Flexible Information Coding in Human Auditory Cortex during Perception, Imagery, and STM of Complex Sounds

Annika C. Linke[1,2] and Rhodri Cusack[1,2]

Abstract

■ Auditory cortex is the first cortical region of the human brain to process sounds. However, it has recently been shown that its neurons also fire in the absence of direct sensory input, during memory maintenance and imagery. This has commonly been taken to reflect neural coding of the same acoustic information as during the perception of sound. However, the results of the current study suggest that the type of information encoded in auditory cortex is highly flexible. During perception and memory maintenance, neural activity patterns are stimulus specific, reflecting individual sound properties. Auditory imagery of the same sounds evokes similar overall activity in auditory cortex as perception. However, during imagery abstracted, categorical information is encoded in the neural patterns, particularly when individuals are experiencing more vivid imagery. This highlights the necessity to move beyond traditional "brain mapping" inference in human neuroimaging, which assumes common regional activation implies similar mental representations. ■

INTRODUCTION

Neurons in human auditory cortex (AC) do not fire only during the perception of sound but also when there is no acoustic input, such as during STM maintenance (Sreenivasan, Curtis, & D'Esposito, 2014; Pasternak & Greenlee, 2005), auditory imagery (Oh, Kwon, Yang, & Jeong, 2013; Zvyagintsev et al., 2013; Kosslyn, Ganis, & Thompson, 2001), and when people are watching silent video clips (Meyer et al., 2010). It is possible that these tasks evoke mental representations that mirror sensory input. However, the subjective experience of an internally generated sound is undoubtedly different from listening to it. This is one reason why the philosophical debate over how similar mental representations during imagery are to perception has continued despite extensive neuropsychological and neuroimaging research addressing this question over the past decades (Kosslyn, 2003; Pylyshyn, 2003). One assumption often made is that the recruitment of primary sensory cortices during imagery would support the hypothesis that representations during mental imagery mirror those during perception. First evidence for this hypothesis came from neuropsychology. Penfield and Perot (1963) electrically stimulated superior temporal gyrus in a patient undergoing treatment for epilepsy and showed that such stimulation in the absence of auditory input caused auditory hallucinations. Some patients with damage to occipital cortex who can no longer see also lose their ability to visually imagine (Farah, 1984), and for some patients, deficits are feature specific in vision as well as visual imagery, for example, to color (De Vreese, 1991) or faces (Young, Humphreys, Riddoch, Hellawell, & de Haan, 1994). Patients with damage to the temporal lobes similarly show deficits in auditory imagery (Zatorre & Halpern, 1993). However, evidence from patients who can either still imagine or have intact perception while the other function is impaired suggests that imagery and perception only partly share the same neural mechanisms (e.g., Behrmann, Winocur, & Moscovitch, 1992).

With the onset of brain imaging and particularly fMRI, it was possible to study imagery in an entirely new way. Being able to observe brain activity while participants were engaged in mental imagery was thought to quickly resolve the question of whether imagery draws upon similar neural mechanisms as perception. Despite these technological advances, it remains unclear how much mental representations during imagery resemble those evoked by perception. In vision, some studies find primary sensory activation during imagery. Others, however, only show activity in secondary visual areas (see Kosslyn et al., 2001, for a short review), and in almost all existing studies on auditory imagery, association but not primary cortices are activated (Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Zatorre & Halpern, 2005; Yoo, Lee, & Choi, 2001; Halpern & Zatorre, 1999). This led to the proposal that primary visual (Kosslyn et al., 2001) and auditory (Kraemer, Macrae, Green, & Kelley, 2005) cortices are recruited during imagery only if a task involves bringing detailed low-level features of the stimulus to mind.

[1] Western University, London, ON, Canada, [2] Medical Research Council, Cambridge, United Kingdom

Journal of Cognitive Neuroscience 27:7, pp. 1322–1333
doi:10.1162/jocn_a_00780

Recent studies have begun to address not only whether perception and imagery recruit the same cortical regions but also whether it is also the same information about a stimulus that is being encoded. Importantly, averaging across voxels—the most ubiquitous analysis method for fMRI—decreases the chances of detecting differences in neural coding. It is possible that the information encoded in a brain region is qualitatively different even if average activity is not (Lee, Kravitz, & Baker, 2013). Multivariate statistical methods such as multivoxel pattern analysis (MVPA) make use of spatially distributed patterns of activity instead and can reveal representational differences even if overall activity is the same across conditions (Kriegeskorte, Goebel, & Bandettini, 2006; Haxby et al., 2001). For instance, using MVPA, Albers, Kok, Toni, Dijkerman, and de Lange (2013) recently showed that activity patterns in early visual cortex were stimulus specific during visual working memory and imagery and resembled activity patterns during perception. From this, they concluded that even top–down processes such as visual imagery are "perception-like," relying on the same representations as bottom–up visual stimulation. Importantly, however, the stimuli they used were simple grayscale gratings that do not contain much information other than low-level features and do not lend themselves to representational abstraction as might be expected during imagery of more complex sensations. To address the issue of whether neural representations of complex sounds contain the same information when different tasks are performed, it is necessary to use a larger set of complex, naturalistic stimuli that share basic perceptual features as well as semantically meaningful characteristics.

The question of which information about a stimulus is processed in a brain region is not only relevant for resolving the century old debate over whether imagery relies on veridical or abstracted representations but also fundamental to our understanding of results from "brain mapping" more generally. Imagery is a prime example for the typical assumption in neuroimaging that common activation of a region in different tasks implies similar information about a stimulus is being processed even if the cognitive demands differ. Here we test whether and how the contents of neural representations in human AC change as individuals engage in auditory imagery of complex, naturalistic sounds. If imagery involves a veridical representation of sensory input, activity patterns during perception and imagery should be very similar, reflecting encoding of the same detailed information about a stimulus. Alternatively, more abstract representations may be present during self-reported imagery. For example, instead of low-level acoustic features, more general semantic information about a sound could be the dominant characteristic encoded. The complex sounds used in this study were therefore selected to deconfound basic acoustic features and semantic category. This allowed us to test whether neural patterns change their information content during self-reported imagery compared to perception.

To further test whether the information about a stimulus encoded in the same brain region changes in more subtly different tasks, participants also performed an STM task in separate blocks of the experiment. Imagery is frequently quoted as a useful strategy to retain information in STM or is even equated with memory rehearsal and retrieval processes (Kraemer et al., 2005). Indeed, any form of imagery requires memory, and it is likely that STM maintenance draws upon or is aided by some form of imagery. However, at least for some memory tasks—such as change detection—reports of sustained firing in sensory regions across species (Pasternak & Greenlee, 2005) during STM maintenance could also reflect a bottom–up process that is more automatic than imagery. This is supported by findings that attention and rehearsal are not always necessary for successful auditory change detection (e.g., Linke, Vicente-Grabovetsky, & Cusack, 2011; Demany, Semal, Cazalets, & Pressnitzer, 2010) suggesting that detailed information about the sound is maintained fairly automatically (e.g., through sustained firing of the same neurons). Importantly, the imagery and STM tasks used the same stimuli and an almost identical task design (Figure 1) and only differed in the instructions people were given and the response they had to make at the end of a trial. Additionally, both tasks required some form of neural representation during the delay period that is dependent on the previously heard sound. This allowed us to compare not only whether neural representations in auditory regions change their informational content during perception and higher-level cognition but also whether the information encoded about the same stimulus and in the same auditory sensory areas of the brain differs depending on the task being performed.

