Cerebral Cortex July 2015;25:1947–1957
doi:10.1093/cercor/bhu003
Advance Access publication January 31, 2014

Representations of Invariant Musical Categories Are Decodable by Pattern Analysis of Locally Distributed BOLD Responses in Superior Temporal and Intraparietal Sulci

Mike E. Klein[1,2] and Robert J. Zatorre[1,2]

[1]Cognitive Neuroscience Unit, Montréal Neurological Institute, McGill University, Montréal, Québec, Canada H3A 2B4 and [2]International Laboratory for Brain, Music and Sound Research, Montréal, Québec, Canada H3C 3J7

Address correspondence to Mike E. Klein, Cognitive Neuroscience Unit, Montréal Neurological Institute, 3801 University Street, Room 276, Montréal, Québec, Canada H3A 2B4. Email: michael.klein@mail.mcgill.ca; michaeleklein@gmail.com

In categorical perception (CP), continuous physical signals are mapped to discrete perceptual bins: mental categories not found in the physical world. CP has been demonstrated across multiple sensory modalities and, in audition, for certain over-learned speech and musical sounds. The neural basis of auditory CP, however, remains ambiguous, including its robustness in nonspeech processes and the relative roles of left/right hemispheres; primary/nonprimary cortices; and ventral/dorsal perceptual processing streams. Here, highly trained musicians listened to 2-tone musical intervals, which they perceive categorically while undergoing functional magnetic resonance imaging. Multivariate pattern analyses were performed after grouping sounds by interval quality (determined by frequency ratio between tones) or pitch height (perceived noncategorically, frequency ratios remain constant). Distributed activity patterns in spheres of voxels were used to determine sound sample identities. For intervals, significant decoding accuracy was observed in the right superior temporal and left intraparietal sulci, with smaller peaks observed homologously in contralateral hemispheres. For pitch height, no significant decoding accuracy was observed, consistent with the non-CP of this dimension. These results suggest that similar mechanisms are operative for nonspeech categories as for speech; espouse roles for 2 segregated processing streams; and support hierarchical processing models for CP.

Keywords: auditory categories, categorical perception, intraparietal sulcus, MVPA, superior temporal sulcus

Introduction

An overarching feature of perception is the awareness of stimuli as "whole" objects, rather than complex amalgams of ambiguous physical signals. A specific aspect of this phenomenon occurs for certain classes of stimuli that are subject to "categorical perception" (CP), whereby continuous physical signals are mapped onto discrete mental categories, mediated by long-term memory. CP was first behaviorally demonstrated in speech perception (Liberman et al. 1957) and later in nonspeech and nonauditory domains, including perception of musical intervals (Burns and Ward 1978; Zatorre and Halpern 1979) and color (Bornstein 1984), implicating it as a more general phenomenon. The neural substrates of CP remain unclear, but increasing evidence indicates that it may be mediated by 2 dissociable streams of information processing: (1) A more perceptual ventral system focused on object identification/recognition and (2) a dorsal system related to motor production, with requisite linkages to the premotor/motor system (Hickok and Poeppel 2007; Rauschecker and Scott 2009).

Over the past decade, functional neuroimaging studies of CP have implicated subregions of the left superior temporal sulcus (STS; Liebenthal et al. 2005; Joanisse et al. 2006; Leech et al. 2009), thought to be part of a ventral stream, as well as portions of the posterior superior temporal gyri (STG) and left parietal and frontal lobes, thought to be nodes in a motor-related dorsal stream (Raizada and Poldrack 2007; Hutchison et al. 2008; Myers et al. 2009). Most of these studies, however, have employed speech (or speech-like) stimuli, leading to what may be an overgeneralization of the predominantly left hemispheric results. A study examining blood oxygen level-dependent (BOLD) responses to categorically perceived musical intervals implicated the right STS and left intraparietal sulcus (IPS; Klein and Zatorre 2011), indicating that these cortical streams may also be recruited for nonspeech categorical processing. The wide variety of intra- and extra-STS peaks is likely due in part to design choices (specific in-scanner experimental tasks, control conditions, and contrasts), leading to differences in networks observed for any one task/contrast (a situation complicated by the range of sensitivity available via univariate and multivariate analysis methodologies). This literature, and the resultant interpretation of imaging results, is further complicated by the strictness with which true CP is behaviorally defined; many studies report data for identification, but not discrimination tasks, while the latter is the only way to ascertain that the processing of category information in some way dominates perception (Repp 1984). Thus, while evidence has begun to mount implicating the STS in categorical processing, the totality of the neural circuitry underlying both speech and nonspeech auditory category perception remains an open question.

