Decoding the sound of hand-object interactions in primary somatosensory cortex

Kerri M Bailey[1], Bruno L Giordano[2], Amanda L Kaas[3], and Fraser W Smith[1]

[1]School of Psychology, University of East Anglia, Norwich, UK

[2]Institut de Neurosciences de la Timone, CNRS and Aix Marseille Université

[3]Department of Cognitive Neuroscience, Maastricht University

Correspondence should be addressed to FWS (e-mail: Fraser.Smith@uea.ac.uk).

School of Psychology

Lawrence Stenhouse Building

University of East Anglia

Norwich Research Park

Norwich, NR4 7TJ

UK

Telephone: +44 (0)1603 591676

Abstract

Neurons, even in earliest sensory regions of cortex, are subject to a great deal of contextual influences from both within and across modality connections. Recently we have shown that cross-modal connections from vision to primary somatosensory cortex (SI) transmit content-specific information about familiar visual object categories. In the present work, we investigated whether SI would also contain content-specific information about sounds depicting familiar hand-object interactions (e.g. bouncing a ball). In a rapid event-related functional magnetic resonance imaging (fMRI) experiment, participants listened attentively to multiple exemplars from three sound categories: hand-object interactions, and control categories of pure tones and animal vocalizations, while performing a one-back repetition counting task. Multi-voxel pattern analysis revealed significant decoding of different hand-object sounds within bilateral post-central gyrus (PCG), whilst no significant decoding was found for either control category. Crucially, in the hand-sensitive voxels defined from an independent tactile localizer, decoding accuracies were significantly higher for decoding hand-object sounds compared to both control categories in left PCG. Our findings indicate that hearing sounds depicting familiar hand-object interactions elicit different patterns of activity in SI, despite the complete absence of tactile stimulation. Thus cross-modal connections from audition to SI transmit content-specific information about sounds depicting familiar hand-object interactions.

Keywords: cortical feedback, multi-voxel pattern analysis, multisensory, S1. (4)

Much traditional neuroscientific research has investigated the function of the primary sensory brain areas (e.g. primary visual, somatosensory, and auditory cortices) with respect to how sensory input is processed within its corresponding sensory modality (see Carandini et al., 2005 for such an example of the visual system). However, it is well known that the majority of input to neurons, even within the primary sensory areas, comes from other cortical sources: either via local or long-range connections (for a review of the visual system, see Muckli & Petro, 2013). We have previously shown that non-stimulated early visual brain regions contain information transmitted via such connections about the surrounding context of natural visual scenes (Smith and Muckli 2010; Muckli et al. 2015). Such early sensory neurons are not only subject to a great deal of contextual influences *within* their respective sensory modality, however, but also via *across* modality connections (for reviews, see Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006). For example, in a recent study we have shown that different visual images of familiar but not unfamiliar object categories could be discriminated in primary somatosensory cortex - SI, despite the absence of any tactile stimulation during the experiment (Smith & Goodale, 2015). In the present work, we investigate whether such effects might also exist when participants are presented with sounds depicting familiar hand-object interactions. We expected this to be possible due to pre-existing associative links that are formed from prior experience of both sensory aspects (i.e. auditory and somatosensory) of interacting with objects (e.g. the sound and haptic stimulation of typing keys on a computer – see Meyer & Damasio, 2009; Smith & Goodale, 2015).

Support for this hypothesis is found in several additional studies that used multi-voxel pattern analysis (MVPA) to reveal content-specific effects in primary sensory cortices of a distal sensory modality to which the stimuli are presented. For example, Meyer et al. (2010) showed that simply viewing a silent yet sound-implying video clip transmits content-specific activity to primary auditory cortex (A1) even in the absence of auditory stimulation. Further

work has shown viewing video clips conveying object interactions with the hands can elicit different patterns of activity in SI in the absence of external tactile stimulation (Meyer et al. 2011). Additionally, when hearing the sounds of rich visual scenes, content-specific activity can be discriminated in early visual cortex, particularly in regions representing the periphery of visual space (Vetter et al. 2014). Finally, a set of classic multisensory studies has shown that primary sensory areas are subject to modulatory influences from other modalities (e.g. Calvert, 1997; McIntosh, Cabeza, & Lobaugh, 1998; see Driver & Noesselt, 2008 for a review).

Here we expand on this body of literature by investigating, for the first time, whether content-specific information can be sent to S1 when beginning from the auditory domain. Therefore, in the present study, participants listened to sound clips of familiar hand-object interactions (e.g. bouncing a ball, typing on a keyboard), in addition to two control categories (pure tones and animal vocalizations), in an event-related functional magnetic resonance imaging (fMRI) experiment. We predicted that MVPA would show significant decoding of sound identity for hand-object interaction sounds in SI, particularly in independently localized hand-sensitive voxels, but not for the two control categories.

Materials and Methods

Participants

Self-reported right handed healthy participants (N = 10; 3 male), with an age range of 18-25 years (M = 22.7, SD = 2.41), participated in this experiment. All participants reported normal or corrected-to-normal vision, and normal hearing, and were deemed eligible after meeting MRI screening criteria, approved by the Scannexus MRI centre in Maastricht. Written consent was obtained in accordance with approval from the Research Ethics

Committee of the School of Psychology at the University of East Anglia. Participants received €24 euros (equivalent to £20 sterling British pounds) for their time.

Stimuli & Design

Three different categories of auditory stimuli were used in a rapid event-related design; sounds depicting hand-object interactions, animal vocalizations, and pure tones. There were five different sub-categories within each of these categories, with two exemplars of each sub-category, thus resulting in 30 individual stimuli in total. The five hand-object interaction sub-categories consisted of bouncing a basketball, knocking on a door, typing on a keyboard, crushing paper, and sawing wood. These were chosen for the reason that participants should have previously either directly experienced rich haptic interactions with such objects, or observed such interactions. Two control categories were also used. Firstly, animal vocalizations were used as familiar sounds not directly involving interactions with the hands. These consisted of birds chirping, a dog barking, a fly buzzing, a frog croaking, and a rooster crowing. An independent ratings experiment confirmed these sounds were matched to the hand-object interactions for familiarity. Sounds from these two categories were downloaded from SoundSnap.com, YouTube.com, and a sound database taken from Giordano, McDonnell, and McAdams (2010). The second control category were non-meaningful sounds, defined as pure tones. These consisted of pure tones of five different frequencies (400Hz, 800Hz, 1600Hz, 3200Hz, and 6400Hz), created in MatLab. All sounds were stored in WAV format, and were cut to exactly 2000ms using Audacity 2.1.2, with sound filling the entire duration. Finally, all sounds were normalised to the root mean square (RMS) level (Giordano et al. 2013). More information regarding how these sounds were selected, including pilot experiments and ratings for the sounds, can be seen in Supplementary Text and Supplementary Table 1.