METHODS

Participants

Twenty-five healthy participants took part in the experiment as paid volunteers. Three participants were excluded from the statistical analysis because of excessive movement (>10 mm). Thus, data from 22 participants (age = 19–35 years, *M* = 24.92, *SD* = 4.4, 14 women) were subsequently analyzed. All participants reported that they had normal hearing and none had any intensive musical training (*M* = 4.43 years of instrument lessons, *SD* = 4.64) or reported having perfect pitch. Participants were recruited from the MRC Cognition and Brain Sciences Unit volunteer panel and gave informed, written consent before beginning the experiment. Ethical approval was obtained from the Cambridge Local Research Ethics Committee.

Stimulus Selection

Sounds were chosen to be as distinct from one another as possible to ensure that any similarities in activity patterns within a category would be attributable to the

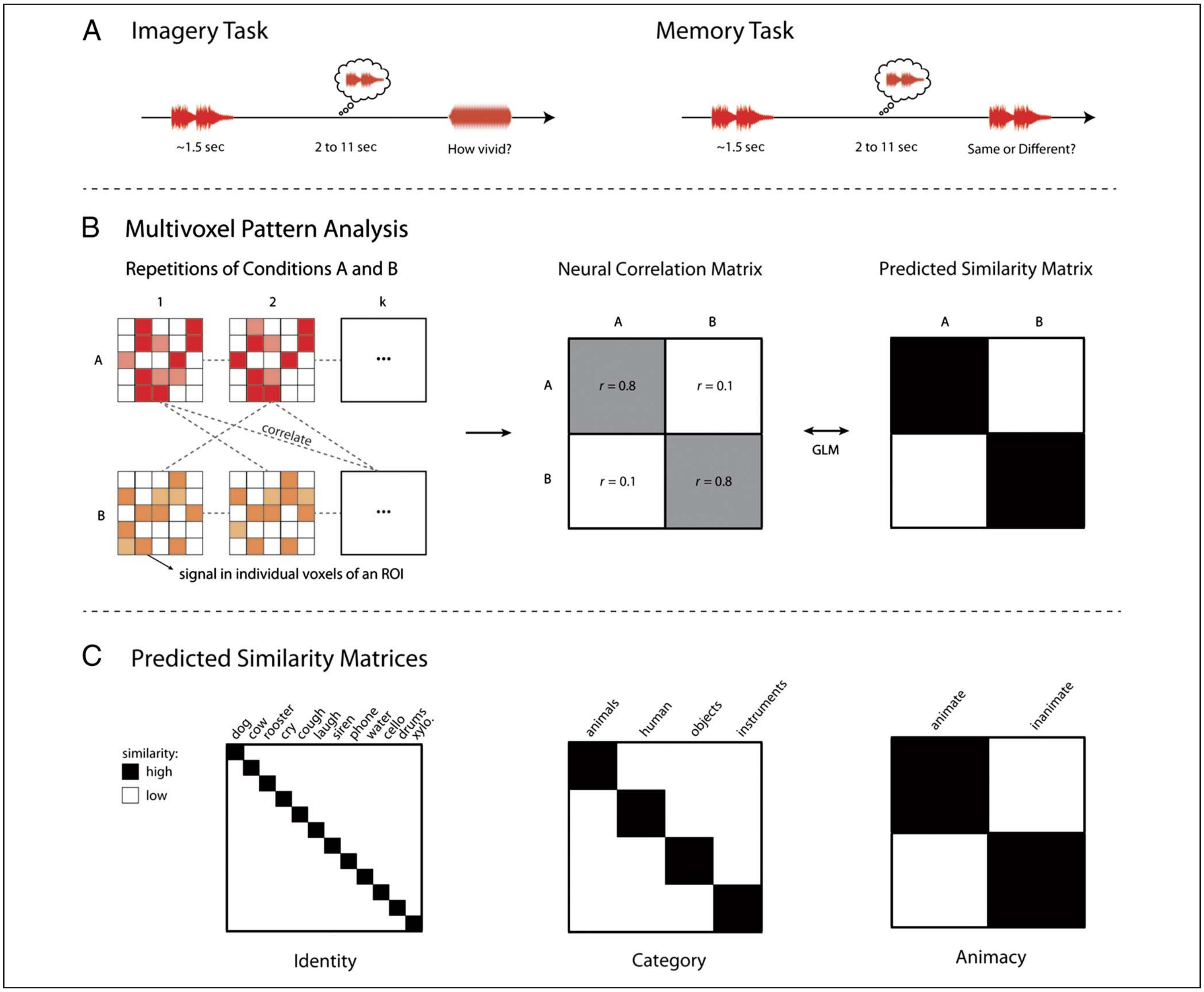


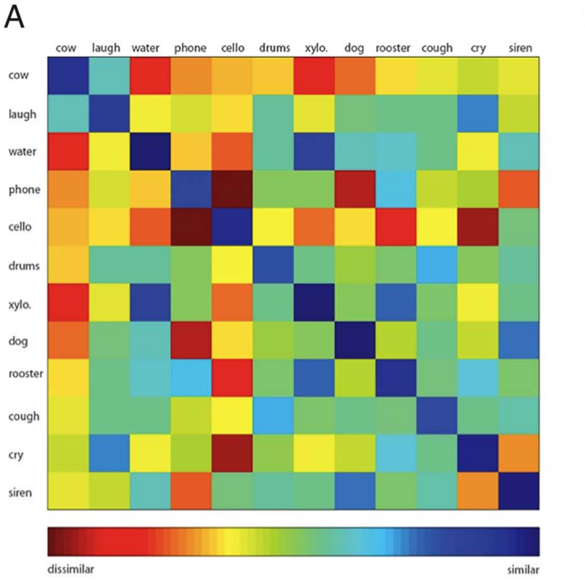
Figure 1. (A) Schematic of the two tasks. In the imagery task, participants listened to a sound and were then asked to imagine the sound they had just heard as vividly and accurately as possible during a variable delay period. A "beep" signaled them to stop and press a button to rate the clarity of their imagery. In the memory blocks, participants similarly listened to a sound and were instructed to hold the sound in memory during the delay period. They then heard the same or a slightly modified version of the same sound again and pressed a button to indicate whether it had changed. The ITI as well as the delay period (during which imagery and memory maintenance were performed) were jittered unpredictably (2–11 sec). (B) MVPA: Voxel-wise data for each ROI and task phase (perception, imagery and memory) were Spearman-correlated and fit to a GLM to test for the consistency of spatial patterns. (C) Three different contrasts were used in the GLM to compare the neural similarity matrix to the predicted matrices and to differentiate what level of information was present in the distributed neural activity patterns. The identity contrast tests the consistency of activity patterns evoked by each individual sound. The category contrast tests whether patterns within a category are more similar than across categories. Lastly, the animacy contrast tests for the highest level of abstraction in which patterns can be distinguished based on whether the sound was produced by an animate or inanimate source.

