



Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes

James W. Lewis^{1,2,3*}, William J. Talkington^{1,2,3}, Katherine C. Tallaksen^{2,4} and Chris A. Frum^{1,2,3}

¹ Center for Neuroscience, West Virginia University, Morgantown, WV, USA

² Center for Advanced Imaging, West Virginia University, Morgantown, WV, USA

³ Department of Physiology and Pharmacology, West Virginia University, Morgantown, WV, USA

⁴ Department of Radiology, West Virginia University, Morgantown, WV, USA

Edited by:

Raphael Pinaud, Northwestern University, USA

Reviewed by:

Sundeeep Teki, University College London, UK

Hirohito M. Kondo, NTT Corporation, Japan

*Correspondence:

James W. Lewis, Department of Physiology and Pharmacology, West Virginia University, PO Box 9229, Morgantown, WV 26506, USA.
e-mail: jwlewis@hsc.wvu.edu

Whether viewed or heard, an object in action can be segmented as a distinct salient event based on a number of different sensory cues. In the visual system, several low-level attributes of an image are processed along parallel hierarchies, involving intermediate stages wherein gross-level object form and/or motion features are extracted prior to stages that show greater specificity for different object categories (e.g., people, buildings, or tools). In the auditory system, though relying on a rather different set of low-level signal attributes, meaningful real-world acoustic events and “auditory objects” can also be readily distinguished from background scenes. However, the nature of the acoustic signal attributes or gross-level perceptual features that may be explicitly processed along intermediate cortical processing stages remain poorly understood. Examining mechanical and environmental action sounds, representing two distinct non-biological categories of action sources, we had participants assess the degree to which each sound was perceived as object-like versus scene-like. We re-analyzed data from two of our earlier functional magnetic resonance imaging (fMRI) task paradigms (Engel et al., 2009) and found that scene-like action sounds preferentially led to activation along several midline cortical structures, but with strong dependence on listening task demands. In contrast, bilateral foci along the superior temporal gyri (STG) showed parametrically increasing activation to action sounds rated as more “object-like,” independent of sound category or task demands. Moreover, these STG regions also showed parametric sensitivity to spectral structure variations (SSVs) of the action sounds—a quantitative measure of change in entropy of the acoustic signals over time—and the right STG additionally showed parametric sensitivity to measures of mean entropy and harmonic content of the environmental sounds. Analogous to the visual system, intermediate stages of the auditory system appear to process or extract a number of quantifiable low-order signal attributes that are characteristic of action events perceived as being object-like, representing stages that may begin to dissociate different perceptual dimensions and categories of every-day, real-world action sounds.

Keywords: signal feature extraction, motion processing, auditory perception, functional MRI, natural sound categorization, entropy, spectral structure variation

INTRODUCTION

For sensory systems, feature extraction models (Laaksonen et al., 2004) represent potential neuronal mechanisms that may develop to efficiently segment and distinguish objects or events based on salient features and components within a scene. Through experience with visual and acoustic scenes, semantically related object groupings or classes of behaviorally relevant objects and/or events (Rosch, 1973; Minda and Ross, 2004) may then become differentially mapped and self-organized across cortical network representations. This in part may lead to the development of cortical regions showing preferential or selective activation to

the various visual and auditory “object categories” reported to date.

In the visual system, several brain regions are reported to be sensitive or selective for different object categories, including human faces (Allison et al., 1994; Kanwisher et al., 1997; McCarthy et al., 1997), animal faces (Mormann et al., 2011; Rutishauser et al., 2011), scenes or places (Epstein and Kanwisher, 1998; Gron et al., 2000), human body parts (Downing et al., 2001), buildings (Hasson et al., 2003), or animals versus tools (Chao and Martin, 2000; Beauchamp et al., 2002). In contrast to object processing, other brain regions (e.g., parahippocampal,

retrosplenial, and some occipital areas) are more sensitive to processing visual scenes (Epstein and Kanwisher, 1998; Epstein et al., 2007; Epstein and Morgan, 2011). However, preceding many of these scene- or object-sensitive stages in cortex are earlier stages that incorporate relatively low-level visual features such as motion and form. For instance, the posterior superior temporal sulci (pSTS) are preferentially activated by biological motion (Johansson, 1973) versus rigid body motion attributes (Frith and Frith, 1999; Lewis et al., 2000; Beauchamp et al., 2002; Pelphrey et al., 2004), which contributes to the segmentation of animate versus inanimate objects. Additionally, portions of the lateral occipital cortices (LOC) are preferentially responsive to object forms as opposed to textures or visual noise patterns, which are otherwise matched for low-level features such as brightness, contrast, and spatial frequencies (Malach et al., 1995; Kanwisher et al., 1996). Portions of the LOC also show relatively invariant responses to object size and/or location in the visual field (Grill-Spector et al., 1998, 1999; Tootell et al., 1998; Doniger et al., 2000; Kourtzi and Kanwisher, 2000). Hence, the pSTS and LOC regions appear to house hierarchically intermediate processing stages or channels for analyzing gross-level visual objects or object-like features by assimilating inputs from earlier areas that represent a variety of low-level visual attributes. This hierarchical processing may thus contribute to the segmentation of a distinct object, or objects, present within a complex visual scene (Felleman and van Essen, 1991; Macevoy and Epstein, 2011).

Parallel processing hierarchies are also known to exist in the primate auditory system (Rauschecker et al., 1995; Kaas et al., 1999). Primary auditory cortical regions (PACs) are known to have a critical role in auditory stream segregation and formation, clustering operations, and sound organization based on primitive acoustic features such as bandwidths, spectral shapes, onsets, and harmonic relationships (Medvedev et al., 2002; Nelken, 2004; Kumar et al., 2007; Elhilali and Shamma, 2008; Woods et al., 2010). The left and right planum temporale (PT) in humans, located posterior and lateral to Heschl's gyrus (HG), are thought to represent subsequent processing stages comprised of computational hubs that segregate spectro-temporal patterns associated with complex sounds, including processing of acoustic textures, location cues, and prelinguistic analysis of speech sounds (Griffiths and Warren, 2002; Obleser et al., 2007; Overath et al., 2010). Subsequent cortical pathways are thought to integrate corresponding acoustic streams over longer time frames, including the posterior portions of the superior temporal gyri (STG) and sulci (STS), which represent processing stages more heavily involved in discriminating and recognizing acoustic events and real-world sounds (Maeder et al., 2001; Zatorre et al., 2004; Griffiths et al., 2007; Leech et al., 2009; Goll et al., 2011; Teki et al., 2011). Additionally, sounds containing vocalizations (human or animal) or strong harmonic content evoke activity along various bilateral STG pathways, which subsequently feed into regions that are relatively specialized for processing speech and/or prosodic information [Zatorre et al., 1992; Obleser et al., 2008; Lewis et al., 2009; Rauschecker and Scott, 2009; Leaver and Rauschecker, 2010; Talkington et al. (in press)].

Many of the above cortical mapping studies have been conducted using stimuli that capture the spectro-temporal

characteristics of natural sounds in an effort to define mechanisms that abstract behaviorally meaningful events. However, given the broader multisensory and supramodal nature of object knowledge representations (Caramazza and Mahon, 2003; Martin, 2007; Lewis, 2010), the concept of an "auditory object" is convenient for more generally addressing issues related to hearing perception and cognition. While its definition remains operational, one principle of auditory object processing is that auditory pattern analyses should allow for perceptual categorization and that auditory objects should be separable by perceptual boundaries (Griffiths and Warren, 2004; Husain et al., 2004). However, beyond representations of components of speech and speech-like sounds, identifying other "bottom-up" acoustic signal attributes and perceptual dimensions that may be used for distinguishing between different real-world sound categories remain poorly understood.

In our earlier studies, we mapped brain regions that were responsive to four distinct semantic ("top-down") categories of behaviorally relevant real-world *action* sounds (devoid of any vocalization content). This included two categories of biological (living) action sounds, human and animal sources, and two categories of non-biological (non-living) action sounds, mechanical, and environmental sources (Engel et al., 2009; Lewis et al., 2011). For the present study, we assumed that the five aforementioned conceptual categories of sound (vocalizations plus four action sound categories) may also be characterized by quantifiable acoustic features. Re-analyzing data from our earlier study (Engel et al., 2009), we focused on examining perceptual features and acoustic signal attributes of the non-biological action sound sources. This included automated machinery (actions perceived as not being directly associated with a human or agent instigating the action) and the natural environment (see **Table A1**).

We restricted our analyses to non-biological action sounds because high-level acoustic features associated with biological action sounds can be strongly tied to motor and multisensory associations (for review see Lewis, 2010). Meaningful biological action sounds may ultimately be processed along specialized pathways that extract or probabilistically compare their acoustic features with representations of the observer's own networks related to sound-producing motor actions (Rizzolatti et al., 1998; Kohler et al., 2002; Rizzolatti and Craighero, 2004), evoking "embodied" representations (Barsalou, 2008) and assessments of motor action intention (Aziz-Zadeh et al., 2004; Bidet-Caulet et al., 2005; Iacoboni et al., 2005; Gazzola et al., 2006; Lewis et al., 2006; Aglioti et al., 2008; de Lucia et al., 2009).

One salient feature of the mechanical and environmental sounds we previously examined was their wide range in spatial scale (Lewis et al., 2011). While there were exceptions, most of the mechanical sounds depicted discrete "object-like" things (e.g., clock, fax machine, laundry machine) while most of the environmental sounds depicted an acoustic scene on a large-scale relative to the size of the observer (e.g., wind, rain, ocean waves). This observation led us to question whether an object-like to scene-like perceptual continuum or boundary might be explicitly represented along intermediate processing stages of the human auditory system, analogous to the parallel hierarchical organizations reported for the visual system. Thus, our first objective

was to test the hypothesis that the auditory system would house intermediate cortical processing stages or channels that are parametrically sensitive to signal attributes characteristic of object-like versus scene-like action sounds. We further hypothesized that any regions sensitive to object-like acoustic features would be located outside of earlier primary auditory cortices (PACs) yet prior to stages sensitive to different “conceptual-level” representations of real-world sound-source categories that we and others have previously reported.

