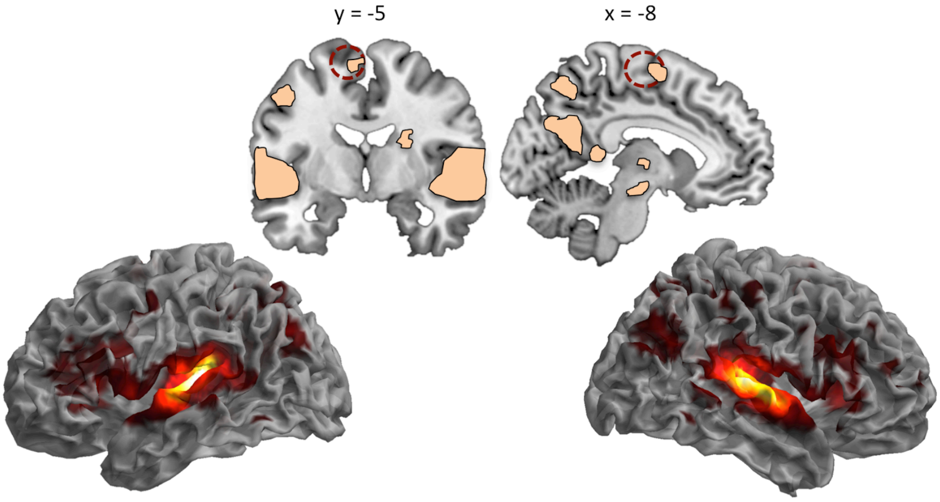


Figure 3. Association between lower representational similarity of functional responses to different types of heard sounds in SMA (i.e., higher specificity/fidelity) and higher reported vividness of auditory imagery.

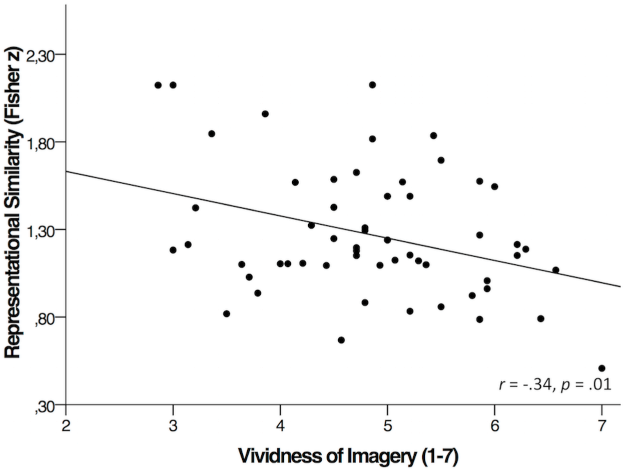
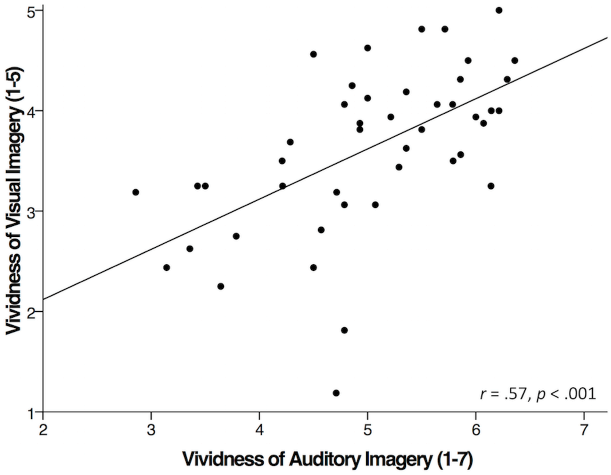


Figure 4. Association between vividness of visual and auditory imagery. Higher vividness corresponds to higher ratings for auditory and visual imagery.


Vividness of Visual Imagery (1-5)
Vividness of Auditory Imagery (1-7)
r = .57, p < .001