Each run began and ended with 12s silent blocks of fixation. After the initial 12s fixation, 60 individual stimuli were played, with each unique sound presented twice per run. Stimuli were played in a pseudo-randomly allocated order at 2s duration with a 3s ISI (5s trial duration). At random intervals, 15 null trials (duration 5s) were interspersed where no sound was played. This resulted in a total run time of 399s. After the main experiment, a somatosensory localiser was included to map the hand region in the somatosensory cortex (see Smith & Goodale, 2015). Piezo-electric Stimulator pads (Dancer Design, UK) were placed against the participant's index finger, ring finger, and palm of each hand using Velcro (six pads total, three per hand; see Supplementary Figure 1 for a visual example on one hand). Each pad contained a 6mm diameter disk centred in an 8 mm diameter static aperture. The disks stimulated both hands simultaneously with a 25Hz vibration in a direction normal to the surface of the disk and skin, at an amplitude within the range of ±0.5mm. Localiser runs consisted of 15 stimulation blocks and 15 baseline blocks (block design, 12s on, 12s off, 366s total run time). Note that for the first two participants, a slightly modified timing was employed (block design, 30s on, 30s off; 10 stimulation blocks, 9 baseline blocks).

Procedure

After signing informed consent, each participant was trained on the experimental procedure on a trial set of stimuli not included in the main experiment, before entering the scan room. Participants were instructed to keep fixated on a black and white central fixation cross presented against a grey background whilst listening carefully to the sounds, which were played at a self-reported comfortable level (as in Leaver & Rauschecker, 2010; Man, Damasio, Meyer, & Kaplan, 2015; Man, Kaplan, Damasio, & Meyer, 2012; Meyer et al., 2010). Participants performed a one-back repetition counting task, and hence counted the number of times they heard any particular sound repeated twice in a row, for example, two

sounds each of a dog barking (randomly allocated from 2 to 6 per run). We chose this task as it was important that the task did not require an explicit motor action such as pressing a button, to prevent a possible confound in somatosensory cortex activity (see Smith & Goodale, 2015). Thus, participants verbally stated the number of counted repetitions they heard at the end of each run, and they were explicitly asked to not make any movements in the scanner unless necessary. Overall, most participants completed either 8 or 9 runs (with the exception of one participant, who completed 7), thus, participants were exposed to approximately 32-36 repetitions per stimulus, and 16-18 repetitions per unique sound exemplar. After the main experiment, participants took part in the somatosensory mapping localiser. Participants were not informed about this part of the experiment until all main experimental runs had been completed. Each participant completed 1 (N = 2) or 2 (N = 8) somatosensory mapping runs, and kept their eyes fixated on a black and white central fixation cross presented against a grey background for the duration of each run. Participants were debriefed after completion of all scanning sessions.

MRI Data Acquisition

Structural and functional MRI data was collected using a high-field 3-Tesla MR scanner (Siemens Prisma, 64 channel head coil, Scannexus, Maastricht, Netherlands). High resolution T1 weighted anatomical images of the entire brain were obtained with a three-dimensional magnetization-prepared rapid-acquisition gradient echo (3D MPRAGE) sequence (192 volumes, 1mm isotropic). Blood-oxygen level dependent (BOLD) signals were recorded using a multiband echo-planar imaging (EPI) sequence: (400 volumes, TR = 1000ms; TE = 30ms; flip angle 77; 36 oblique slices, matrix 78 x 78; voxel size = 2.5mm^3; slice thickness 2.5mm; interslice gap 2.5mm; field of view 196; multiband factor 2). A short five volume posterior-anterior opposite phase encoding direction scan was acquired before

the main functional scans, to allow for subsequent EPI distortion correction (Jezzard and Balaban 1995; Fritz et al. 2014). Slices were positioned to cover somatosensory, auditory, visual, and frontal cortex. Sounds were presented via an in-ear hi-fi audio system (Sensimetrics, Woburn MA, USA), and the visual display was rear projected onto a screen behind the participant via an LCD projector. Finally, a miniature Piezo Tactile Stimulator (mini-PTS; developed by Dancer Design, UK) was used to deliver vibro-tactile stimulation to the hands, using the same fMRI sequence with a modified number of volumes (366s for the majority, slightly longer for the first two participants due to slightly different design – see Stimuli & Design above).

MRI Data Pre-processing

Functional data for each main experimental run, in addition to somatosensory localiser runs, was pre-processed in Brain Voyager 20.4 (Brain Innovation, Maastricht, The Netherlands; Goebel, Esposito, & Formisano, 2006), using defaults for slice scan time correction, 3D rigid body motion correction, and temporal filtering. Functional data were intra-session aligned to the pre-processed functional run closest to the anatomical scan of each participant. Distortion correction was applied using COPE 1.0 (Fritz et al. 2014), using the 5 volume scan acquired in the opposite phase encode direction (posterior to anterior) for each participant. Voxel displacement maps (VDM)'s were created for each participant, which were applied for EPI distortion correction to each run in turn. Functional data were then coregistered to the participant's ACPC anatomical scan. Note no Talairach transformations were applied, since such a transformation would remove valuable fine-grained pattern information from the data that may be useful for MVPA analysis (Fischl et al. 1999; Argall et al. 2006; Goebel et al. 2006; Kriegeskorte and Bandettini 2007). For the main MVPA analyses (described further below) we conducted a GLM analysis independently per run per

participant, with a different predictor coding stimulus onset for each stimulus presentation convolved with a standard double gamma model of the haemodynamic response function (see Greening, Mitchell, & Smith, 2018; Smith & Muckli, 2010). The resulting beta-weight estimates are the input to the pattern classification algorithm described below (see Multi-Voxel Pattern Analysis).