To examine the neural basis of nonspeech auditory CP while minimizing potential confounds due to the nature of tasks and control stimuli, we utilized multivariate pattern analyses (MVPAs), which consider data from spatially distributed patterns of brain activity to differentiate between experimental conditions (Haynes and Rees 2006; Mur et al. 2008; Pereira et al. 2009). MVPA's enhanced sensitivity over univariate General Linear Model (GLM) analyses allows for (a) comparison between "sibling" conditions of interest from the same underlying continuum, as opposed to use of "null" conditions lacking some essential quality (e.g., direct comparison of 2 speech phonemes without the need for acoustically matched controls that are not perceived as speech sounds) and (b) utilization of fairly passive scanning protocols, free of major behavioral task requirements. Using a local pattern analysis "searchlight" approach (Kriegeskorte et al. 2006), we sought to distinguish between brain regions carrying decodable information about the categorical quality of musical intervals from any regions underlying noncategorical processing of pitch height. Compared with speech stimuli, musical intervals are

nonlinguistic, acoustically simple, and allow for experimental and orthogonal differentiability based on the same feature (tone frequency). Thus, the use of musical intervals allows for the possibility to dissociate bottom-up, absolute pitch-based effects (present in both stimuli dimensions in roughly equal quantity) from top-down, categorical memory-based effects (present in the interval quality—but not the absolute frequency—dimension).

Unlike prior imaging studies of CP, we employed a combination of (a) behavioral identification and discrimination tasks to be certain that true CP was demonstrated; (b) 3 categories per continuum, in order to be certain that observations were not due to anchoring/range effects (Simon and Studdert-Kennedy 1978); and (c) an orthogonal control dimensions, which circumvent confounds due to differences in the physical features of stimuli. Because analyses decoding only single exemplars of musical intervals would not allow us to dissociate which component of the results were due to categorical differences as opposed to acoustic differences between the sounds, the classifiers were trained and tested on multiple exemplars of each interval varying in absolute pitch (i.e., roved in the orthogonal dimension), and these MVPA results were compared with those from the orthogonal analysis based on the pitch height dimension, which was not predicted to be categorically perceived. Classification of categorical qualities was hypothesized to occur in the superior temporal and intraparietal sulci, with successful pitch height decoding predicted in the STG.

Materials and Methods

Study Participants

We recruited 37 trained musician participants (22 females, minimum 5 years formal training and currently practicing or performing); the majority of whom came from McGill University's undergraduate and graduate music student populations and none of whom possessed absolute pitch abilities. Of this cohort, we selected 10 participants (4 females, average 13 years of musical training, 8 instrumentalists, and 2 singers) who showed the greatest degree of CP, as determined by discrimination task performance (see "prescanning behavioral tasks" below). All participants gave their informed consent. Ethical approval was granted by the Montreal Neurological Institute Ethics Review Board.

Pretest Sound Stimuli

Each experimental stimulus was composed of a 2-tone melodic (i.e., sequential) interval. Each 750-ms complex tone was synthesized in Audacity and Max/MSP software out of 5 harmonics with amplitudes inversely proportional to the harmonic number. A volume envelope was applied (initial 50 ms ramp from 0% to 100% and final 50 ms ramp from 100% to 0%) in order to avoid onset and offset percussive clicks, and sound intensity was adjusted to each subject's comfort level. The two 750-ms tones in a given interval were separated by a 500-ms silent gap, resulting in 2000-ms long intervals (only 1500 ms of which contained sound). The second tone always the higher-pitched of the two.

A musical interval in common Western musical practice is defined by the frequency ratio (measured in terms of a logarithmic frequency variable termed "cents") between its constituent tones, rather than by the absolute frequencies of the tones. This feature allows us to construct intervals that are invariant in the category they belong to, but are made from tones with different frequencies. The stimulus set we constructed thus varied along 2 orthogonal dimensions. In the first dimension ("interval quality"), the frequency ratio between the higher- and lower-pitched tones varied, with ratios derived from equally tempered semitones (in which each 100 cents corresponds to a semitone, and the 3 intervals we used, minor third, major third, and perfect fourth, correspond to 300, 400, and 500 cents, respectively). These values ranged from 287.5 to 512.5 cents, with stimuli generated at 12.5-cent increments (see Fig. 1). This range spanned and included minor thirds, major thirds, and perfect fourths, all of which are common and important intervals in western music.