semantic meaning and not the physical characteristics of the sounds. Twelve complex diotic sounds were chosen from a large sound set comprising 140 different natural sounds to cover human nonspeech vocalizations (coughing, crying, laughing), animal vocalizations (cow, dog, rooster), sounds originating from nonliving sources (ambulance siren, phone ringing, water flowing), as well as instrument sounds (cello, drums, xylophone). First, the larger set was reduced to 54 sounds by excluding all sounds that were uncommon or potentially hard to imagine (e.g., "punching"). The number of sounds per semantic category was matched. This set was then rated by six additional participants (four women, 24–27 years old, M = 25.5, SD = 1.38) to find those sounds that were judged as being the most imaginable. Each participant listened to each of the 54 sounds and rated how difficult it was to imagine the sound on a scale from 1 (*very difficult*) to 9 (*very easy*). On the basis of these ratings, the set was narrowed down to the 12 sounds that were rated to be the most imaginable and could easily be modified in their acoustic parameters without introducing audible artifacts. Next, three additional exemplars of each of the 12 sounds were created by modifying each original sound on three characteristics (frequency, playback speed, or loudness) in turn. The frequency of the sound was changed in Audacity (audacity.sourceforge.net/) with the tempo of the sound remaining largely unchanged. Similarly, a vocoder algorithm was used to change the tempo of the sounds without substantially altering frequency (labrosa.ee.columbia.edu/matlab/pvoc). Sounds had a mean duration of 1.54 sec (*SD* = 0.46) and varied widely in their acoustic features (Figure 2).

Experimental Design

Participants performed two different tasks while being in the scanner—a change detection and an imagery task (Figure 1). During the change detection task, one sound was played followed by a silent maintenance/delay period. The maintenance period and the intertrial interval (ITI) were jittered (2–11 sec) to ensure that the different phases of the task could be modeled separately in the fMRI analysis. Participants then heard the same sound again. In 50% of the trials, the same exemplar was played. If a change occurred in the other half of the trials, the new sound was a different exemplar of the same sound. This way, participants were forced to encode the sound as a whole, preventing them from focusing on a few, salient, or task-relevant features. Participants were instructed to respond as soon as the probe sound had finished playing by pressing one of two buttons to indicate a "same" or "different" response.

During the imagery task, participants again heard one sound during the encoding period. Like in the change detection task, the following delay period was jittered (2–11 sec) and participants were instructed to imagine, as vividly as possible, the sound they had just heard, repeating it in their head for the duration of the delay. The end of the delay was signaled by a 0.3-sec beep (523 Hz) and a 2-sec visual display, asking them to rate how vivid/clear their imagery had been during the preceding delay by pressing one of four buttons (*very clear*, *clear*, *blurred*, *absent*). Like for the change detection task, the response phase was followed by a variable ITI (2–11 sec). Importantly, the timings for the two different tasks were identical, except for the presentation of the probe during the response phase.



B

sound	mean pitch (Hz)	mean intensity (dB SPL)	duration (sec)	sound	mean pitch (Hz)	mean intensity (dB SPL)	duration (sec)
cow	270	83	2.0	xylophone	999	75	1.6
	315	83	2.0		793	74	1.6
	272	79	1.4		1006	69	1.4
	270	76	2.0		999	66	1.6
laugh	768	79	1.1	dog	814	81	1.1
	679	79	1.1		689	81	1.1
	739	75	0.9		897	76	0.9
	767	67	1.1		814	71	1.1
water	760	73	1.9	rooster	799	81	2.2
	1005	72	1.9		687	81	2.2
	760	70	1.4		803	77	1.7
	760	64	1.9		799	72	2.2
phone	615	80	1.5	cough	550	77	0.9
	541	80	1.5		416	77	0.8
	747	74	1.1		470	71	1.3
	615	73	1.5		550	67	0.9
cello	116	75	1.7	cry	479	79	2.4
	128	75	1.6		402	79	2.4
	110	71	1.3		485	74	1.5
	116	67	1.7		479	71	2.4
drums	146	72	1.7	siren	762	84	1.1
	118	72	1.7		856	84	1.0
	165	69	1.5		765	81	0.8
	146	64	1.7		762	76	1.1

Figure 2. Sound stimuli used in the imagery and short-term memory tasks. Sounds varied widely on their acoustic features within a category: human nonspeech vocalizations (coughing, crying, laughing), animal vocalizations (cow, dog, rooster), sounds originating from nonliving sources (ambulance siren, phone ringing, water flowing), as well as instrument sounds (cello, drums, xylophone). (A) Similarity matrix of the sounds' physical characteristics (pitch, brightness, timbre, and attack timings; analyzed using the MIR toolbox [https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox] for Matlab). For each sound characteristic, the pairwise cosine distance was calculated across all sounds in the set yielding a 48 × 48 similarity matrix (12 sounds × 4 exemplars) for each of the four sound features. The similarity matrices for each sound feature were then summed, with the four sound features being weighted equally, and a mean similarity value across the four exemplars (see B) was calculated resulting in the 12 × 12 similarity matrix. (B) Four exemplars of each sound were presented in the experiment, modified either in frequency, playback speed, or loudness without causing distortion. Sound characteristics for each exemplar are listed.

In both tasks, sounds were played via custom-built NordicNeuroLab headphones and presented using Matlab (www.themathworks.com) with the Psychtoolbox (Brainard, 1997). Participants completed two blocks of each task. To reduce the intention of using an imagery strategy during the change detection task, an ABBA design that always started with a change detection block was used. Participants were only informed and instructed about the imagery task once they had completed this first change detection block. Before entering the scanner, the different exemplars of a randomly selected sound from the set were played to the participants to demonstrate that the change detection task was easy and did not require active rehearsal during the maintenance interval to further discourage imagery strategies. It was not mentioned which characteristics of the sounds had been changed but only that these were some examples of slight variations of the same sound.

Each block consisted of 36 trials and took approximately 10 min. To allow for multivariate pattern analysis to be conducted, each block of 36 trials was divided into three subblocks of 12 trials—one for each of the 12 sounds—to ensure that repetitions of the same sound would be separated in time. Each delay/ITI jitter was presented once per subblock, yielding 12 different delay and ITI durations. Similarly, one exemplar of each sound was played in each subblock. Which precise exemplar was presented during the encoding period of the tasks was randomized. Throughout the tasks, participants were instructed to fixate on a white cross presented on an otherwise black screen and the hand used to make responses was counterbalanced across participants.