The somatosensory mapping localiser data was analysed using a GLM approach, with one predictor defining stimulation onset convolved with the standard double gamma model of the haemodynamic response function. The *t*-values were defined from the somatosensory localiser by taking the contrast of stimulation vs baseline in each participant. This allowed us to define the 100 voxels showing the strongest hand-related response in each individual's hand-drawn anatomical mask (see below) of the post-central gyrus (PCG).

Anatomical mask of Post-Central Gyri

In order to accurately capture the potential contribution from each sub-region of SI – i.e. area 3a, 3b, 1 or 2, hand-drawn masks of the PCG were created in each individual participant (Meyer et al. 2011; Smith and Goodale 2015). This allowed us to go beyond the capabilities of the somatosensory localiser alone, by enabling inclusion of all the information potentially available – i.e. both tactile and proprioceptive - in SI for the pattern classification algorithms (see Smith & Goodale, 2015 for further information). Furthermore, such an approach concords exactly with the key previous studies in this area (Smith and Goodale, 2015; Meyer, et al., 2011), since anatomical masks were defined in both these studies.

Anatomical masks were created using MRIcron 6 (Rorden et al. 2007) using each participant's anatomical MRI scan in ACPC space. As in Meyer et al. (2011) and Smith and Goodale (2015), the latero-inferior border was taken to be the last axial slice where the corpus callosum was not visible. From anterior to posterior the masks were defined by the

floors of the central and post-central sulci. Furthermore, masks did not extend to the medial wall in either hemisphere (Meyer, et al., 2011; Smith & Goodale, 2015). This resulted in an average of 41 slices (total range 39 to 46) for each hemisphere per participant (see Figure 1 for an example). The average voxel count was 1969 (*SD* = 229) for the right PCG, and 2106 (*SD* = 215) for the left PCG, which did not significantly differ from one another (p = .084). The size of each mask per participant is reported in Supplementary Table 2.

FIGURE 1 HERE

Additional Regions of Interest

Additional regions of interest (ROIs) were created using the Jüelich Anatomy toolbox (Eickhoff et al. 2005) as in Smith and Goodale (2015). Regions included Primary Auditory Cortex (Morosan et al. 2001; Rademacher et al. 2001), Pre-Motor Cortex (Geyer 2003), Primary Motor Cortex (Geyer et al. 1996), and Primary Visual Cortex (Amunts et al. 2000). We used the 30% probability cut-off for each map as this produces a roughly comparable number of voxels as in the anatomical masks of PCG (Smith & Goodale, 2015; Eickhoff et al., 2005). See Supplementary Figure 2 for examples of the masks.

Multi-Voxel Pattern Analysis

For the multi-voxel pattern analysis (MVPA; e.g. Haynes, 2015), we trained a linear support vector machine (SVM) to learn the mapping between the spatial patterns of brain activation generated in response to each of the five different sounds within a particular sound category (for example: for hand-object interactions, the classifier was trained on a five way discrimination between each relevant sub-category: typing on a keyboard, knocking on a door, crushing paper and so on; Greening et al., 2018; Smith & Goodale, 2015; Smith &

Muckli, 2010; Vetter et al., 2014). The classifier was trained and tested on independent data, using a leave one run out cross-validation procedure (Smith and Muckli 2010; Smith and Goodale 2015). The input to the classifier was always single trial brain activity patterns (beta weights) from a particular ROI while the independent test data consisted of an average activity pattern taken across the repetitions of specific exemplars in the left out run (e.g. the single trial beta weights of the four presentations of 'bouncing a ball' in the left out run were averaged). We have used this approach successfully in previous studies, as averaging effectively increases the signal to noise of the patterns (Smith and Muckli 2010; Vetter et al. 2014; Muckli et al. 2015). For similar approaches applied to EEG and MEG data, see Smith and Smith (2019) and Grootswagers, Wardle, and Carlson (2017) respectively. We note that a supplementary analysis using single trial activity patterns as the test data revealed a very similar pattern of results but lower accuracies as expected. These are reported in the Supplementary Results, and Supplementary Figure 3.

Finally, we used the LIBSVM toolbox (Chang and Lin 2011) to implement the linear SVM algorithm, using default parameters (C = 1). The activity pattern estimates (beta weights) within each voxel in the training data was normalised within a range of -1 to 1, prior to input to the SVM. The test data were also normalised using the same parameters as in the training set, in order to optimise classification performance. To test whether group level decoding accuracy was significantly above chance, we performed non-parametric Wilcoxon signed-rank tests using exact method on all MVPA analyses, against the expected chance level of 1/5 (Formisano et al. 2008; Greening et al. 2018), with all significance values reported two-tailed. Effect sizes for the Wilcoxon tests are calculated as $r = Z / \sqrt{N}$, when N = number of observations (Rosenthal 1991), to be identified as small (> .1), moderate (> .3), and large (> .5), according to Cohen's (1988) classification of effect sizes.

Results

Bilateral anatomical masks of the lateral post-central gyri (PCG) were defined in each participant (see Figure 1). We then computed cross-validated decoding performance of sound identity for each sound category (familiar hand-object interactions, animal vocalizations, and pure tones) independently in right, left and pooled PCG.

Post-Central Gyri

As expected, significantly above-chance decoding was found for hand-object interaction sounds in right PCG (Med = 23.65%; Z = -2.199, p = .025, r = .492), left PCG (Med = 30.56%; Z = -2.383, p = .016, r = .533), and pooled PCG (Med = 28.75%; Z = -2.490, p = .012, r = .557); signed rank, two-tailed test, chance = 20% (see Figure 2A). Crucially, the same analyses for our two control categories of familiar animal vocalizations and unfamiliar pure tones did not show any significant decoding above chance in right, left, or pooled PCG (all *p's* > .4). Further pairwise comparisons revealed decoding performance for hand-object interactions was significantly higher than pure tones in pooled PCG (Z = -2.380, p = .016, r = .532), showing the same trend in left PCG (p = .053). Decoding accuracies across the right and left hemisphere were not significantly different from one another for hand-object interaction sounds (p = .105*)*. Thus, the PCG carries content-specific information only for the familiar hand-object interaction sounds which convey haptic properties with the hands, regardless of hemisphere.