Assuming that some cortical regions would show either parametric sensitivity or a sharp categorical boundary to object-like versus scene-like non-biological action sounds, a second objective of this study was to identify specific acoustic signal attributes that might quantitatively characterize this perceptual dimension. Environmental sounds have previously been modeled as distinguishable sound textures using relatively simple time-averaged statistics (McDermott and Simoncelli, 2011). Additionally, quantitative characterizations using measures of spectral dynamics are reported to represent a possible scheme for categorizing natural sounds (Reddy et al., 2009). Thus, we further hypothesized that some of these relatively low-order signal attributes of our ecologically valid sound stimuli would show a parametric correlation with the perceptual ratings of object saliency and/or the activation of cortical regions sensitive to sounds rated more as object-like versus scene-like.

MATERIALS AND METHODS

PARTICIPANTS

The functional magnetic resonance imaging (fMRI) data for this study draws from earlier publications (Engel et al., 2009; Lewis et al., 2011), which provide additional details of the sound stimuli, psychophysical attributes of the sounds, and imaging methods used. For the present study, we included neuroimaging results from 31 right-handed participants (19–36 years of age, 16 women). All participants were native English-speakers with no previous history of neurological or psychiatric disorders, or auditory impairment, and had a self-reported normal range of hearing. Informed consent was obtained for all participants following guidelines approved by the West Virginia University Institutional Review Board.

SOUND STIMULUS CREATION AND PRESENTATION

The sound stimuli were compiled from professionally recorded action sounds (Sound Ideas, Inc, Richmond Hill, ON, Canada) including 64 sounds in each of four conceptual categories of sound sources (human, animal, mechanical, and environmental). The mechanical and environmental sounds retained for primary analyses in the present study are included in **Table A1**, and a complete list of the sounds is detailed in our earlier study (Engel et al., 2009). Sound stimuli were edited to 3.0 ± 0.5 s duration, matched for total root mean-squared (RMS) power, with 25 ms onset/offset ramps (Cool Edit Pro, Syntrillium Software Co., owned by Adobe). Sound stimuli were retained from one channel (mono, 44.1 kHz, 16-bit), and these single channel stimuli were used for acoustic signal processing analyses. For participants, monaural sounds were presented to both ears, which precluded the presence of binaural spatial cues, yet allowed the sounds

to be heard more clearly. During fMRI scanning, high fidelity sound stimuli were presented using a Windows PC computer (Presentation software version 11.1, Neurobehavioral Systems Inc.) and delivered via MR compatible electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax LTD., Gardena, CA) worn under sound attenuating ear muffs.

SCANNING PARADIGMS

Each scanning session consisted of eight separate functional imaging runs, across which the sound stimuli and silent events were presented in random order. Participants randomly assigned to Group A ($n = 12$) were instructed to press a response box button immediately at the offset of each sound stimulus (from Engel et al., 2009). They were unaware of the purposes of the study and had not heard these particular sound stimuli before. Participants in Group B ($n = 19$), also unfamiliar with the specific sound stimuli, were instructed to silently determine in their head (no overt responses) whether or not a human was directly involved with the production of the action sound (from Engel et al., 2009 and Lewis et al., 2011). Based on post-scanning assessments by participants, we censored responses to 45 of the 256 sound stimuli *post-hoc* for all participant data-sets to be certain that the sounds fell clearly within a given category and were perceived to be devoid of any vocalization content. Brain responses to sounds that were incorrectly categorized, based on the individual's scanning responses (Group B) or post-scanning responses (Group A), were excluded from all analyses for that individual. Additionally, the mean entropy or spectral structure variation (SSV) measures could not be derived for some sound stimuli (see below), and responses to those sounds were excluded from all analyses.

MAGNETIC RESONANCE IMAGING AND DATA ANALYSIS

Scanning was completed on a 3 Tesla General Electric Horizon HD MRI scanner using a quadrature bird-cage head coil. We acquired whole-head, spiral in-and-out images of blood-oxygenated level dependent (BOLD) signals (Glover and Law, 2001) using a clustered-acquisition fMRI design. This allowed sound stimuli to be presented during silent periods (at a comfortable level between 80–83 dB C-weighted) without the presence of scanner noise (Edmister et al., 1999; Hall et al., 1999). A sound or silent event occurred every 9.3 s. At 6.8 s after event onset BOLD signals were collected as 28 axial brain slices with $1.9 \times 1.9 \times 4$ mm³ spatial resolution (TR = 9.3 s, TE = 36 ms, OPTR = 2.3 s volume acquisition, FOV = 24 cm). In a subsequent imaging sequence, whole brain T1-weighted anatomical MR images were collected using a spoiled GRASS pulse sequence (SPGR, 1.2 mm slices with 0.94×0.94 mm² in-plane resolution).

Acquired data were analyzed using volumetric-based registration techniques with AFNI software (<http://afni.nimh.nih.gov/>) and related plug-ins (Cox, 1996). For each participant's data, the eight scans were concatenated into a single time series and brain volumes were corrected for baseline linear drift and for global head motion translations and rotations. BOLD signals were normalized to a percent signal change on a voxel-by-voxel basis relative to responses to the silent events that were presented randomly throughout each scanning run (Belin et al., 1999; Hall et al., 1999). Several multiple linear regression models (using

3dDeconvolve) identified voxels showing preferential activation related either to the Likert scale object-vs.-scene ratings of sounds, the category of sound, or parametric measures of acoustic signal attributes (addressed below). Regression coefficients were spatially low-pass filtered (4 mm box filter), and subjected to *t*-test and thresholding.

For whole-brain correction, we estimated the spatial structure of the noise in BOLD signal in voxels outside the brain (using AFNI plug-ins AlphaSim and 3dFWHMx) after the residuals left over from linear modeling fitting was subtracted from each voxel's time series. This yielded an estimated $2.0 \times 2.1 \times 3.4 \text{ mm}^3$ spatial smoothness in *x*, *y*, and *z* dimensions (full-width half-max Gaussian filter widths). Using the estimated 2.4 mm^3 spatial blur in brain voxels, together with a minimum cluster size of 20 voxels, and voxel-wise *p*-value of $p < 0.05$ yielded a whole-brain correction at $\alpha < 0.05$. Anatomical and functional imaging data were transformed into standardized Talairach coordinate space (Talairach and Tournoux, 1988). Data were then projected onto the PALS atlas cortical surface models (in AFNI-*tlrc*) using Caret software (<http://brainmap.wustl.edu>) (van Essen et al., 2001; van Essen, 2005).

ACOUSTIC SIGNAL ATTRIBUTES OF MECHANICAL AND ENVIRONMENTAL SOUNDS

The mechanical and environmental action sounds retained for analyses in the current study had been matched overall for low-level acoustic attributes including loudness (RMS intensity) and duration ranges. To assess changes in the spectro-temporal dynamics of the action sounds, we measured the mean entropy (Wiener entropy) in the acoustic signal (Tchernichovski et al., 2001) using freely available phonetic software (Praat, <http://www.fon.hum.uva.nl/praat/>). We further derived the SSVs of the sounds (using Praat), which is a measure of changes in signal entropy over time that has been shown to have utility in categorizing natural sound signals (Reddy et al., 2009). The natural log of SSV measures provided a more widespread distribution of values relative to the Likert scale ratings, and thus we used $\ln(\text{SSV})$ values for linear regression analyses. Both the entropy and $\ln(\text{SSV})$ measures were *z*-normalized based on the mean and standard deviation of the entropy measures $[(x - \mu)/\sigma]$ of the retained mechanical and environmental sounds.

PERCEPTUAL ATTRIBUTES OF SOUND STIMULI

All of the 64 mechanical and 64 environmental sound stimuli were presented in random order to a group of participants ($n = 18$) not included in the fMRI scanning paradigms. They rated the sounds using a Likert scale (1–5) with written responses, assessing the degree to which they perceived the sound-source as a distinct object (low rating) versus part of an acoustic scene (high rating). As examples, they were instructed that hearing the hum of traffic when you are in a neighborhood that is near an interstate highway might be rated more as an acoustic scene (response 4 or 5), whereas hearing a stopwatch ticking might be perceived more as a distinct object (response 1 or 2). The ratings were averaged across the group (Figure 1A). Seven of the environmental sounds rated as object-like (Figures 1A,B) fell below the overall average Likert ratings of 3.08. Using this number of

sounds as a threshold, we opted to identify cortical regions most sensitive to the object-vs.-scene perceptual dimension by examining (1) seven extreme object-like environmental (EO7) sounds versus seven extreme scene-like mechanical (MS7) sounds, and conversely (2) cortical responses to the seven extreme object-like mechanical (MO7) sounds versus the seven extreme scene-like environmental (ES7) sounds (28 sounds total, see Table A1 bold text entries). To validate the reliability of the Likert ratings of the retained 54 mechanical and 57 environmental sounds (Table A1) we calculated Cronbach's alpha scores (Cronbach, 1951) using multivariate methods (JMP 9.0 software, SAS Institutes, Inc.). Including ratings of all 111 sounds (54 mechanical plus 57 environmental) by the entire set of 18 participants yielded a value of 0.9474. As a more conservative measure, including only the 28 most extreme object-like and scene-like sounds (mentioned above) yielded a value of $\alpha = 0.9784$, and subsequent removal of each participant individually from the group data consistently produced values between 0.9763 and 0.9784, which were well above the accepted consistency score of 0.7 (Nunnally, 1978).