Behavioral Analysis

A paired samples *t* test was carried out to assess whether performance differed in the first and second block of the change detection task. A change in performance could suggest that participants were using a different strategy after they had been exposed to the imagery task or that repeated exposure to the sounds made the task easier to perform over time. On the basis of the results of a post-experimental questionnaire in which participants indicated whether they used a specific strategy during the change detection task, we also performed an independent samples *t* test to check whether performance differed for those participants that had and had not engaged in imagery during the change detection blocks. Lastly, to make sure that any differences in activity we would find during the change detection and imagery blocks were not because of task difficulty, we compared participants' ratings of attentiveness and how hard they had found the two different tasks with an additional paired-samples *t* test.

Functional Imaging

Scanning was performed at the MRC Cognition and Brain Sciences Unit on a Siemens (Erlangen, Germany) TIM Trio 3T scanner. At the beginning of each session, a whole-brain T1-weighted high-resolution structural image was acquired with an MPRAGE sequence (matrix size = 256 × 240 × 160, flip angle = 9°, repetition time = 2250 msec, echo time = 2.99 msec, 1 mm isotropic resolution). Functional imaging data covering most of the brain (small parts of the frontal and temporal poles were missing because of the field of view) were acquired using a quiet EPI sequence (Schmitter et al., 2008) with the following parameters: 32 slices, matrix size = 64 × 64, 3 mm slice thickness, including a 25% gap, flip angle = 83°, repetition time = 2640 msec, echo time = 44 msec, bandwidth 1220 Hz/Px, 3 mm × 3 mm resolution. Eight dummy scans were discarded in the analysis to allow for T1 equilibrium. This sequence was chosen to reduce interference from scanner noise without the trade-off of a reduced number of volume acquisitions and fixed assumptions about the precise shape of the hemodynamic response function as would have been necessary when using a sparse EPI sequence. The quiet EPI sequence implemented at the MRC Cognition and Brain Sciences Unit reduces noise by approximately 24 dB and has been shown to be particularly well suited for experiments involving auditory presentations of stimuli (Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010).

Functional Imaging Preprocessing

Imaging data were preprocessed (including slice-time correction, realignment to a reference image, nonlinear normalization to the Montreal Neurological Institute template brain, and, for the univariate analysis only, spatial filtering with a 10-mm FWHM Gaussian kernel) and analyzed using SPM5 software (Wellcome Department of Imaging Neuroscience, London, UK) and the automatic analysis library developed in our laboratory (https://github.com/rhodricusack/automaticanalysis/wiki).

Univariate Analysis

A general linear model (GLM) was fit to the acquired data with separate regressors for each of the three task phases (perception, imagery/memory, response), averaging across all sounds. Event onsets were defined as the onset time of the sound for the perception phase, the end of the sound for the imagery/memory phase, and the response probe signaling the end of the imagery/memory phase for the response period. Durations of events were based on the exact length of the sound during perception, the length of the jittered imagery/memory period, and, for the response phase, the period from its onset until a button press had been recorded. The time course was convolved with the canonical hemodynamic response function as defined by SPM. The jittered ITIs served as the baseline. A contrast for each phase versus baseline as well as a perception versus imagery, perception versus memory, and imagery versus memory contrast were tested in a standard univariate analysis. All results were multiple-comparison (FDR) corrected at $p < .05$.

ROI Selection

A Heschl's gyrus (HG) ROI, representative of primary AC, and a larger noncore AC ROI were created for the multivariate analysis by masking activity generated when participants were listening to different pure tones (compared to a silent baseline, whole-brain family-wise error corrected at $p < .005$; peak activity in Montreal Neurological Institute coordinates at [56, −14, 2] on the right and [−52, −24, 6] on the left) with HG and superior temporal regions as defined in the MarsBar AAL ROI package (Brett, Anton, Valbregue, & Poline, 2002), respectively. This sound versus silence contrast was derived from a previous study using pure tones (Linke et al., 2011). Additionally, a middle temporal gyrus (MTG) ROI was derived from the speech versus scrambled speech contrast (family-wise error corrected at $p < .05$) from Rodd, Davis, and Johnsrude (2005). These ROIs were chosen based on the a priori hypothesis that categorical information, particularly during imagery, might be represented in regions involved in processing semantic information (i.e., MTG) but not in regions usually thought of as mainly processing physical sound characteristics (i.e., HG and noncore AC). All ROIs were back-normalized to the space of the individual participants' brains for the multivariate analysis described below.

Multivariate Analysis

MVPA was used to establish whether representations in auditory regions differed during perception, self-reported imagery, and memory maintenance in three ROIs (HG, larger AC, and MTG). Importantly, the correlational method chosen is insensitive to differences in average activation magnitude providing orthogonal information about how a stimulus is processed (Figure 1B). A new GLM was fit to the data with individual regressors modeling each sound separately in each subblock. To ensure all comparisons were made across equal temporal distributions, MVPA was restricted to comparisons across subblocks and was carried out for the two phases of interest, perception and imagery, only. All data were gray matter masked. For each individual participant, beta values for all events were extracted for the three ROIs. The voxel-wise data were then Spearman correlated. These correlations were normalized to assure that each subblock contributed equally to the average correlations. By taking the mean across all subblocks, the data were then condensed into a 24 × 24 (12 sounds, two task phases—perception, imagery/memory) correlation matrix and contrasted with the identity, category, and animacy matrices (illustrated in Figure 1C and explained in more detail below) for each task phase using a GLM (Figures 3 and 4). *Identity contrast*: First, we tested whether activity patterns for repetitions of the same sound were more similar to one another than to repetitions of the other sounds in the set. This would show that it can be decoded from the individual patterns of activity which of the 12 sounds a participant was listening to. *Category contrast*: Second, we grouped sounds according to their semantic category membership and tested whether activity patterns of sounds within a category were more similar than activity patterns across categories. *Animacy contrast*: Third, we tested whether the activity patterns evoked by animate sounds (human and animal vocalizations) and inanimate sounds (object sounds and instruments) were more similar to sounds of their respective category than to sounds of the other animacy category. Because it was taken care to choose sounds within a category to differ in their physical features, significant pattern similarities for the last two contrasts would suggest that