FIGURE 2 HERE

In subsequent analyses, when selecting the top 100 most active voxels in PCG from the somatosensory hand localiser, significant decoding for hand-object interactions was found

only in left PCG (Med = 29.45%; Z = -2.504, p = .008, r = .560); signed rank, two-tailed test, chance = 20% (see Figure 2B; see also Figure 2G for single participant data). Critically, decoding accuracies for hand-object interactions in left PCG were significantly higher than both control categories (Hands vs Animals: Z = -2.346, p = .016, r = .525; Hands vs Tones: Z = -2.603, p = .006, r = .582). In addition, decoding of hand-object interactions was significantly higher in the left than the right PCG (Z = -2.199, p = .027, r = .492). These results show the classifier could reliably decode hand-object interaction sounds above chance when constrained to the hand-sensitive voxels in left PCG, which were significantly higher than both control categories. Thus sound identity was reliably decoded above chance when restricting the MVPA analysis to voxels with high responses to tactile stimulation of the right, but not left, hand.

Additional Regions of Interest

Primary Auditory Cortex. As would be expected, decoding in primary auditory cortex (A1) was robustly significant for all sound categories (all Meds ≥ 64.72%, all Z's ≤ -2.601, all p's ≤ .002, all r's ≥ .627; signed rank, two-tailed test, chance = 20%; see Figure 2C). Further pairwise comparisons showed in right A1, decoding of pure tones (Med = 83.65%) was significantly higher than both animal vocalizations (Med = 71.25%, Z = -2.431, p = .012, r = .544) and hand-object interactions (Med = 66.25%, Z = -2.666, p = .004, r = .596), in addition to animal vocalizations being significantly higher than hand-object interactions (Z = -2.668, p = .004, r = .597). In pooled A1, pure tones (Med = 84.29%) were decoded significantly better than hand-object interactions (Med = 73.75%, Z = -2.552, p =.008, r = .571), and animal vocalizations (Med = 84.45%) were decoded significantly better than hand-object interactions (Z = -2.243, p = .023, r = .502). Thus in A1, all sound categories were highly discriminated with the specific pattern of decoding performance being the opposite to

that in PCG, with better decoding of pure tones, followed by animal vocalizations, then hand-object interaction sounds.

Pre-Motor Cortex. In pre-motor cortex (PMC), significantly above chance decoding was found for hand-object interactions in right PMC (Med = 30.56%; Z = -2.601, p = .006, r = .582), left PMC (Med = 24.72%, Z -2.527, p = .008, r = .565) and pooled PMC (Med = 31.25%, Z = -2.666, p = .004, r = .596); signed rank, two-tailed test, chance = 20% (see Figure 2D). Interestingly, further tests showed decoding for hand-object interactions was significantly higher than pure tones in right PMC (Z = -2.449, p = .012, r = .548), left PMC (Z = -2.197, p = .031, r = .491), and pooled PMC (Z = -2.807, p = .002, r = .628). Finally there was a weak effect of decoding animal sounds in pooled PMC (Med = 27.22%, Z = -1.963, p = .047, r = .439). Supplementary analyses – using single trial test data – also supported the robustness of this effect. Thus overall it appears that PMC may contain information about both types of familiar sound, but not the pure tone control category.

Primary Motor Cortex. Decoding accuracies in primary motor cortex (M1) revealed significant decoding for hand-object interactions only in left M1 (Med = 27.09, Z = -2.245, p = .021, r = .502; signed rank, two-tailed, chance = 20%; see Figure 2E). There were no reliable differences in decoding across categories or hemispheres, however hand-object interactions showed a strong trend to be significantly higher than pure tones (Med = 20.00%) in left M1 (p = .064, two-tailed). Supplementary analyses also showed weak evidence for decoding of animal calls in left M1.

Primary Visual Cortex. Decoding accuracies in primary visual cortex (V1) revealed no significant decoding. However, hand-object interactions in pooled V1 revealed a non-

significant trend for above-chance decoding performance (Med = 24.72%, p = .061; signed rank, two-tailed, chance = 20%; see Figure 2F).

Discussion

In the present study we show that hearing sounds depicting familiar hand-object interactions elicit significantly different patterns of activity in primary somatosensory cortex (SI), despite the complete absence of external tactile stimulation. Crucially, such decoding was not found for two control categories of familiar animal vocalizations, and unfamiliar pure tones. Moreover, further analysis showed decoding accuracies were significantly higher for hand-object interaction sounds compared to both control categories when restricted to the hand-sensitive voxels of the left PCG. Thus these results strongly suggest cross-modal connections from audition to SI transmit content-specific information about familiar hand-object sounds. We further show that pre-motor cortex (PMC) also contains content-specific information about sounds depicting familiar hand-object interactions, as well as weaker evidence for a similar effect with animal vocalizations.

Triggering cross-modal content-specific information in SI from audition

The present study agrees with a set of studies that show cross-modal connections can trigger content-specific activity even in the earliest regions of supposedly unimodal sensory cortex, across multiple sensory domains (Meyer et al., 2010, 2011; Vetter et al., 2014; Smith & Goodale, 2015). Our results significantly extend this previous body of work by demonstrating, for the first time, that a similar effect is present across the domains of audition and touch for particular sound categories. Similar to Smith & Goodale (2015), we demonstrate that such effects are not present for all sound categories, but only for specific categories where some correspondence is known to exist across sensory modalities (i.e. hand-

object sounds activate lateral PCG where hand sensitive voxels are located). In fact, we demonstrate greater decoding of hand-object interaction sounds than both pure tones and animal vocalizations in the hand sensitive voxels of left PCG. It is important to note, furthermore, that the pattern of decoding performance was different in PCG and in auditory cortex, with better decoding of hand sounds than pure tones in PCG (and also in PMC) but vice versa in auditory cortex. Taken together this suggests that the present results in PCG reflect high-level information being activated in this region, as opposed to passive relay of low level acoustic features from auditory cortex.