RESULTS

In our earlier studies examining these same data we reported that the medial two-thirds of HG, the approximate location of PACs, were strongly activated by both the mechanical and environmental sound stimuli; there was no differential activation to these different conceptual categories of sound in these regions (Engel et al., 2009; Lewis et al., 2011). Rather, mechanical action sounds preferentially activated the bilateral anterior superior temporal gyri (aSTG) and parahippocampal regions, while environmental action sounds preferentially activated bilateral medial prefrontal cortices, precuneus, retrosplenial cortex, and the right hemisphere visual motion processing area hMT/V5 (Engel et al., 2009; Lewis et al., 2011). For the present study, we examined cortical responses to the same mechanical and environmental sound stimuli but “re-grouped” them according to their perceptual ratings along a putative continuum of object-like to scene-like; psychophysical ratings of the mechanical and environmental sounds were derived from non-imaging listeners ($n = 18$) who rated the sounds on a Likert scale (Figure 1A; range 1 = object-like to 5 = scene-like; refer to Methods).

To assess extremes in response to the object-like versus scene-like sounds, we charted the power spectra of the 28 most extreme-rated sounds for each category (Figure 1B; seven in each subset, see Methods). Inspection of these spectra revealed greater roughness of the contours for the sounds rated as more object-like and smoother contours for the sounds rated as more scene-like. We averaged the power spectra of each of these four subsets of sound (not shown) and fit them with a logarithmic function ($y = a \times \ln(x) + b$). This revealed a systematic increase in the amplitude of the slope of the exponential fit with increasing scene-like ratings (Figure 1B, the value of “*a*” shown in parentheses). These power spectrum features are addressed later in the context of signal attribute processing (see Discussion).

We mapped regions showing significantly preferential activity to the 28 action sounds that were rated at the extremes of the object-to-scene perceptual dimension. Our first analysis entailed a conjunction contrasting (1) the seven mechanical action sounds

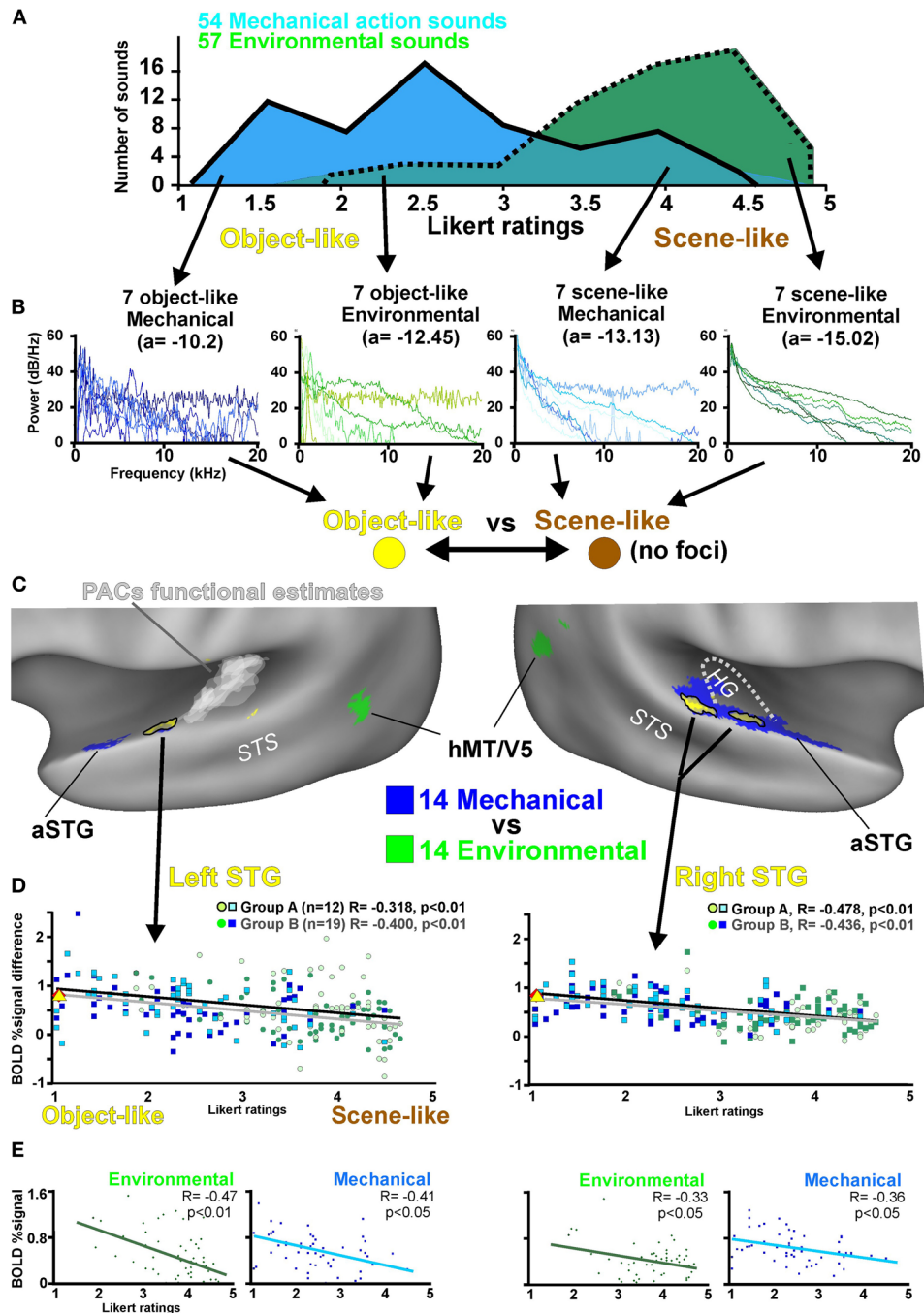


FIGURE 1 | Cortical sensitivity to the perception of auditory “objects” versus acoustic scenes, using real-world non-biological action sounds. (A) Frequency of Likert ratings (1–5) of the Mechanical (M; blue, $n = 54$ sounds retained) and Environmental sound stimuli (E; green, $n = 57$). See **Table A1** bolded entries for a list of these sounds. **(B)** Power spectra of the 28 action sounds with the most extreme object-vs-scene group-averaged activation in each conceptual category of action sound (refer to color key). **(C)** Volume-based group-averaged activation common to both Groups A and B (conjunction analyses; yellow with black outlines) that showed preferential activation to sounds judged to be object-like (MO7 and EO7) versus scene-like (MS7 and ES7). Cortical responses to the same sounds were used to define regions preferential for mechanical (blue) versus environmental (green) sounds. Transparent white

patches in the left hemisphere depict an overlapping “heat map” of tonotopically organized regions (disregarding orientation of the tonotopic gradient) derived from eight individuals. STS = superior temporal sulcus. **(D)** Charts illustrating the BOLD percent signal change response profiles as a function of Likert scale rating for both Groups (refer to color key). Blue squares depict mechanical sounds and green circles depict environmental sounds. The group-averaged BOLD percent signal change responses to the human action sounds (red diamonds; left STG 0.62% BOLD signal differential, right 0.73%) and animal action sounds (yellow triangle; left 0.61%, right 0.72%) are also depicted for comparison. **(E)** Charts separately illustrating BOLD responses to environmental and mechanical action sounds as a function of Likert scale ratings. Refer to text for other details.

(**Table A1**) rated as being the most object-like (Likert rating range of 1.1–1.4; dark blue traces in **Figure 1B**) versus the seven environmental sounds that were most scene-like (range 4.5–4.7; dark green), together with (2) regions sensitive to the seven environmental sounds that were most object-like (range 1.9–2.8; light green) versus the seven mechanical sounds that were rated as most scene-like (range 3.6–4.5; light blue). Thus, for the fMRI participants the cortical responses to sounds generally judged as being object-like versus scene-like were balanced for correctly categorized sound source stimuli, mechanical, or environmental.

The above fMRI analysis had been conducted for two different groups of listeners: Group A participants ($n = 12$) pressed a button as quickly as possible immediately at the end of each sound, and Group B participants ($n = 19$) silently responded in their head whether or not the sound was directly produced by a human (no overt responses). Both groups of listeners revealed significant bilateral activation along the STG that was preferential for sounds perceived as object-like as opposed to scene-like, independent of the category of sound (data not shown). Consequently, we combined those data-sets using a second conjunction analysis to reveal activation foci common to both Groups A and B (**Figure 1C**, yellow with black outlines), which provided a more conservative localization of cortical regions showing sensitivity to object-like sounds, independent of sound category and listening task.

These auditory object-sensitive STG foci (Talairach coordinates: left STG $x = -54$, $y = -12$, $z = 1$, volume = 148 μl ; right STG 54, -21 , 7, 783 μl) fell well outside of the estimated locations of primary auditory cortices (PACs), which are typically located along the medial two-thirds of HG (**Figure 1C**, right hemisphere dotted white line) (Morosan et al., 2001; Rademacher et al., 2001). We additionally charted the functionally estimated locations of PACs of eight participants incorporating results from our earlier frequency-dependent response (“tonotopy”) mapping studies (**Figure 1C**, left hemisphere white heat map) using the same MRI scanner and same basic clustered acquisition fMRI design [Lewis et al., 2009; Talkington et al. (in press)]. This further indicated that the STG foci were outside of primary auditory cortices, which were functionally defined here as contiguous stretches of cortex that were differentially responsive to high, medium, and low frequency pure tones and band pass noises.