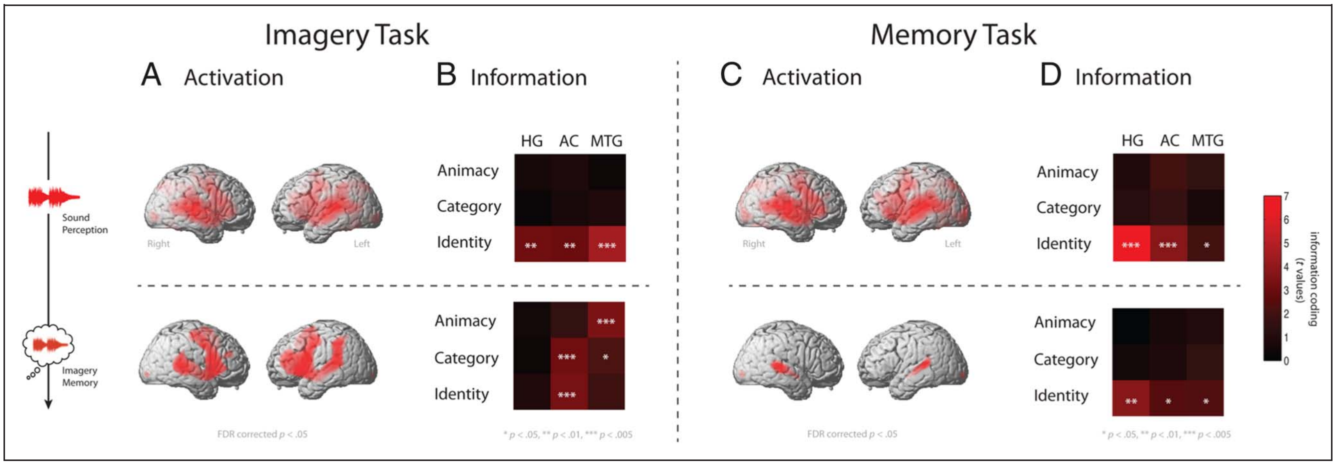
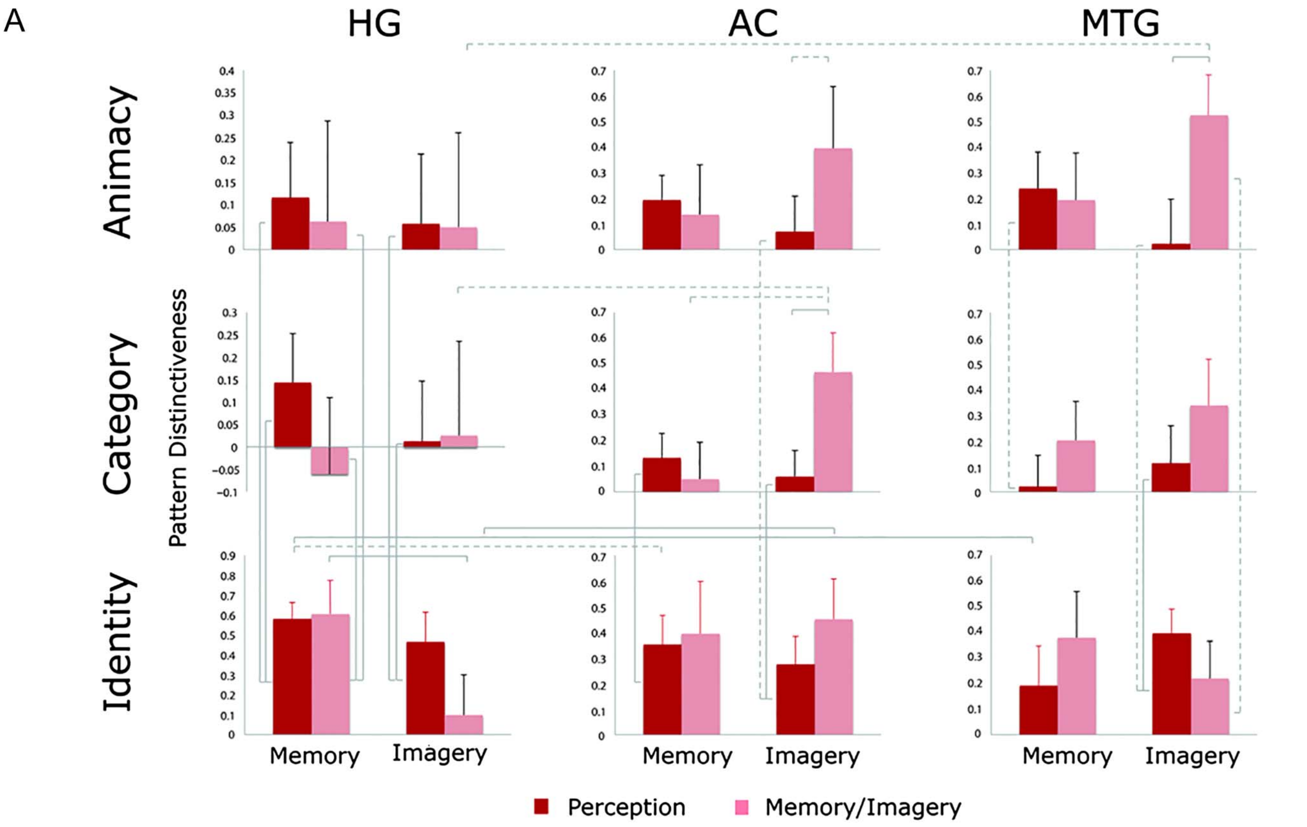


Figure 3. The brain rendering shows regional average activation during perception (top row) and imagery/memory (bottom row) for each task (FDR corrected for multiple comparisons, $p < .05$). The matrix shows MVPA results for the three different ROIs (HG, AC, and MTG) and each task. For each participant, a GLM contrasted neural pattern similarity in each task phase and ROI with the identity, category, and animacy matrices (Figure 1C). Higher values indicate more distinct information coding (expressed in t values; also see Figure 4 for error bars and results of additional repeated-measures ANOVA).



B

	Memory				Imagery			
	Perception		Delay		Perception		Delay	
	p Value	*t* Value	*p* Value	*t* Value	*p* Value	*t* Value	*p* Value	*t* Value
Identity								
HG	.001	6.63	.001	3.74	.005	3.14	.342	0.97
AC	.001	3.74	.022	2.48	.008	2.96	.003	3.31
MTG	.051	2.07	.031	2.32	.000	4.49	.074	1.88
Category								
HG	.286	1.10	.650	0.46	.750	0.32	.738	0.54
AC	.112	1.66	.392	0.87	.469	0.74	.005	1.63
MTG	.150	0.67	.165	1.44	.351	0.95	.038	3.44
Animacy								
HG	.361	0.93	.920	0.10	.565	0.59	.594	0.54
AC	.055	2.03	.467	0.74	.443	0.78	.118	1.63
MTG	.150	1.50	.302	1.06	.714	0.37	.002	3.44

Figure 4. MVPA results. (A) Post hoc results of repeated-measures ANOVA. Pattern distinctiveness is expressed in beta values derived, for each participant, by fitting a GLM contrasting neural pattern similarity in each task phase and ROI with the identity, category, and animacy matrices (Figure 1C). Red error bars signify significant coding (also see matrices in Figure 3); dashed lines show significant post hoc *t* test ($p < .05$, one-tailed); solid lines show significant post hoc *t* test ($p < .05$ two-tailed). (B) Statistics for the three ROIs, three contrasts, and two tasks.

the ROI under investigation processed information on a semantic level in addition to simply processing the physical characteristics of a sound.