The current results expand on our earlier study that showed simply viewing images of familiar graspable objects led to reliable decoding in SI (Smith & Goodale, 2015; see also the related study of Meyer et al., 2011). Thus we show either viewing images or hearing sounds of appropriate objects/events – i.e. those related to haptic interactions with the hands - triggers content-specific activity in the earliest regions of somatosensory cortex. Crucially, both the current study and Smith & Goodale (2015) demonstrate that these effects are present when the analyses are limited to independently-localized hand- (or finger)-sensitive voxels in the PCG, and that the effects are higher in magnitude than for appropriate control categories. What then might be the function of this cross-sensory information being present in SI? If Predictive Coding is assumed to be the general computational function of the brain (Friston et al. 2009; Clark 2013), then it may be the case that either seeing a familiar graspable object, or hearing the sounds associated with a familiar hand-object interaction, leads to content-specific activity in SI that is useful for future (or concurrent) interaction with the specific object. It should be possible to directly test Predictive Coding accounts by using appropriately designed paradigms where specific sensory cues (e.g. visual or auditory) predict forthcoming 3D objects (see e.g. Rossit, McAdam, Mclean, Goodale, & Culham, 2013) in a target modality such as somatosensory cortex (see Kok, Jehee, & de Lange, 2012;

see also Zhou & Fuster, 2000). Alternatively it might be the case that the activation present in SI is not useful for future object interaction but rather reflects the broader representation of object knowledge in the haptic domain (e.g. Man et al., 2012; Martin, 2016; Meyer & Damasio, 2009). Such accounts propose that the representation of object concepts is distributed across the perceptual, action and emotion systems in the brain (Martin 2016), and that conceptual processing involves *neural re-use* of the same systems for *perception* and for *conceptual processing* (Barsalou 2016). While these accounts do not generally invoke the primary sensory areas as being involved in the representation of object knowledge, Martin's (2016) account, for example, proposes that such regions may become involved under specific task conditions. Such accounts would predict that the effects we report in SI should be modified as a function of task constraints – with stronger effects for tasks where somatosensory properties of objects and/or actions are more vs less prominent.

How does auditory stimulation lead to content-specific information being present in SI? There are several possible routes through which this could be accomplished (see e.g. Driver & Noesselt 2008). First, auditory information may first arrive at high level multisensory convergence zones, such as pSTS, posterior parietal cortex or ventral premotor areas, before such areas then send feedback to SI (see Driver & Noesselt 2008; Meyer et al., 2009; see also Smith & Goodale, 2015, for a comparable view). Second, auditory information may be directly projected to SI, without passing through such higher multisensory regions. Such direct connections have been found in animal models between certain sensory pairings: e.g. from primary auditory to primary visual cortex (e.g. Falchier, Clavagnier, Barone, & Kennedy, 2002; Falchier et al., 2010) and between primary auditory and somatosensory cortex (Budinger, Heil, Hess, & Scheich, 2006; see also Cappe & Barone, 2005). However it has been proposed that such direct connections are relatively sparse as opposed to the amount of feedback arriving from higher multisensory areas (Driver and Noesselt 2008). Finally a

third possibility is the involvement of lower tier multisensory regions (Driver and Noesselt 2008) that are anatomically located next to primary sensory areas: for instance auditory regions located close to SII may be bimodal responding to both auditory and tactile information (see e.g. Cappe & Barone, 2005; Wallace, Ramachandran, & Stein, 2004). Indeed recent studies in humans have indicated the presence of auditory frequency information in SII area OP4 (Pérez-Bellido et al. 2018) and the influence of auditory information at very fast time-scales in SI indicative of a feedforward effect (Sugiyama et al. 2018). From the present data we can only speculate on the involvement of these potential routes by which information could be transmitted from auditory to somatosensory cortex. We note that high temporal resolution techniques – such as EEG or MEG – may reveal differences in the timing of cross-sensory effects that would indicate the involvement of direct or feedback-mediated mechanisms (see e.g. Sugiyama et al., 2018). In addition, laminar fMRI may be used to indicate which layers of SI manifest cross-sensory effects, with implications for the relative role of feedforward vs feedback-mediated connections (see e.g. Lawrence, Formisano, Muckli, & de Lange, 2019; Muckli et al., 2015).

Hemispheric Differences in auditory and visually triggered cross-sensory information in SI

We note that both Smith and Goodale (2015) and Meyer et al. (2011) used visual stimuli (either images of familiar graspable objects, or videos depicting haptic explorations of objects with the hands, respectively) and found stronger decoding accuracies in the right hemisphere PCG whereas in the present study with auditory stimuli we found stronger decoding in the left hemisphere PCG. There are several potential reasons for the greater involvement of hand-sensitive voxels in the left PCG in the present study. First, some of our sounds depict bimanual actions (e.g. typing on a keyboard) and previous studies have found greater activation in the left hemisphere for bimanual action sounds (Aziz-Zadeh et al. 2004).

In addition, much research concerning the neural processing of tools has reported strong left lateralization of the tool network in right-handed participants (as our participants were; Ishibashi, Pobric, Saito, & Lambon Ralph, 2016; Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005; Lewis, Phinney, Brefczynski-Lewis, & DeYoe, 2006) although this would suggest left-lateralization for both sounds and images/videos. However it may be the case that such left lateralization depends upon object directed action content being present (or strongly implied), as was the case in the current study. Finally one further important difference between Smith and Goodale (2015) and the present study is that in the earlier study we mapped hand (finger) sensitive voxels in SI for each hand independently, which permitted considering the relative influence of contra- and ipsi-lateral influences, whereas in the present study both hands were mapped simultaneously. Hence in the present study selected voxels may have reflected a stronger contra-lateral bias, and therefore reflect the relatively earlier sub-regions of SI, such as area 3b. Future work will be necessary, ideally in the same participants with the same localizers, to determine the relative role of hand sensitive voxels in left and right SI to visually and auditory triggered information.

Decoding action related information in Pre-Motor Cortices

In pre-motor cortex (PMC), we also found reliable decoding of hand-object interaction sounds in both the left and right hemisphere. We also found greater decoding of such sounds compared to pure tones. Finally, decoding of animal vocalizations was significant in PMC when pooling across hemispheres (and in further analyses using single trials as test data, the effect was robustly present in each hemisphere). PMC is known to play a large role in processing action related information (Gallese et al. 1996). For instance, PMC has been found to be preferentially activated for object-related hand actions and non-object-related mouth actions (Buccino et al. 2001). Both our familiar sounds of hand-object

interactions and animal vocalizations imply such an action, therefore it would seem reasonable for them to trigger activity in pre-motor areas.