We also charted cortex preferential for the 14 mechanical versus 14 environmental action sounds (from **Figure 1B**), which revealed regions more sensitive to category membership at a conceptual level (**Figure 1C**, blue versus green regions). While the 14 mechanical sounds were overall more object-like than the 14 environmental sounds, there nonetheless was a double dissociation that supported our earlier finding. In particular, the anterior portions of the left and right STG (aSTG) were preferentially activated by the mechanical action sounds, and the hMT/V5 region, among other cortices, were preferentially activated by the environmental action sounds. Thus, the STG foci sensitive to sounds rated more as object-like (yellow) were in locations distinct from many of the regions that were preferential for environmental (green) or mechanical (blue) action sounds at a categorical level. While this 2×2 analysis design was inherently non-orthogonal (using the same four subsets of sound), both the anatomical

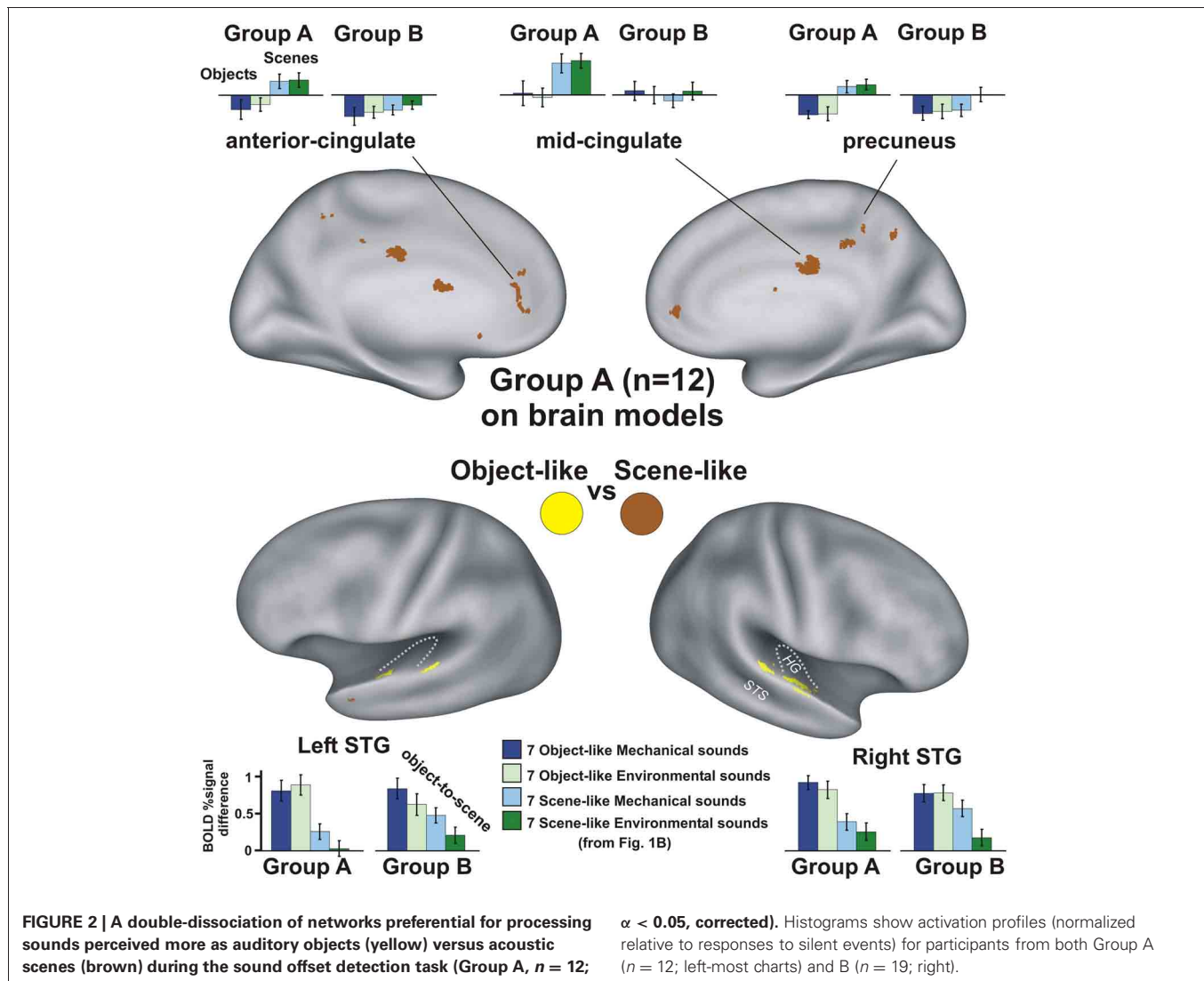
and functional placement of the bilateral STG preferential for object-like qualities were consistent with representing intermediate processing stages within the cortical networks subserving hearing perception (see Discussion).

Using the STG foci as regions of interest, we next charted the averaged BOLD signal response (across all subjects; $n = 31$) relative to the Likert scale rating of each sound (**Figure 1D**). These results further indicated that a roughly *linear* parametric correlation with the left and right STG activation existed, which was greater for object-like sounds and lower for scene-like sounds for both Group A (right STG yielded $R = -0.478$, Steiger’s Z -test 111 df, $Z = 3.72$, $p < 0.01$; left STG $R = -0.318$, $p < 0.01$) and Group B listeners (right STG $R = -0.436$, $p < 0.01$; left STG $R = -0.400$, $p < 0.01$). This correlation with object-like Likert ratings persisted separately for both mechanical and environmental sound categories (**Figure 1E**), in both the left STG (Environmental sounds, $R = -0.47$, $p < 0.01$; Mechanical sounds $R = -0.41$, $p < 0.01$) and right STG (Environmental sounds, $R = -0.33$, $p < 0.05$; Mechanical sounds $R = -0.36$, $p < 0.01$).

We further assessed cortical activation showing differential BOLD signal in response to the remaining four pairings of four extreme-rated sound groups along the object-to-scene continuum (i.e., **Figure 1B** pairs MO7vsEO7, MO7vsMS7, EO7vsES7, and MS7vsES7): For both Groups A and B, these pair-wise comparisons consistently resulted in activation that was either significantly preferential for the more object-like subset of sounds or at least trended toward significance within or near the bilateral STG (data not shown). These differential activation contrasts were generally stronger and more expansive for Group B, who performed a task that required sound categorization. Thus, while the bilateral STG (**Figure 1C**) were significantly more responsive to sounds rated as more object-like for both of our listening tasks, task demands could modulate the relative degree and cortical expanse of activation associated with processing auditory object salience.

Group A participants, who performed a non-categorization task (pressing a button at the end of each sound), revealed a double-dissociation of networks sensitive to object-like versus scene-like action sounds (**Figure 2**, yellow vs. brown; $n = 12$, $\alpha < 0.05$, corrected). Relative to hearing silent events, the scene-like sounds with this task preferentially activated bilateral anterior cingulate (TLRC $x = 0.5$, $y = 41$, $z = 6$, 643 μl), mid-cingulate (2, -24 , 29; 800 μl), and precuneus cortices (2, -49 , 40; 1219 μl) for both the mechanical and environmental sounds (**Figure 2**, light blue and dark green histograms). This double-dissociation did not meet statistical significance in these or any other brain region for Group B (see histograms), who performed the task of indicating if the sounds were directly produced by a human or not—correctly indicating “not” for both the mechanical and environmental sounds based on post-scan testing. Thus, preferential activation to sounds rated as scene-like, in contrast to object-like, depended heavily on task demands.