Individual Differences Analysis

To test whether individual differences in imagery ability and memory capacity are related to the magnitude of neural activity in early auditory and higher-level regions, mean beta values were extracted from the same three ROIs for each participant and task phase. The behavioral measures (the individuals' mean rating of imagery clarity after each trial and a measure of memory capacity, Cowan's $K = n \times (H - \text{FA})$, where n = number of items to be remembered, H = proportion of hits, and FA = proportion of false alarms; Cowan, 2001) were then standardized (z-scored) and Pearson-correlated with the mean beta value in each of the ROIs, separately for perception, imagery, and memory maintenance.

To test whether individual differences in abstraction (as measured by how consistently identity, category, and animacy were encoded) were related to participants' ability to engage in clear imagery or performance in the memory task, we extracted the individual participants' MVPA results (beta values), standardized them (z scores), and Pearson-correlated them with their mean trial-by-trial ratings of imagery clarity and memory capacity (z scores) for each level of abstraction (identity, category, and animacy), ROI (HG, AC, and MTG), and task phase (perception, imagery, and memory maintenance).

RESULTS

Behavioral Results

Missing trials (8.2% of trials in the change detection and 4.5% of trials in the imagery task) were excluded in the behavioral results reported here.

In the imagery blocks, participants rated the clarity of their imagery on a scale from 1 (*very clear*) to 4 (*absent*) after every trial. There were no reliable differences in the average trial-by-trial clarity of self-reported imagery ratings based on a sound's category [human nonspeech vocalization, animal vocalization, inanimate sounds or instruments; $F(1, 3) = .755, p = .524$] or whether they originated from animate or inanimate sources [$t(20) = 1.029, p = .315$, two-tailed].

All participants performed well in the change detection task but not at ceiling ($M = .87, SD = .34$). Performance did not differ in change versus no-change trials [$t(20) = .40, p = .69$, two-tailed] or in the first versus second change detection block [$t(20) = .64, p = .53$, two-tailed]. False alarms, that is, reporting that a change had occurred when the sound had remained the same, and misses, that is, responding "same" when in fact a different sound exemplar had been played, were similarly frequent (7.2% and 6.2% of change detection trials, respectively).

Participants were asked to indicate in the postexperimental questionnaire whether they had used a specific strategy to perform the change detection task. If they answered "yes," they described the strategy they had used in detail. They also indicated whether they had changed their strategy when performing the change detection task for the second time. We were particularly interested in how many of the participants had used imagery during STM maintenance and whether the intermittent imagery blocks had changed their behavior during the second change detection block. Thirteen of the 22 participants responded that they had used a strategy, nine of which explicitly stated that they used imagery or described an imagery-like strategy (such as having replayed the sounds in their head). The remaining four participants used an alternative strategy that could not easily be related to imagery, for example, "feeling the location of the sound in the brain" or "focusing on the start and end of the sound." Only six participants indicated that they changed their strategy in the second change detection block (four of which had not used a strategy during the first block but used an imagery strategy during the second block, and two who switched from an alternative strategy described above to imagery). For the purpose of the analysis, we grouped participants that reported having used a strategy in both change detection blocks (n = 13) and those that had not used any strategy in at least one of the two blocks (n = 9, with seven participants not having used a strategy during the first or second block). Performance did not differ depending on whether participants reported using a strategy or not to keep the sounds in STM [$t(20) = .10, p = .92$, two-tailed], replicating results of a previous study (Linke et al., 2011) that showed auditory change detection to be independent of cognitive strategy used and in accordance with other findings suggesting that auditory change detection might be an automatic process (Demany et al., 2010).

To be able to accurately compare activity patterns during STM maintenance and self-reported imagery, it is important that the two tasks were equally difficult and observed differences not due to one of the tasks being significantly easier to perform. Participants rated attentiveness during and difficulty of the two tasks on a 1 (*not attentive/very easy*) to 5 (*very attentive/very difficult*) scale after the experiment. The tasks did not differ in this attentiveness (MCD = 3.36, SDCD = 0.95, MIMG = 3.45, SDIMG = 1.01; $t(20) = 0.34, p = .74$, two-tailed) or difficulty rating (MCD = 2.55, SDCD = 0.86, MIMG = 2.27, SDIMG = 0.94; $t(20) = 1.0, p = .33$, two-tailed).

Imagery Task

As found previously (Herholz, Halpern, & Zatorre, 2012; Kraemer et al., 2005; Zatorre & Halpern, 2005), self-reported auditory imagery activated similar regions as sound perception (Figure 3A). To assess whether these regions also encoded the same information about a sound, we used MVPA as described above. First, we determined whether neural activity patterns of individual sounds (Figure 1C, identity coding) could be distinguished in three cortical regions along the auditory processing pathway—HG, noncore AC, and MTG. During perception, MVPA revealed identity coding in all three ROIs, indicating that the sounds' individual properties were encoded (Figure 3B). We found no evidence of category or animacy coding in the three ROIs during perception (Figures 3B and 4). However, consistent with the hypothesis that representations during imagery might be abstracted, our MVPA results show that during imagery, categorical information was encoded in noncore AC [$t(20) = 3.11, p < .005$, two-tailed] and MTG [$t(20) = 2.21, p < .05$, two-tailed]. Additionally, neural patterns of activity in MTG also contained information at the highest level of semantic abstraction, reflecting whether a sound came from an animate or inanimate source [$t(20) = 3.44, p < .005$, two-tailed].

To directly compare the similarity of representations during the different task phases, we then correlated activity patterns during perception and self-reported imagery but the comparison was not significant for any of the three ROIs.

STM Task

During the perception stage of the STM task, the same network of regions as during the imagery task was activated and all three ROIs showed identity coding (Figure 3C). However, activation during memory maintenance was constrained to AC. This difference between STM maintenance and self-reported imagery was not driven by higher cognitive demands in one of the tasks and was also reflected in which information about a stimulus was encoded in the three auditory ROIs (Figure 3D). When participants were holding information in STM, identity coding persisted, replicating results from visual STM research (Harrison & Tong, 2009). Unlike during imagery, category and animacy were not encoded in any of the ROIs (Figure 3D).