In addition, decoding in PMC for both hand-object interactions and animal vocalizations could be part of a somatotopic auditory mirror neuron system, since PMC has previously been found to be active in response to both performing an action and hearing the corresponding action sound (Kohler et al. 2002). Furthermore, Gazzola, Aziz-Zadeh, and Keysers (2006) found overlap at the voxel level between left PMC activation when human participants executed a motor action, or listened to the sound of the action. Crucially, they found a somatotopic pattern, whereby a dorsal cluster within PMC was involved in listening to and executing hand actions, and a ventral cluster within PMC was involved in listening to and executing mouth actions. Therefore, both hand- and mouth-specific clusters within PMC may contribute to the decoding found in the present experiment. We would predict no significant decoding for the hand-object sounds if PMC analyses were limited to mouth-selective voxels in PMC, and likewise for animal vocalizations in hand-selective voxels in PMC. Hence in future work it would be optimal to include an additional mouth (and hand) movement localizer to test these predictions. Overall, the decoding effects we see for action-related information in PMC (and also M1) suggest these regions also receive content-specific information regarding the action properties of familiar hand-object interaction sounds (with weaker evidence of an effect also for animal vocalizations).

Conclusion

We have shown that the identity of familiar hand-object interaction sounds can be discriminated in SI, in the absence of any concurrent tactile stimulation. Thus cross-modal connections from audition to SI transmit content-specific information about such sounds. Our work provides converging evidence that activity in supposedly modality-specific primary

sensory areas can be shaped in a content-specific manner by relevant contextual information transmitted across sensory modalities (Meyer et al. 2010, 2011; Vetter et al. 2014; Smith and Goodale 2015). Such an effect is in keeping with the rich range of contextual effects expected in primary sensory areas under the Predictive Processing framework (Clark, 2013).

References

Amunts K, Malikovic A, Mohlberg H, Schormann T, Zilles K. 2000. Brodmann's areas 17 and 18 brought into stereotaxic space - Where and how variable? Neuroimage. 11:66–84.

Argall BD, Saad ZS, Beauchamp MS. 2006. Simplified intersubject averaging on the cortical surface using SUMA. Hum Brain Mapp. 27:14–27.

Aziz-Zadeh L, Iacoboni M, Zaidel E, Wilson S, Mazziotta J. 2004. Left hemisphere motor facilitation in response to manual action sounds. Eur J Neurosci. 19:2609–2612.

Barsalou LW. 2016. On Staying Grounded and Avoiding Quixotic Dead Ends. Psychon Bull Rev. 23:1122–1142.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: An FMRI study. Eur J Neurosci. 13:400–404.

Budinger E, Heil P, Hess A, Scheich H. 2006. Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. Neuroscience. 143:1065–1083.

Calvert G. 1997. Activation of auditory cortex during silent lip-reading. Science. 276:593–596.

Cappe C, Barone P. 2005. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci. 22:2886–2902.

Carandini M, Demb J, Mante V, Tolhurst D, Dan Y, Olshausen B, Gallant J, Rust N. 2005. Do We Know What the Early Visual System Does? J Neurosci. 25:10577–10597.

Chang C, Lin C. 2011. LIBSVM: a library for support vector machines. ACMTrans Intell Syst Technol. 2:27.

Clark A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive

science. Behav Brain Sci. 181–253.

Cohen JD. 1988. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum.

Driver J, Noesselt T. 2008. Multisensory Interplay Reveals Crossmodal Influences on "Sensory-Specific" Brain Regions, Neural Responses, and Judgments. Neuron. 57:11–23.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage. 25:1325–1335.

Falchier A, Clavagnier S, Barone P, Kennedy H. 2002. Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci. 22:5749–5759.

Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF. 2010. Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. Cereb Cortex. 20:1529–1538.

Fischl B, Sereno MI, Tootell RBH, Dale AM. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp. 8:272–284.

Formisano E, De Martino F, Bonte M, Goebel R. 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science (80-). 322:970–973.

Friston K, Kiebel S, Barlow H., Feynman R., Neal R., Hinton G., Neisser U. 2009. Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci. 364:1211–1221.

Fritz L, Mulders J, Breman H, Peters J, Bastiani M, Roebroeck A, Andersson J, Ashburner J, Weiskopf N, Goebel R. 2014. Comparison of EPI distortion correction methods at 3T and 7T. In: OHBM Annual Meeting, Hamburg, Germany.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. 1996. Action recognition in the premotor

cortex. Brain. 119:593–609.

Gazzola V, Aziz-Zadeh L, Keysers C. 2006. Empathy and the Somatotopic Auditory Mirror System in Humans. Curr Biol. 16:1824–1829.

Geyer S. 2003. The microstructural border between the motor and the cognitive domain in the human cerebral cortex. 1st ed. Wien: Springer.

Geyer S, Ledberg A, Schleicher A, Kinomura S, Schormann T, Bürgel U, Klingberg T, Larsson J, Zilles K, Roland PE. 1996. Two different areas within the primary motor cortex of man. Nature. 382:805–807.

Ghazanfar AA, Schroeder CE. 2006. Is neocortex essentially multisensory? Trends Cogn Sci. 10:278–285.

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P. 2013. Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex.

Giordano BL, McDonnell J, McAdams S. 2010. Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds. Brain Cogn. 73:7–19.

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

Greening SG, Mitchell DGV, Smith FW. 2018. Spatially generalizable representations of facial expressions: Decoding across partial face samples. Cortex. 101:31–43.

Grootswagers T, Wardle SG, Carlson TA. 2017. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J Cogn Neurosci. 29:677–697.

Ishibashi R, Pobric G, Saito S, Lambon Ralph MA. 2016. The neural network for tool-related

cognition: An activation likelihood estimation meta-analysis of 70 neuroimaging contrasts. Cogn Neuropsychol. 33:241–256.