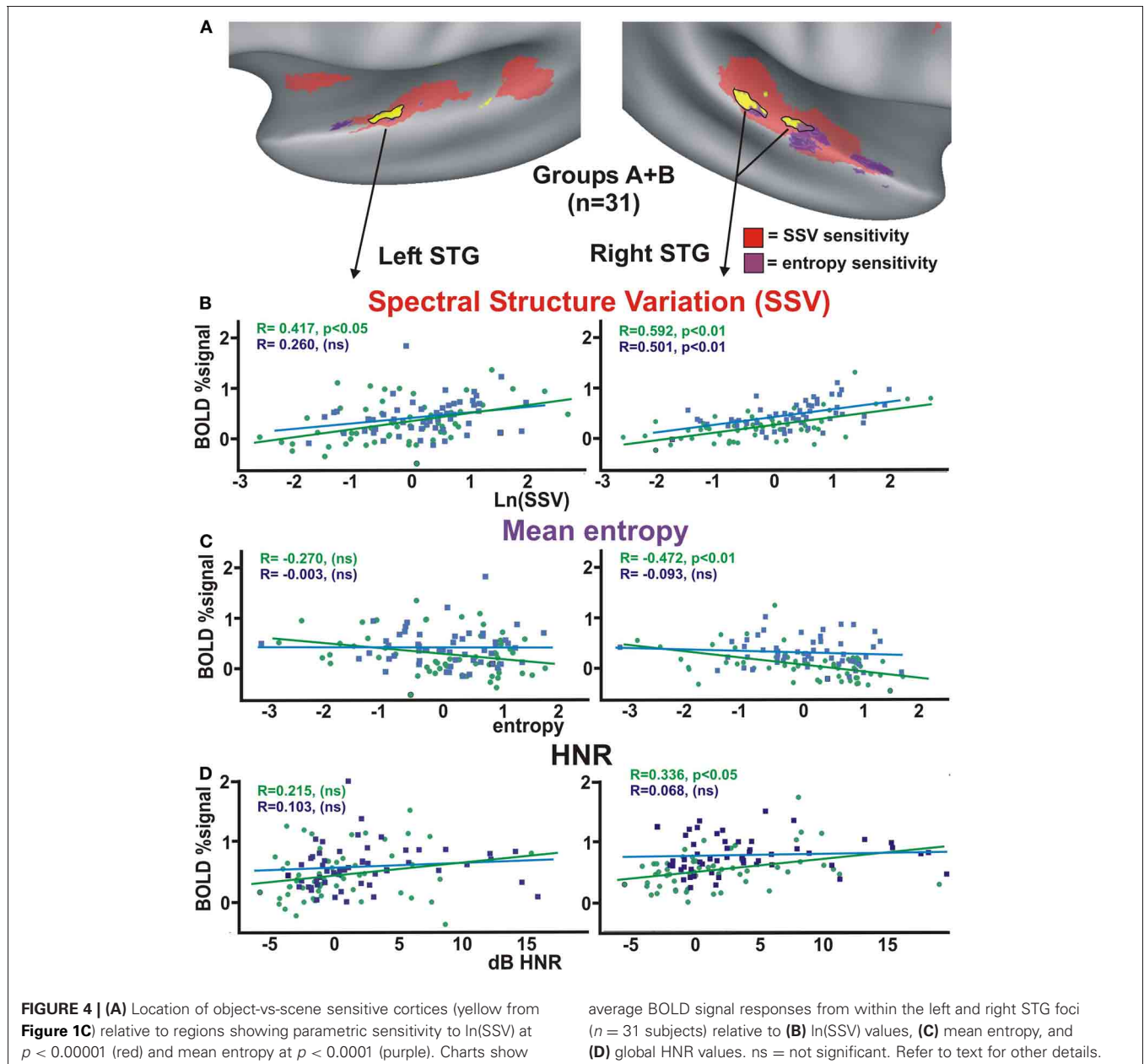
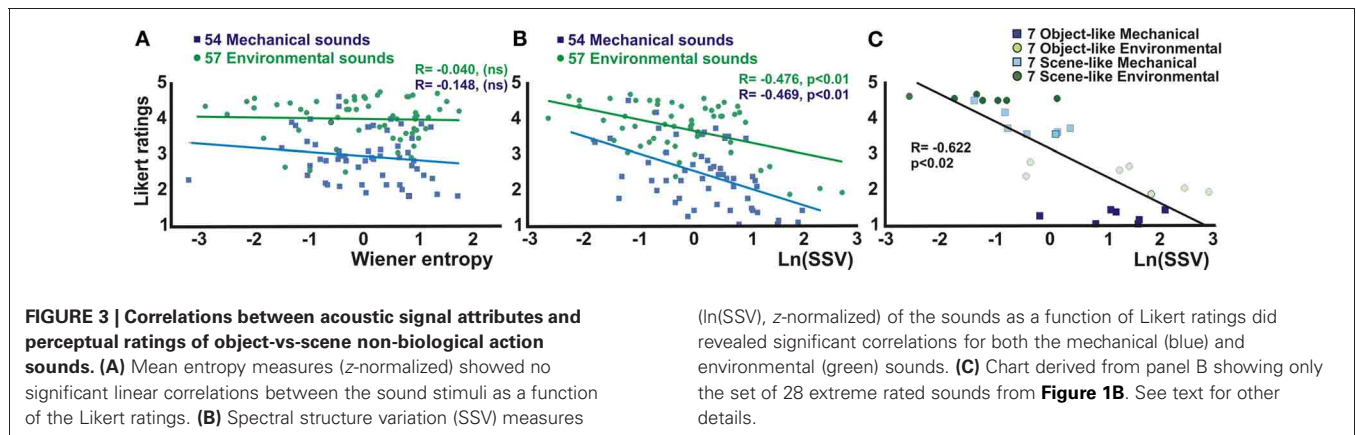
We next sought to identify quantifiable acoustic signal attributes that might correlate with the perception of object-like versus scene-like sound stimuli (Likert ratings) and/or the cortical response profiles of the STG foci depicted in **Figure 1C**. Both the mechanical and environmental action sounds had been



matched in loudness and duration, and binaural spatial cues had been removed from all sound stimuli. Qualitatively, our selection of scene-like sounds tended to be more homogeneous in acoustic temporal structure over time (e.g., the whooshing of wind, or slow droning sound of rainfall) and were characterized by relatively smoother $1/f^\alpha$ structure in their power spectra (see **Figure 1B**), where f = frequency and α ranges from 1 to 2. Inspired by earlier studies, we sought to quantify aspects of these signal features by deriving measures of both mean spectral entropy and changes in entropy dynamics over time (Reddy et al., 2009). Measures of the mean entropy (**Figure 3A**) showed no correlation with the object-like versus scene-like perceptual ratings of the mechanical or environmental sounds. However, changes in entropy over time, quantified by SSV measures, did reveal a significant relationship with the object-to-scene perceptual dimension; this relationship held for both categories of sound when examining all sounds within each category (**Figure 3B**; environmental sounds $R = -0.476$, $p < 0.01$; mechanical sounds $R = -0.469$, $p < 0.01$) or just the 28 extreme-rated sounds

(**Figure 3C**; $R = -0.622$, $p < 0.02$). Further quantification and approaches for assessing the $1/f^\alpha$ signal attributes, or “roughness” distributions (Antal et al., 2002), were beyond the scope of the present study.

Based on the correlations between object-to-scene Likert ratings with SSV signal attributes, we re-analyzed the fMRI data for both Groups A and B testing for regions showing parametric linear sensitivity to SSV of the 54 mechanical and 57 environmental sounds. This parametric fMRI analysis (initially combining data from both groups based on the rationale described for **Figure 1C**) revealed bilateral SSV-sensitive regions (**Figure 4A**, red; $p < 0.00001$, corrected) along large expanses of the superior temporal plane and STG, and this overlapped with the ROIs sensitive to object-like sounds (yellow with black outlines). The right STG focus preferential for object-like sounds showed a significant correlation of increasing activation with increasing SSV measures for both the environmental and mechanical sounds (**Figure 4B**; environmental $R = +0.592$, $p < 0.01$ two-tailed; mechanical $R = +0.501$, $p < 0.01$), while the left STG showed



SSV-sensitivity to the environmental sounds ($R = +0.417$, $p < 0.05$), but only a trend toward SSV-sensitivity for the mechanical sounds. Separately, Group A and B showed a very similar fMRI BOLD response profile to SSV (not shown) for both the environmental action sounds (right STG: Group A, slope = 0.1352, $R = +0.468$, $p < 0.02$; Group B, slope = 0.1589, $R = +0.588$, $p < 0.01$) and mechanical action sounds (Group A, $R = 0.390$, $p < 0.05$; Group B, $R = 0.469$, $p < 0.02$). Thus, task factors did not significantly affect the correlations between SSV measures and the BOLD fMRI responses within the bilateral STG foci.

Parametric sensitivity to mean entropy (Figure 4A, purple; $p < 0.0001$, corrected) was also evident along the bilateral STG (left: -53 , -6 , 5 , $567 \mu\text{l}$, and right: 50 , 3 , -5 and 60 , -13 , 2 , $3326 \mu\text{l}$ combined volume). These foci showed partial overlap with regions identified as being sensitive to object-like sounds (Figure 4A, overlap colors). The right STG foci sensitive to more object-like sounds (yellow with black outlines) showed a significant linear parametric decrease in activation with increasing mean entropy measures of the environmental sounds (Figure 4C; $R = -0.472$, $p < 0.01$), but this did not reach statistical significance for the mechanical action sounds. This result with the environmental sounds held separately for both Groups A (right STG, $R = -0.467$, $df = 57$, $p < 0.02$) and Group B ($R = -0.376$, $p < 0.05$). Thus, the different task demands did not have a strong effect on this basic finding.

We previously assessed human cortex for parametric sensitivity to a harmonics-to-noise ratio (HNR) of vocalizations and artificially constructed sounds, which revealed sensitivity to harmonic content along portions of the bilateral STG (Lewis et al., 2009). The harmonic content of the 54 mechanical action sounds (average = 2.22 ± 4.84 dB HNR; mean plus standard deviation) and 57 environmental sounds (0.23 ± 4.23 dB HNR) did reveal significant differences from one another [t -test(109) = -2.31 ; $p = 0.023$ two-tail]. The non-biological action sounds we examined were substantially lower in HNR measures than typical vocalization sounds (roughly $+4$ to $+20$ dB HNR), thereby precluding a systematic, objective comparison between vocalizations and action sounds. Nonetheless, within the right STG focus for object-like sounds there was a significant correlation of increasing activation with increasing HNR values of the environmental action sounds (Figure 4D).

In sum, a variety of relatively low-level signal attributes (SSV, entropy, and HNR) of real-world sounds showed parametric correlations of cortical activity along various portions of the bilateral STG. Within the STG foci sensitive to object-like perceptual judgments (Figure 4A, yellow), the right hemisphere foci showed a bias for stronger parametric sensitivity to these attributes. Moreover, the SSV measures of our ecologically valid sound stimuli showed a robust correlation with both perceptual ratings along an object-to-scene continuum (Figure 3C) as well as with cortical activation profiles of the left and right STG (Figures 1C, 4A) that were preferentially activated by sounds rated as more object-like.

DISCUSSION

The findings of the present study supported our hypothesis that intermediate stages of auditory cortex are sensitive to an

object-like versus scene-like perceptual dimension of real-world non-biological action sounds. In particular, bilateral STG regions showed increasing parametric sensitivity to action sounds judged as being increasingly more object-like in quality. This parametric activation persisted both for mechanical and environmental sound sources and was independent of listening task. Conversely, cortical regions preferentially activated by scene-like sounds showed dependence on the listening task. This suggested that a double-dissociation of cortical networks representing the perceptual dimension of scene-like to object-like sounds may exist, but depends heavily on top-down task demands rather than solely on bottom-up acoustic signal features inherent to these sounds. An analysis of SSV measures of the object-to-scene perceptual continuum further demonstrated that the bilateral STG regions were parametrically sensitive to quantifiable measures related to acoustic signal entropy. This finding suggests that the STG regions may serve as a general-purpose channel or hub for extracting a number of relatively low-order signal attributes that may alert the auditory system to the presence of a distinct acoustic event, sound source, or “auditory object” emerging from the listener’s ambient acoustic background. Collectively, these results are addressed below in the context of hierarchical processing stages of the auditory system, acoustic scene processing networks, and analogies to visual object processing stages in cortex.