Statistical Comparison of Perception, Imagery, and STM MVPA Results

A four-way repeated-measures ANOVA (Task [CD, Imagery] × ROI [HG, AC, MTG] × Contrast [identity, category, animacy] × Task phase [perception, imagery/memory]) was carried out on the MVPA results to test whether differences in identity, category, and animacy coding were significant within and across the two tasks. Results revealed a significant main effect of Contrast [$F(2, 42) = 5.12, p < .01$], a significant ROI × Contrast interaction [$F(4, 84) = 3.72, p < .01$], and Task × Contrast × Task phase interaction approaching significance [$F(2, 42) = 2.60, p = .08$]. Results of post hoc paired-samples t tests are shown in Figure 4. *STM memory task*: For the STM task, identity coding was significantly higher than animacy coding [in HG: $t(21) = 3.82, p < .001$] and category coding [in HG: $t(21) = 4.02, p < .001$, and AC: $t(21) = 2.32, p < .05$] during perception. Similarly, during memory maintenance, identity coding was higher than animacy coding [HG: $t(21) = 2.10, p < .05$] and category coding [in HG: $t(21) = 3.12, p < .005$]. The degree of identity, category, and animacy coding, however, did not differ significantly between perception and memory maintenance for any of the ROIs, indicating that representations during perception and memory maintenance were similar. *Imagery task*: Similar to the results from the STM task, identity coding was higher than animacy coding [in HG: $t(21) = 2.40, p < .05$, and MTG: $t(21) = 2.26, p < .05$, and approaching significance in AC: $t(21) = 1.78, p = .09$, two-tailed] and category coding [in HG: $t(21) = 3.00, p < .01$; AC: $t(21) = 2.15, p < .05$, and MTG $t(21) = 2.31, p < .05$] during perception. Unlike for the STM task, however, this was not the case during self-reported imagery. During imagery, category and animacy coding was significantly higher than during perception in AC [category: $t(21) = 2.88, p < .01$; with animacy approaching significance: $t(21) = 1.84, p = .08$, two-tailed] and MTG [animacy: $t(21) = 2.10, p < .05$]. This further confirms the pattern of the MVPA results reported above which show that representations become more abstracted during self-reported imagery compared to perception or memory maintenance.

Individual Differences

Next, we assessed whether the content of neural representations was related to behavioral measures of imagery clarity and memory capacity. We correlated the magnitude of activity in the three auditory ROIs with mean trial-by-trial ratings of imagery clarity and a measure of STM capacity (Cowan, 2001). Imagery clarity ratings correlated positively with mean overall activity in MTG [$r(20) = .51, p < .01$, two-tailed] during perception and with activity in AC [$r(20) = .44, p < .05$, two-tailed] and MTG [$r(20) = .43, p < .05$, two-tailed] during imagery. More importantly, how distinctly category information was coded for in AC while participants were imagining the sounds (as revealed by MVPA) was positively correlated with imagery clarity ratings and approached significance [$r(20) = .39, p = .07$, two-tailed; Figure 5]. The degree to which activity patterns contained abstract information during the perception stage of the imagery task also correlated with perceived clarity of imagery with approaching significance [category: $r(20) = .40, p = .06$, two-tailed; animacy: $r(20) = .43, p < .05$, two-tailed], suggesting that even during perception future task demands might influence how sensory information is encoded. Although the correlations of information coding with imagery clarity were only approaching significance and additional studies or a larger subject pool are necessary to draw final conclusions, these results suggest that abstract mental representations are essential for successful imagery, contrary to the predictions of a model of imagery

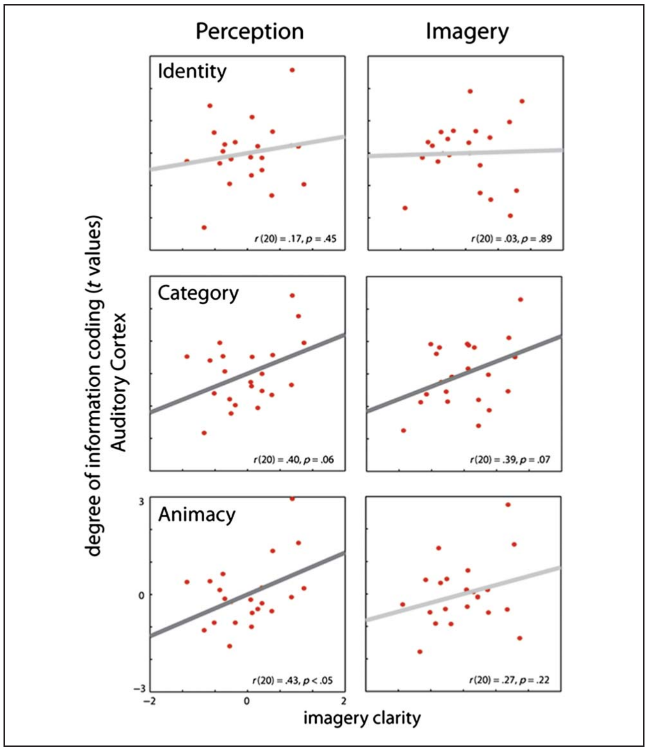


Figure 5. Pearson correlations of individual participants' MVPA results (z-scored beta values) and mean trial-by-trial imagery clarity ratings (z-scored) for the three levels of information represented in AC. Two-tailed statistics are reported.

as a veridical representation of the percept. Memory capacity, on the other hand, did not correlate with overall activity or the MVPA results in any of the three auditory ROIs.

DISCUSSION

How the brain manages to encode vastly different sensory information under the pressure to quickly form a stable percept of the environment and act upon it is the core question of cognitive neuroscience. Neuroimaging has been utilized in thousands of studies over the last few decades to map function to different areas of the brain. Importantly however, common activation in a brain area does not necessarily imply the same information about a stimulus is being processed. In the current study, we showed that neural representations of complex sounds differ in their informational content depending on the task participants were performing. Similar to previous studies on imagery, regions activated while participants were imagining complex, natural sounds largely overlapped with regions activated during perception. In the MVPA group analysis, we found, however, that semantic information was represented in AC and MTG during imagery only. Furthermore, activity patterns during perception and self-reported imagery were not the same in auditory regions even though they showed significant univariate activation during both task phases. During STM maintenance of the same complex sounds, the same abstraction was not observed. This shows that information is represented in an abstracted way during self-reported imagery but not during STM and implies that, although sensory regions can be activated in the absence of sensory input, the nature of the representations might not be the same as during perception. This has important implications for our understanding of how sounds are processed in auditory regions of the human brain. Traditionally, sensory information has been thought of as passing through a hierarchy of feed-forward processing steps with primary cortices analyzing basic properties of the sensory signal (Nelken, 2008; Wessinger et al., 2001), but this hierarchy was violated in the current study in two ways: during perception, information about specific sounds was found across levels of the hierarchy, but more abstract representations were absent; and during auditory imagery, abstract coding was found even in earlier levels of the hierarchy, but detailed sensory information was not. It could be argued that imagery simply evokes a less precise representation than memory maintenance despite stronger overall activity during imagery. However, given the lack of perceptual similarity across items within a category, there would be no reason to expect a less precise physical representation to allow for the presence of a categorical code. It is, however, possible that the memory task encouraged finer coding of perceptual details compared to the imagery task. This only strengthens the conclusion that neural representations in AC are flexible and task dependent.