Jezzard P, Balaban RS. 1995. Correction for geometric distortion in echo planar images from B0 field variations. Magn Reson Med. 34:65–73.

Kohler E, Keysers C, Alessandra Umilta A, Fogassi L, Gallese V, Rizzolatti G. 2002. Hearing sounds, understanding actions: Action representation in mirror neurons. Science (80-). 297:846–848.

Kok P, Jehee JFM, de Lange FP. 2012. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron. 75:265–270.

Kriegeskorte N, Bandettini P. 2007. Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage. 38:649–662.

Lawrence SJD, Formisano E, Muckli L, de Lange FP. 2019. Laminar fMRI: Applications for cognitive neuroscience. Neuroimage. 197:785–791.

Leaver AM, Rauschecker JP. 2010. Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. J Neurosci. 30:7604–7612.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. 2005. Distinct Cortical Pathways for Processing Tool versus Animal Sounds. J Neurosci. 25:5148–5158.

Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe E a. 2006. Lefties Get It “Right” When Hearing Tool Sounds. J Cogn Neurosci. 18:1314–1330.

Man K, Damasio A, Meyer K, Kaplan JT. 2015. Convergent and invariant object representations for sight, sound, and touch. Hum Brain Mapp. 36:3629–3640.

Man K, Kaplan JT, Damasio A, Meyer K. 2012. Sight and sound converge to form modality-invariant representations in temporoparietal cortex. J Neurosci. 32:16629–16636.

Martin A. 2016. GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain.

Psychon Bull Rev. 23:979–990.

McIntosh AR, Cabeza RE, Lobaugh NJ. 1998. Analysis of Neural Interactions Explains the Activation of Occipital Cortex by an Auditory Stimulus. J Neurophysiol. 80:2790–2796.

Meyer K, Damasio A. 2009. Convergence and divergence in a neural architecture for recognition and memory. Trends Neurosci. 32:376–382.

Meyer K, Kaplan JT, Essex R, Damasio H, Damasio A. 2011. Seeing touch is correlated with content-specific activity in primary somatosensory cortex. Cereb Cortex. 21:2113–2121.

Meyer K, Kaplan JT, Essex R, Webber C, Damasio H, Damasio A. 2010. Predicting visual stimuli on the basis of activity in auditory cortices. Nat Neurosci. 13:667–668.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. 2001. Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage. 13:684–701.

Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E. 2015. Contextual Feedback to Superficial Layers of V1. Curr Biol. 25:2690–2695.

Muckli L, Petro LS. 2013. Network interactions: Non-geniculate input to V1. Curr Opin Neurobiol. 23:195–201.

Norman KA, Polyn SM, Detre GJ, Haxby J V. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10:424–430.

Pérez-Bellido A, Anne Barnes K, Crommett LE, Yau JM. 2018. Auditory Frequency Representations in Human Somatosensory Cortex. Cereb Cortex. 28:3908–3921.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage. 13:669–683.

Rorden C, Karnath H-O, Bonilha L. 2007. Improving Lesion–Symptom Mapping. J Cogn Neurosci. 19:1081–1088.

Rosenthal R. 1991. Meta-analytic procedures for social research. 2nd ed. Newbury Park, CA: Sage.

Rossit S, McAdam T, Mclean DA, Goodale MA, Culham JC. 2013. FMRI reveals a lower visual field preference for hand actions in human superior parieto-occipital cortex (SPOC) and precuneus. Cortex. 49:2525–2541.

Smith FW, Goodale MA. 2015. Decoding visual object categories in early somatosensory cortex. Cereb Cortex. 25:1020–1031.

Smith FW, Muckli L. 2010. Nonstimulated early visual areas carry information about surrounding context. Proc Natl Acad Sci U S A. 107:20099–20103.

Smith FW, Smith ML. 2019. Decoding the dynamic representation of facial expressions of emotion in explicit and incidental tasks. Neuroimage. 195:261–271.

Sugiyama S, Takeuchi N, Inui K, Nishihara M, Shioiri T. 2018. Effect of acceleration of auditory inputs on the primary somatosensory cortex in humans. Sci Rep. 8:1–9.

Vetter P, Smith FW, Muckli L. 2014. Decoding sound and imagery content in early visual cortex. Curr Biol. 24:1256–1262.

Wallace MT, Ramachandran R, Stein BE. 2004. A revised view of sensory cortical parcellation. Proc Natl Acad Sci. 101:2167–2172.

Zhou YD, Fuster JM. 2000. Visuo-tactile cross-modal associations in cortical somatosensory cells. Proc Natl Acad Sci. 97:9777–9782.