HIERARCHICAL PROCESSING STAGES OF THE AUDITORY SYSTEM

The primary auditory cortices and immediately surrounding regions (e.g., PT) were comparably activated by all of our action sound stimuli (effectively subtracted out in our contrasts, cf. Figures 1, 2); there was no differential activation in these early cortical processing stage regions, neither for the perceptual dimension of object-like versus scene-like sounds nor at a conceptual category level for mechanical versus environmental sound sources. This may partially be a result of either ceiling level BOLD measurement effects, the use of relatively long duration stimuli (~ 3 s), and/or the timing parameters of our fMRI clustered acquisition paradigm. Nonetheless, the results of the present study were consistent with the idea that the PACs and PT represent earlier hierarchical cortical processing stages (see Introduction). Both of these earlier stages may have been performing comparable degrees of processing operations on our mechanical and environmental action sounds, which across categories contained many complex spectro-temporal features and were matched overall for duration and intensity.

Beyond the PACs and PT, the bilateral STG region’s preference for the object-like non-biological action sounds were consistent with depicting higher-order intermediate processing stages. This was due in part to their location, reported circuitry, and response latencies both in non-human primates (Rauschecker et al., 1995; Kaas and Hackett, 1998; Kaas et al., 1999; Rauschecker and Scott, 2009) and humans (Howard et al., 2000; Woods et al., 2010). Additionally, the fMRI activation profiles of the STG foci correlated parametrically with quantifiable acoustic signal features, suggestive of bottom-up influences that may be predominantly associated with auditory (as opposed to multi-sensory or amodal) processing. Although we did not directly

manipulate attentional demands in this study, Group B listeners (who performed a categorization task) versus Group A listeners (who performed an end-of-sound task) did show differences in the expanse and/or relative amplitude of BOLD signal levels in the STG (e.g., **Figure 2**). Hence, the STG were modulated by task demands, consistent with hierarchical placement at intermediate stages of the auditory system (Fritz et al., 2007a,b).

The bilateral STG foci for object-like sounds appeared to represent stages prior to those sensitive to more conceptual-level category network representations. While our analysis examining the 28 extreme-rated sounds for both conceptual category membership and object-vs-scene qualities were not fully independent dimensions, the results nonetheless were consistent with our earlier reports using the full range of action sounds. In particular, portions of the cortical foci located further anterior along the STG (aSTG), plus parahippocampal regions, were preferentially activated by mechanical action sounds relative not only to the environmental sounds (mostly scene-like sounds) but also relative to the object-like human and animal action sound categories (Engel et al., 2009; Lewis et al., 2011). Additionally, as a conceptual-level category, environmental sounds activated various midline cortical regions plus the bilateral visual motion processing areas hMT/V5 (Engel et al., 2009; Lewis et al., 2011). Other studies have reported involvement of the parietal cortices in auditory object detection and segmentation (Cusack, 2005; Dykstra et al., 2011; Teki et al., 2011). Collectively, these findings are consistent with the emerging idea that regions outside the conventional auditory system play a significant role in hearing perception germane to non-vocal action sounds (Lewis et al., 2004). The present results did not address the temporal dynamics of when object-like versus scene-like signal processing was taking place in the aforementioned cortical stages (hierarchically or in parallel). Nonetheless, the above results were consistent with placing the object-like sensitive STG foci at a hierarchically intermediate cortical stage of sound processing in the broader context of multimodal and cognitive networks subserving real-world auditory object recognition and identification. These findings provide new insights regarding how the mammalian auditory system may become organized to efficiently detect a given complex sound stream (an object-like sound) and permit it to pop out from an acoustic background scene, including complex scenes that may be composed of multiple “auditory objects” or sound sources, as addressed next.

ACOUSTIC SCENE PROCESSING

An important role of the properly functioning auditory system is to dynamically filter out the drone of “uninteresting” background acoustic noise (Bregman, 1990). While the scene-like and object-like sound stimuli we used were matched overall in loudness, duration, and spatial location, only the scene-like sounds revealed preferential activation of cortical foci along the midline structures, and only for one of our listening task conditions (**Figure 2**). Based on ablation studies, one interpretation of these findings is that the activation of the midline cortices may

have been related to monitoring sensory events relative to the listener’s own behavior for purposes of spatial orientation and memory (Vogt et al., 1992). A related possibility is that downstream imagery and retrieval of episodic memories related to the acoustic scene may have preferentially led to activation of these midline regions (Hassabis et al., 2007). However, it remains unclear how these interpretations would fully account for the strong modulations we observed due to task demands (indicating end of sound versus indicating if the sound was produced by a human).

An alternative or additional possibility is that the activation profile we observed for scene-like versus object-like sounds along cortical midline structures was related to “default mode” network processing (Raichle et al., 2001; Greicius et al., 2003; Fransson and Marrelec, 2008). Acoustic scenes, which may be comprised of one or multiple sound textures (e.g., a ventilation and heating system, or sounds of rain and wind heard amidst a forest) often convey sensory information that the auditory system may dynamically and adaptively “filter out” or represent as background acoustic context (Maeder et al., 2001; Gygi et al., 2004; Overath et al., 2010), thereby freeing up attentional resources for other sensory or cognitive processes. This could include freeing up “default mode” processing that becomes suspended during specific goal-directed tasks.

In contrast to the object-like sounds, the scene-like mechanical and environmental sounds of the present study were characterized by relatively smoother $1/f^\alpha$ functions (**Figure 1B**), consistent with earlier reports (Voss and Clarke, 1975; Attias and Schreiner, 1997). As the distance between an observer and sound-source (or sources) increases there is a greater filtering of the sound pressure waves such that amplitude modulations in the acoustic signal becomes smoother. Perceptually, sound producing actions that are located further away from an observer’s focus of attention are arguably more likely to represent events that can be relegated as sensory “background.” Thus, sounds with relatively smoother $1/f^\alpha$ (among other attributes) are probabilistically more likely to be judged as scene-like, as opposed to object-like, even though the same sound-source may be judged as object-like when it is very close to the observer and/or when attention is directed to it.

The bilateral STG foci for object-like sounds were also significantly activated by the scene-like sounds relative to silent events, and the degree of activation exhibited a trend toward greater activation during a listening task that required sound categorization (human or not; i.e., **Figure 2**, Group B vs. A STG histograms). This response profile was consistent with the view that auditory scene analysis is a dynamic process that optimizes its representations of sound input depending on task demands (Hughes et al., 2001; Fritz et al., 2007a,b). Hence, the bilateral STG may be under top-down attentional control to channel specific acoustic features (such as those reflected by SSV, mean entropy, HNR, and other measures related to $1/f^\alpha$ profiles) as a means for directing attention to particular types or categories of anticipated sound (auditory objects or acoustic background scenes) based on past listening experiences. In the absence of an explicit sound categorization task, incoming signal input with

scene-like signal attributes (e.g., relatively low SSV, spectral flatness, smooth $1/f^\alpha$ profile) may be processed in a manner that more rapidly leads to acoustic accommodation, which in turn serves to recalibrate the listener to a new ambient noise “background.” Listening for sounds with the goal of categorizing them (i.e., Group B) may have led to decreased activation of default mode networks regardless of the sound category, and possibly regardless of whether or not a sound was even presented (i.e., hearing a “silent event” when *anticipating* a sound stimulus). Conversely, the relatively simpler task of determining the sound offset (i.e., Group A) may have permitted a relatively greater degree of activity related to default mode processing when hearing the scene-like sounds (**Figure 2**, brown regions). Given these interpretations, activation of the midline structures seems unlikely to be directly related to the processing of acoustic signals *per se*.

ANALOGIES BETWEEN VISUAL AND AUDITORY OBJECT PROCESSING

In the visual system, objects may be segregated from a background scene based on a number of different and converging features, including object motion, self-motion cues (head and eye movements), borders, textures, colors, etc. (Malach et al., 1995; Grill-Spector et al., 1998; Macevoy and Epstein, 2011). For the auditory system, action sounds necessarily imply the presence of some form of dynamic motion, ostensibly leading to the production of the sound pressure waves, whether or not those action sources can also be viewed. Thus, from a more general perspective of sensory processing, the ability to extract salient physical attributes such as changes in signal energy or entropy likely represents an efficient and common neuro-computational means for representing the presence of distinct objects and meaningful events in the environment. While direct comparisons with the visual system are not always straight forward (King and Nelken, 2009), some potential common principles in signal processing were revealed by the present study.

One signal processing computation that may generalize across sensory systems is time averaged mean entropy measures. Somewhat surprisingly, the mean entropy measures of environmental sounds, which showed no correlation with object-vs-scene Likert ratings (**Figure 3A**), did show a significant parametric correlation with activity in portions of the bilateral STG cortices, including the right hemisphere object-sensitive STG region. We speculate that these attributes may correlate with other perceptual dimensions, including judgments that emphasize discrimination of acoustic “textures” (Reddy et al., 2009; Overath et al., 2010; McDermott and Simoncelli, 2011), as opposed to other features such as object size or object-motion attributes. Sound and visual texture perception has been proposed to involve similar types of signal attribute computations in cortex (Warren et al., 1972; Julesz, 1980; Cusack and Carlyon, 2003; McDermott and Oxenham, 2008; Sathian et al., 2011). Together with the above studies, the present results are consistent with implicating entropy measures as one neuro-computational signal attribute that could be used to help segment, stream, or define objects (auditory, visual, or tactile) as distinct from other objects and from ambient background scenes.