The difference in activation and the informational content of neural activity patterns during self-reported imagery and STM maintenance is particularly striking as many previous studies have implicitly or explicitly made the assumption that imagery and rehearsal rely on the same cognitive and neural mechanisms (Kaiser et al., 2010). From our group analysis results, however, it becomes clear that this is not the case even when the tasks and stimuli are closely matched. Furthermore, in both tasks, the sound to be held in STM and the sound to be imagined were played to the participant right before the delay period started. After the delay ended, participants compared a second presentation of the sound to the first in the change detection task, whereas in the imagery task, participants rated how clear their imagery had been. The imagery task, therefore, also contained a memory component, that is, in order to respond participants had to compare their imagery to the memory of the actual sound. This makes the two tasks even more similar, yet differences between the tasks are obvious in the univariate as well as multivariate analysis, indicating that neural mechanisms differ substantially between the two. Importantly, activity also goes into opposite directions with some of the regions activated during the imagery delay being suppressed during STM maintenance. This as well as the behavioral ratings of the perceived difficulty and attentiveness during the imagery and STM tasks that participants completed after the main experiment indicates that differences in activity during the delay periods of the two tasks were not due to differences in difficulty or attention. Although impossible to perform an imagery task without some involvement of memory, from these results it appears that self-reported imagery is most likely dependent on top–down, long-term rather than purely STM representations (for a discussion, see Hubbard, 2010) whereas change detection, necessary for keeping track of changes in the perceptual scene, is performed fairly automatically (Demany et al., 2010) and on an ad hoc basis that requires the short-term storage of information that new information can be compared to.

Previous studies have implied many different regions to be involved in auditory STM (Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003) that we did not observe in the current study. However, these studies commonly use stimuli that are easy to vocalize (such as musical sequences or speech) and instruct participants to actively rehearse. They are, thus, much more similar to the imagery condition in the current study. Our results imply that sustained, stimulus-specific coding in auditory regions is sufficient to maintain sounds in STM, in accordance with previous studies showing that auditory changes can be detected without much conscious effort (Linke et al., 2011; Demany et al., 2010; Pavani & Turatto, 2008).

Lastly, the current study has important implications for studying internally generated mental representations. Using complex stimuli that can be compared on physical features as well as abstract semantic characteristics, we were able to address the century-old question of whether imagery relies on veridical or abstracted mental representations. Our results support Pylyshyn's theory that self-reported imagery is not a precise reconstruction of perception but a top–down, cognitive process that draws on the abstract knowledge about what is being "imagined." This does not only have implications for the study of imagery. The ability of the human mind to consciously and vividly relive previously experienced and envision future scenarios makes it possible to extend sensory reality in a way not many other species are thought capable of and is implicated in a wide range of mental processes such as word learning, spatial navigation, and problem solving. Individuals vary widely in their reported ability to engage in imagery and hold information in STM, which impacts performance on other cognitive tasks. The ability to imagine natural sounds in postlingually deaf patients, for example, is predictive of clinical outcome after fitting a cochlear-implant that partially restores hearing (Lazard, Giraud, Truy, & Lee, 2011). Conversely, uncontrolled imagery, such as auditory hallucinations in schizophrenia, which evoke activity in AC (Dierks et al., 1999), can be extremely disruptive. This study was designed for group analysis, but the trends from our additional individual differences analysis imply that it is important to study not just how individuals and patients differ in which networks of the brain they recruit during imagery but also what information about a stimulus they are processing.

In summary, our results show that the information present in activity patterns in human AC is flexible on a short timescale and that the degree to which representations adapt to task demands reflects individual differences in performance. This questions whether specific information coding can easily be mapped to a particular region of the brain without taking task demands into account and highlights the highly plastic and flexible nature of neural coding in the human brain.

Acknowledgments

This work was supported by the Medical Research Council (UK).

Reprint requests should be sent to Annika Linke, The Brain and Mind Institute, The Natural Sciences Centre, Western University, London, ON, N6A 5B7, Canada, or via e-mail: annikalinke@gmail.com.

REFERENCES

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23,* 1427–1431.

Behrmann, M., Winocur, G., & Moscovitch, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature, 359,* 636–637.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 443–446.

Brett, M., Anton, J. L., Valbregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage, 16,* 1140–1141.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.-J., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *Neuroimage, 26,* 1119–1127.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural Brain Sciences, 24,* 87–114.

De Vreese, L. P. (1991). Two systems for colour-naming defects: Verbal disconnection vs colour imagery disorder. *Neuropsychologia, 29,* 1–18.

Demany, L., Semal, C., Cazalets, J., & Pressnitzer, D. (2010). Fundamental differences in change detection between vision and audition. *Experimental Brain Research, 203,* 261–270.

Dierks, T., Linden, D. E. J., Jandi, M., Formisano, E., Goebel, R., Lanfermann, H., et al. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron, 22,* 615–621.

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition, 18,* 245–272.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory—An fMRI study with sparse temporal sampling. *Neuroimage, 19,* 1417–1426.

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

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458,* 632–635.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Herholz, S. C., Halpern, A R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience, 24,* 1382–1397.

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin, 136,* 302–329.

Kaiser, S., Kopka, M.-L., Rentrop, M., Walther, S., Kronmüller, K., Olbrich, R., et al. (2010). Maintenance of real objects and their verbal designations in working memory. *Neuroscience Letters, 469,* 65–69.

Kosslyn, S. (2003). Mental imagery: Against the nihilistic hypothesis. *Trends in Cognitive Sciences, 7,* 109–111.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2,* 635–642.

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

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.

Lazard, D. S., Giraud, A. L., Truy, E., & Lee, H. J. (2011). Evolution of non-speech sound memory in postlingual deafness: Implications for cochlear implant rehabilitation. *Neuropsychologia, 49,* 2475–2482.

Lee, S.-H., Kravitz, D. J., & Baker, C. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience, 16,* 997–999.

Linke, A. C., Vicente-Grabovetsky, A., & Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences, 108,* 12961–12966.

Meyer, K., Kaplan, J. T., Essex, R., Webber, C., Damasio, H., & Damasio, A. (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature Neuroscience, 13,* 667–668.

Nelken, I. (2008). Processing of complex sounds in the auditory system. *Current Opinion in Neurobiology, 18,* 413–417.

Oh, J., Kwon, J. H., Yang, P. S., & Jeong, J. (2013). Auditory imagery modulates frequency-specific areas in the human auditory cortex. *Journal of Cognitive Neuroscience, 25,* 175–187.

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience, 6,* 97–107.

Pavani, F., & Turatto, M. (2008). Change perception in complex auditory scenes. *Perception & Psychophysics, 70,* 619.

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *Neuroimage, 52,* 1410–1419.

Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience. A final summary and discussion. *Brain, 86,* 595–696.

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences, 7,* 113–118.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15,* 1261–1269.

Schmitter, S., Diesch, E., Amann, M., Kroll, A., Moayer, M., & Schad, L. R. (2008). Silent echo-planar imaging for auditory fMRI. *Magma, 21,* 317–325.

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences, 18,* 82–89.

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 13,* 1–7.

Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: Event-related functional MRI study. *NeuroReport, 12,* 3045–3049.

Young, A. W., Humphreys, G. W., Riddoch, M. J., Hellawell, D. J., & de Haan, E. H. (1994). Recognition impairments and face imagery. *Neuropsychologia, 32,* 693–702.

Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia, 31,* 221–232.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron, 47,* 9–12.

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience, 37,* 1421–1434.