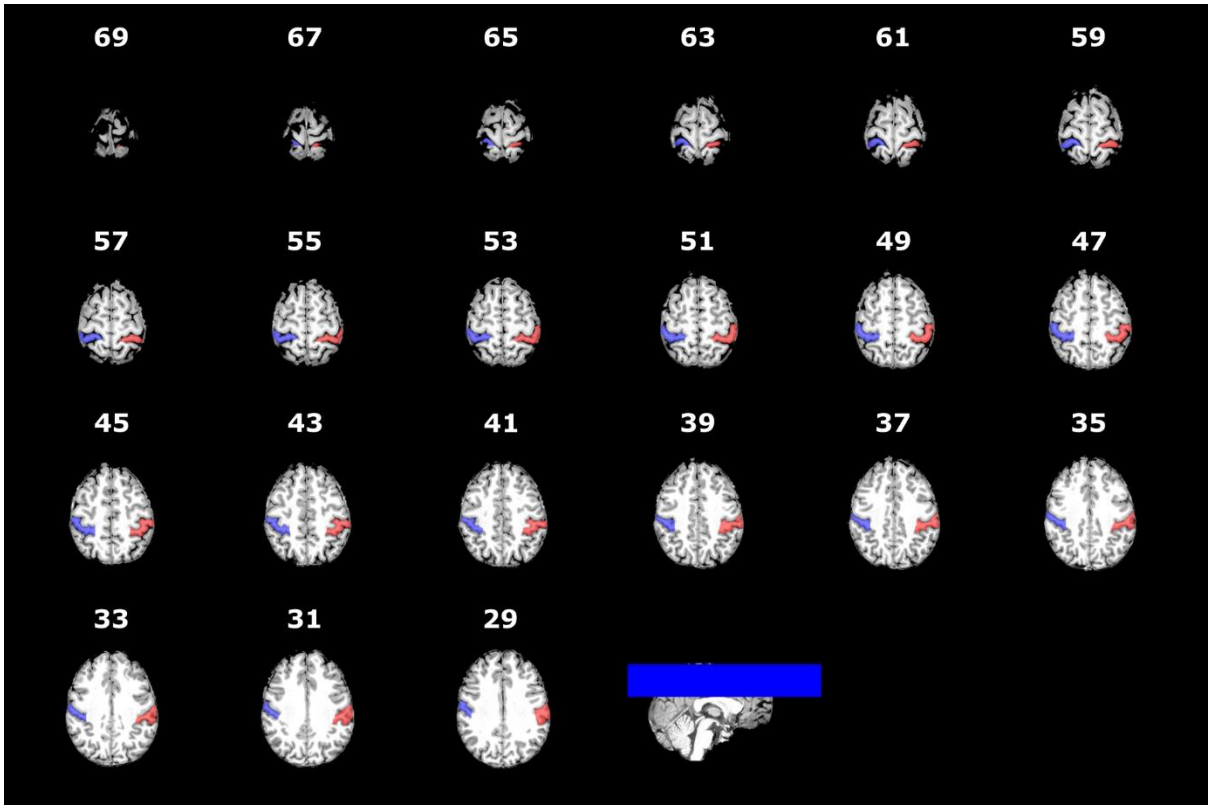

Figure 1. Anatomical masks of the lateral post-central gyrus for a representative participant. The numbers in white refer to slices through the Z plane. The box in the lower right image depicts the slices of the brain on which the PCG was marked (see Materials and Methods).

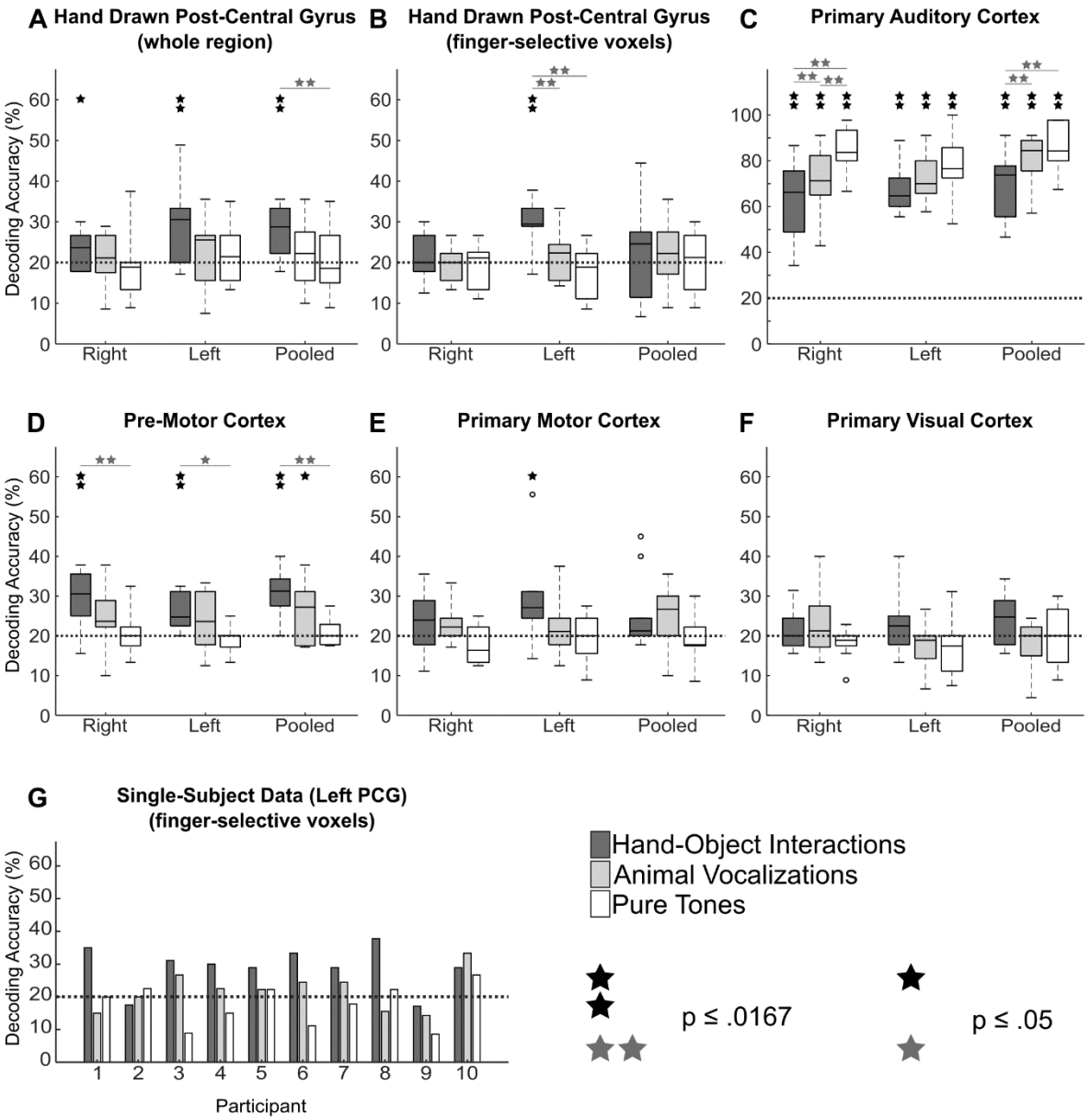

Figure 2. Decoding of sound identity. (A) Cross-validated 5AFC decoding performance for each stimulus category (hand-object interaction sounds, animal vocalizations and pure tones) for the right and left post-central gyri independently and pooled across hemispheres. Double stars: $P < 0.0167$, single star: $P < 0.05$. (B) As in A but for the top 100 voxels that were responsive to tactile stimulation of the hands in an independent localizer session. (C–F) As in A but for several additional, anatomically defined, regions of interest. (G) As in B left post-central gyri but single participant data.