Another potential analogy between auditory and visual processing strategies relates to “stationary” motion cues. The visual system includes pathways for processing first-order attributes, such as local luminance changes or changes in motion direction, as well as more subtle second- or third-order motion cues (e.g., contrast or spatial frequency deviations from the background, isoluminant chromatic motion), which are thought to rely on separate pathways (Chubb and Sperling, 1988; Cavanagh, 1992; Huddleston et al., 2008). In the auditory system, earlier neuroimaging studies demonstrated that sound motion processing, including explicit interaural intensity or time differences robustly activate primary auditory cortices (Griffiths et al., 1994; Mäkelä and McEvoy, 1996; Murray et al., 1998; Baumgart et al., 1999; Lewis et al., 2000; Warren et al., 2002). In our action sound stimuli, binaural spatial cues were entirely absent, and acoustic motion information depicting spatial excursions were not prevalent, with the exception of a few sounds containing motion-in-depth cues (looming or receding). Thus, we speculate that the measure of SSV in our collection of real-world sounds may be comparable to second- or third-order motion cues that are predominantly processed at stages hierarchically beyond, or at least distinct from, primary auditory cortices. More specifically, the SSV measures may capture physical motion features of real-world sounds-sources (monaural motion cues) that could alert the auditory system to the presence of an auditory object (e.g., a drying machine or ticking clock) even though the object as a whole may not be moving about in the space of one’s environment *per se*.

In the present study, sounds were presented in a relatively artificial acoustic environment—through ear-buds with the participant’s head held still while they were lying in an MRI scanner in the presence of a relatively low acoustic noise floor. Of course, the acoustic contexts in which an individual typically becomes familiar with real-world sound-sources, auditory objects, and acoustic scenes are within a wide variety of noisy acoustic backgrounds. Moreover, the freedom to make frequent head movements helps to entrain the auditory system to disambiguate the location of different sound sources as well as the acoustic features that might uniquely characterize the identity or category of those sources. Accordingly, we further speculate that acoustic attributes such as SSV measures may reflect an acoustic dimensionality reduction that the auditory system can use to probabilistically detect a “stationary” sound-producing object. Such processing would be robust against streaming interference due to different background ambiances, changes in spatial location of the source, and variations in monaural and binaural acoustic cues that occur during normal head movements by the listener. The processing of spectral signal structure variations characteristic of auditory objects may thus share some analogy with size and location invariant properties observed in intermediate visual object processing stages (e.g., the LOC regions), which are important feature extraction stages for figure-ground segregation processing of gross-level object form (Grill-Spector et al., 1998; Doniger et al., 2000; Kourtzi and Kanwisher, 2000). In sum, portions of the bilateral STG appear to incorporate SSV attributes, among various other low-level quantifiable signal

attributes, which may enable the brain to efficiently distinguish salient auditory “objects” and/or events that can emerge in complex acoustic scenes.

ACKNOWLEDGMENTS

We thank Drs. Robert Cox and Ziad Saad for continual development of AFNI and related software for cortical surface data

analyses, and Dr. Kristin Ropella for suggestions on acoustic signal processing.

FUNDING

This work was supported by the NCRN NIH COBRE grant E15524 (to the Sensory Neuroscience Research Center of West Virginia University).

REFERENCES

- Aglioti, S. M., Cesari, P., Romani, M., and Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nat. Neurosci.* 11, 1109–1116.
- Allison, T., McCarthy, G., Nobre, A., Puce, A., and Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb. Cortex* 5, 544–554.
- Antal, T., Droz, M., Gyorgyi, G., and Racz, Z. (2002). Roughness distributions for 1/f alpha signals. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65, 046140.
- Attias, H., and Schreiner, C. E. (1997). Temporal low-order statistics of natural sounds. *Adv. Neural Info. Process. Syst.* 9, 27–33.
- Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., and Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. *Eur. J. Neurosci.* 19, 2609–2612.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.
- Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H.-J., and Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature* 400, 724–725.
- Beauchamp, M., Lee, K., Haxby, J., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159.
- Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., and Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage* 10, 417–429.
- Bidet-Caulet, A., Voisin, J., Bertrand, O., and Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage* 28, 132–139.
- Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Caramazza, A., and Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn. Sci.* 7, 354–361.
- Cavanagh, P. (1992). Attention-based motion perception. *Science* 257, 1563–1565.
- Chao, L. L., and Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12, 478–484.
- Chubb, C., and Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. A* 5, 1986–2007.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17, 641–651.
- Cusack, R., and Carlyon, R. P. (2003). Perceptual asymmetries in audition. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 713–725.
- de Lucia, M., Camen, C., Clarke, S., and Murray, M. M. (2009). The role of actions in auditory object discrimination. *Neuroimage* 48, 475–485.
- Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., and Javitt, D. C. (2000). Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J. Cogn. Neurosci.* 12, 615–621.
- Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473.
- Dykstra, A. R., Halgren, E., Thesen, T., Carlson, C. E., Doyle, W., Madsen, J. R., Eskandar, E. N., and Cash, S. S. (2011). Widespread brain areas engaged during a classical auditory streaming task revealed by intracranial EEG. *Front. Hum. Neurosci.* 5:74. doi: 10.3389/fnhum.2011.00074
- Edmister, W. B., Talavage, T. M., Ledden, P. J., and Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Hum. Brain Mapp.* 7, 89–97.
- Elhilali, M., and Shamma, S. A. (2008). A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *J. Acoust. Soc. Am.* 124, 3751–3771.
- Engel, L. R., Frum, C., Puce, A., Walker, N. A., and Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage* 47, 1778–1791.
- Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598–601.
- Epstein, R. A., Higgins, J. S., Jablonski, K., and Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *J. Neurophysiol.* 97, 3670–3683.
- Epstein, R. A., and Morgan, L. K. (2011). Neural responses to visual scenes reveals inconsistencies between fMRI adaptation and multivoxel pattern analysis. *Neuropsychologia* 50, 530–543.
- Felleman, D. J., and van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.
- Fransson, P., and Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation network analysis. *Neuroimage* 42, 1178–1184.
- Frith, C. D., and Frith, U. (1999). Interacting minds—a biological basis. *Science* 286, 1692–1695.
- Fritz, J. B., Elhilali, M., and Shamma, S. A. (2007a). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* 98, 2337–2346.
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007b). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear. Res.* 229, 186–203.
- Gazzola, V., Aziz-Zadeh, L., and Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Curr. Biol.* 16, 1824–1829.
- Glover, G. H., and Law, C. S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn. Reson. Med.* 46, 515–522.
- Goll, J. C., Crutch, S. J., and Warren, J. D. (2011). Central auditory disorders: toward a neuropsychology of auditory objects. *Curr. Opin. Neurol.* 23, 617–627.
- Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258.
- Griffiths, T. D., Bench, C. J., and Frackowiak, R. S. J. (1994). Human cortical areas selectively activated by apparent sound movement. *Curr. Biol.* 4, 892–895.
- Griffiths, T. D., Kumar, S., Warren, J. D., Stewart, L., Stephan, K. E., and Friston, K. J. (2007). Approaches to the cortical analysis of auditory objects. *Hear. Res.* 229, 46–53.
- Griffiths, T. D., and Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353.
- Griffiths, T. D., and Warren, J. D. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203.
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., and Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum. Brain Mapp.* 6, 316–328.
- Gron, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., and Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nat. Neurosci.* 3, 404–408.
- Gygi, B., Kidd, G. R., and Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *J. Acoust. Soc. Am.* 115, 1252–1265.

- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., and Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223.
- Hassabis, D., Kumaran, D., and Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *J. Neurosci.* 27, 14365–14374.
- Hasson, U., Harel, M., Levy, I., and Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37, 1027–1041.
- Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Stein-schneider, M., Reale, R. A., Hind, J. E., and Brugge, J. F. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* 416, 79–92.
- Huddleston, W. E., Lewis, J. W., Phinney, R. E. Jr., and de Yoe, E. A. (2008). Auditory and visual attention-based apparent motion share functional parallels. *Percept. Psychophys.* 70, 1207–1216.
- Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., and Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. *Neuroimage* 13, 1073–1089.
- Husain, F. T., Tagamets, M. A., Fromm, S. J., Braun, A. R., and Horwitz, B. (2004). Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. *Neuroimage* 21, 1701–1720.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. 529–535. doi: 10.1371/journal.pbio.0030079
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211.
- Julesz, B. (1980). Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 83–94.
- Kaas, J. H., and Hackett, T. A. (1998). Subdivisions of auditory cortex and levels of processing in primates. *Audiol. Neurootol.* 3, 73–85.
- Kaas, J. H., Hackett, T. A., and Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170.
- Kanwisher, N., Chun, M. M., McDermott, J., and Ledden, P. J. (1996). Functional imaging of human visual recognition. *Brain Res. Cogn. Brain Res.* 5, 55–67.
- Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- King, A. J., and Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat. Neurosci.* 12, 698–701.
- Kohler, E., Keyzers, C., Umiltà, A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848.
- Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *J. Neurosci.* 20, 3310–3318.
- Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., and Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Comput. Biol.* 3:e100. doi: 10.1371/journal.pcbi.0030100
- Laaksonen, J. T., Markus Koskela, J., and Oja, E. (2004). Class distributions on SOM surfaces for feature extraction and object retrieval. *Neural Netw.* 17, 1121–1133.
- Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.
- Leech, R., Holt, L. L., Devlin, J. T., and Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *J. Neurosci.* 29, 5234–5239.
- Lewis, J. W. (2010). "Audio-visual perception of everyday natural objects – hemodynamic studies in humans," in *Multisensory Object Perception in the Primate Brain*, eds J. Marcus and P. J. K. Naumer, (Springer Science+Business Media, LLC), 155–190.
- Lewis, J. W., Beauchamp, M. S., and de Yoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* 10, 873–888.
- Lewis, J. W., Phinney, R. E., Brefczynski-Lewis, J. A., and de Yoe, E. A. (2006). Lefties get it "right" when hearing tool sounds. *J. Cogn. Neurosci.* 18, 1314–1330.
- Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., and Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *J. Cogn. Neurosci.* 23, 2079–2101.
- Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., and Brefczynski-Lewis, J. A. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *J. Neurosci.* 29, 2283–2296.
- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., and de Yoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cereb. Cortex* 14, 1008–1021.
- Macevoy, S. P., and Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nat. Neurosci.* 14, 1323–1329.
- Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14, 802–816.
- Mäkelä, J. P., and McEvoy, L. (1996). Auditory evoked fields to illusory sound source movements. *Exp. Brain Res.* 110, 446–453.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., and Tootell, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8135–8139.
- Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.
- McCarthy, G., Puce, A., Gore, J. C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9, 605–610.
- McDermott, J. H., and Oxenham, A. J. (2008). Spectral completion of partially masked sounds. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5939–5944.
- McDermott, J. H., and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71, 926–940.
- Medvedev, A. V., Chiao, F., and Kanwal, J. S. (2002). Modeling complex tone perception: grouping harmonics with combination-sensitive neurons. *Biol. Cybern.* 86, 497–505.
- Minda, J. P., and Ross, B. H. (2004). Learning categories by making predictions: an investigation of indirect category learning. *Mem. Cognit.* 32, 1355–1368.
- Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., Kraskov, A., Quiroga, R. Q., Adolphs, R., Fried, I., and Koch, C. (2011). A category-specific response to animals in the right human amygdala. *Nat. Neurosci.* 14, 1247–1249.
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., and Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13, 684–701.
- Murray, S. O., Newman, A. J., Roder, B., Mitchell, T. V., Takahashi, T., and Neville, H. J. (1998). Functional organization of auditory motion processing in humans using fMRI. *Soc. Neurosci. Abstr.* 24, 1401.
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.
- Oblerser, J., Eisner, F., and Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8123.
- Oblerser, J., Zimmermann, J., van Meter, J., and Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb. Cortex* 17, 2251–2257.
- Overath, T., Kumar, S., Stewart, L., von Kriegstein, K., Cusack, R., Rees, A., and Griffiths, T. D. (2010). Cortical mechanisms for the segregation and representation of acoustic textures. *J. Neurosci.* 30, 2070–2076.
- Pelphrey, K. A., Morris, J. P., and McCarthy, G. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J. Cogn. Neurosci.* 16, 1706–1716.
- Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., and Zilles, K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 13, 669–683.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682.
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the

- auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114.
- Reddy, R. K., Ramachandra, V., Kumar, N., and Singh, N. C. (2009). Categorization of environmental sounds. *Biol. Cybern.* 100, 299–306.
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Rizzolatti, G., Luppino, G., and Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalogr. Clin. Neurophysiol.* 106, 283–296.
- Rosch, E. H. (1973). Natural categories. *Cogn. Psychol.* 4, 328–350.
- Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A. N., Heller, A. C., Ross, I. B., Philpott, L., Sutherling, W. W., and Adolphs, R. (2011). Single-unit responses selective for whole faces in the human amygdala. *Curr. Biol.* 21, 1654–1660.
- Sathian, K., Lacey, S., Stilla, R., Gibson, G. O., Deshpande, G., Hu, X., Laconte, S., and Glielmi, C. (2011). Dual pathways for haptic and visual perception of spatial and texture information. *Neuroimage* 57, 462–475.
- Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme Medical Publishers.
- Talkington, W. J., Rapuano, K. M., Hitt, L., Frum, C. A., and Lewis, J. W. (in press) Humans mimicking animals: a cortical hierarchy for human vocal communication sounds. *J. Neurosci.*
- Tchernichovski, O., Mitra, P. P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291, 2564–2569.
- Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci.* 31, 164–171.
- Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Liu, A. K., and Dale, A. M. (1998). The representation of the ipsilateral visual field in human cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* 95, 818–824.
- van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28, 635–662.
- van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* 8, 443–459.
- Vogt, B. A., Finch, D. M., and Olson, C. R. (1992). Functional heterogeneity in cingulate cortex: the anterior executive and posterior evaluative regions. *Cereb. Cortex* 2, 435–443.
- Voss, R. E., and Clarke, J. (1975). 1/f noise in music and speech. *Nature* 258, 317–318.
- Warren, J., Zielinski, B., Green, G., Rauschecker, J., and Griffiths, T. (2002). Perception of sound-source motion by the human brain. *Neuron* 34, 139–148.
- Warren, R. M., Obusek, C. J., and Ackroff, J. M. (1972). Auditory induction: perceptual synthesis of absent sounds. *Science* 176, 1149–1151.
- Woods, D. L., Herron, T. J., Cate, A. D., Yund, E. W., Stecker, G. C., Rinne, T., and Kang, X. (2010). Functional properties of human auditory cortical fields. *Front. Syst. Neurosci.* 4:155. doi: 10.3389/fnsys.2010.00155
- Zatorre, R. J., Bouffard, M., and Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *J. Neurosci.* 24, 3637–3642.
- Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2011; accepted: 01 April 2012; published online: 09 May 2012.

Citation: Lewis JW, Talkington WJ, Tallaksen KC and Frum CA (2012) Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. *Front. Syst. Neurosci.* 6:27. doi: 10.3389/fnsys.2012.00027

Copyright © 2012 Lewis, Talkington, Tallaksen and Frum. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

APPENDIX

Table A1 | List of sound stimuli, ordered by object-like to scene-like Likert ratings.

Environmental (57 sounds)	Likert rating	Mechanical (54 sounds)	Likert rating
Mud bubbling	1.9	Clock ticking #2	1.1
Water bubbling #1	1.9	Stopwatch ticking	1.1
Water bubbling #2	2.1	Clock ticking #3, grandfather (cavernous)	1.1
Fire in fire place, loud cracks	2.4	Clock ticking, medium size	1.2
Water dripping, quietly	2.6	Egg timer, ticking	1.3
Water dripping in cave #1	2.7	Fax machine	1.4
Water running, stream	2.8	Printer, slow rate	1.4
Water bubbling, hot tub #2	3.0	Scanner adjusting	1.4
Glacier break;	3.1	Antique clock chiming and ticking	1.4
Heavy rain	3.1	Paint can lid rolling on floor	1.4
Water dripping in cave #2	3.1	Fax or copy machine adjusting	1.5
Lake water wave ashore	3.2	Church bell chimes	1.6
Waves, large lake	3.2	Money falling out of slot machine	1.7
Fire cackling, big forest	3.3	Church bells ringing	1.8
Fire in fire place	3.3	Industry generator, compressor	1.8
Wind blowing #1	3.3	Printer, rotor movements	1.8
Large river, flowing	3.4	Film projector	2.0
Rain fall, medium hard	3.4	Clocks, several ticking	2.1
Small waterfall	3.4	Fax machine	2.1
Forest fire	3.5	Printer, dot matrix	2.1
Fire crackling #1	3.6	Airplane, propeller	2.3
Fire crackling #2	3.7	Helicopter passing	2.3
River flowing	3.7	Machinery, chugging sounds	2.3
Small brush fire	3.8	Printer, office	2.3
Water waves coming ashore	3.8	Helicopter #2	2.3
Wind blowing, cold	3.8	Office machine, handling paper	2.3
Bubbling water in hot tub	3.8	Police car with siren passing by	2.3
Rocks falling/sliding	3.8	Fax machine, paper coming out	2.4
Wind gusting	3.9	Office printer, printing	2.4
River medium flow	3.9	Drying machine #2	2.4
Wind blowing #2	3.9	Fireworks going off	2.4
Water flow, stream	4.0	printer, feeding paper	2.4
Rockslide	4.1	Airplane, propeller #2	2.6
Rolling thunder	4.1	Air conditioner motor turning on	2.6
Water dripping	4.1	Drying machine #1	2.6
Wind blowing, low pitch	4.1	Puncher metal	2.7
Wind blowing, quietly	4.1	Helicopter #3	2.7
Wind, cold breeze #2	4.1	Conveyor belt moving	2.8
Glacier break	4.2	Windshield wipers	2.8
Wind blowing, high pitch	4.2	Airline fly by	2.9
Wind blowing, whistling #2	4.2	Large newspaper print press, chugging	3.1
Heavy wind through doorway	4.3	Mechanical conveyer moving	3.3
Glacier break	4.3	Garage door opening 2	3.4
Raining falling, with thunder	4.3	Pressbook, chugging sound	3.4
Wind blowing #5	4.3	Train, freight passing	3.4
Wind blowing #6	4.3	Garage door opening	3.5
Wind blowing #7	4.4	Train squeal breaks to a stop	3.5
Avalanche	4.4	Exhaust fan automatic turn on and blow	3.6
Glacier break	4.4	Industry, large press	3.6
Ocean waves #1	4.5	Jet airplane engine starting	3.6
Ocean waves #2	4.5	Heavy machine, quiet	3.7
Wind blowing, gusty	4.5	Train passing by	3.7
Wind, fast, wispy	4.5	Industry, flywheel	4.2
Heavy rainstorm with thunder	4.6	Industry, machinery	4.5
Ocean waves #3	4.6		
Rain, medium hard	4.6		
Wind blowing, whistling #1	4.7		

Bold text refers to extreme rated sounds used in Figure 1B.