In the second ("pitch height") dimension, which is orthogonal to the first, the frequency values of the intervals were roved in absolute pitch space (e.g., a 400-cent major third can be generated with base

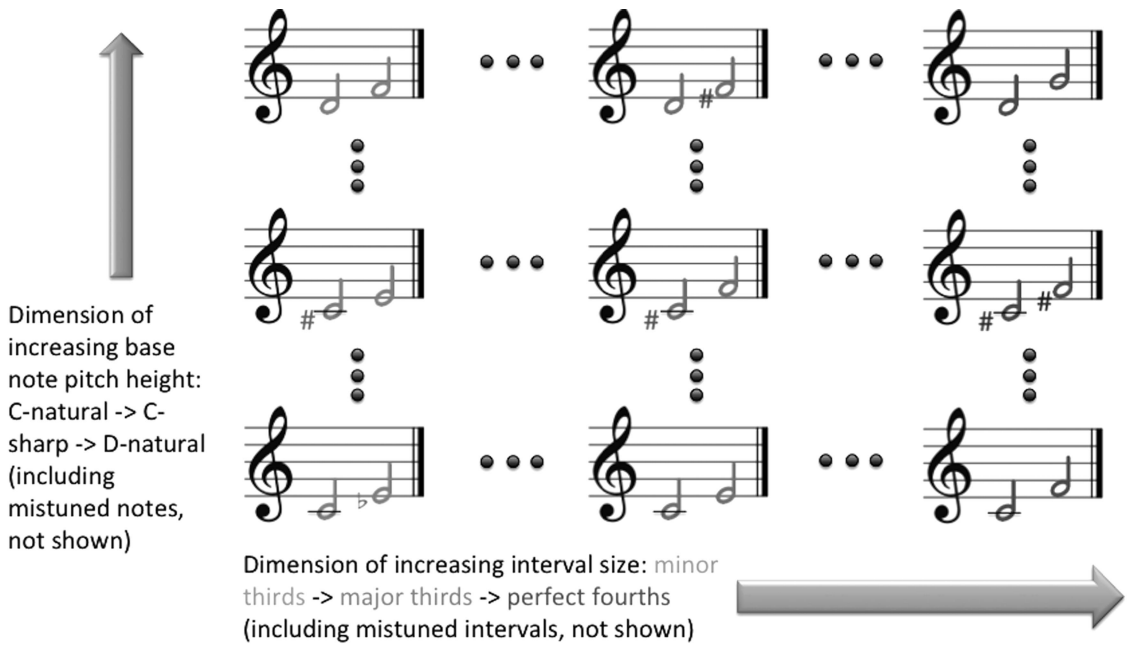

Figure 1. (sound stimuli) Schematic of auditory stimuli used in the behavioral and imaging experiments. The imaging study used only 9 pictured stimuli, while the behavioral pretest included those 9 in addition to many additional sounds that were "mistuned" between standard frequencies and standard frequency ratios (indicated by ellipses). Movement along the *x*-axis indicates a change in interval size (i.e., frequency ratio between 2 notes), but no change in the pitch of base notes. Movement along the *y*-axis indicates a change in the pitch of both notes, but no change in the frequency ratio of an interval's 2 notes.

notes of C-natural, C-sharp, mistuned notes between C-natural and C-sharp, or any other frequency). Thus, without affecting the quality of the intervals along a minor third <-> major third <-> perfect fourth dimension, intervals were generated with base notes that varied from 259.7 Hz (slightly below middle C) to 295.8 Hz (slightly above the D 2 semitones above middle C). The second note of each interval was then independently calculated according to whichever frequency ratio (from the interval quality dimension) we wished to implement. Thus, interval quality could be manipulated without affecting the absolute pitch of intervals, and vice versa.

In general, the chosen approach to examining CP was to (a) create a set of sounds that were shown to be perceived categorically, (b) create an orthogonal extension of this first set that was acoustically well matched but not categorically perceived, (c) take the "hallmark" exemplars from each spectrum and present them within an functional magnetic resonance imaging (fMRI) paradigm, and (d) examine differences in how well machine learning algorithms were able to decode within the experimental versus the orthogonal sets. CP, specifically, was screened for in the behavioral experiment [(a) and (b)] by making use of the continuous feature space in both the interval size and pitch height dimensions. Afterwards, a subset of 9 of these sounds was used during the fMRI experiment: the 3 "true" (nonmistuned) intervals (300-cent minor third; 400-cent major third; and 500-cent perfect fourth), each of which were synthesized with base notes of exactly C-natural, C-sharp, and D-natural (3 × 3 design, see Fig. 1). We chose to use an approach comparing and contrasting primary and orthogonal stimulus dimensions (interval quality vs. absolute pitch), as both could be manipulated via the same simple feature: frequency of constituent tones. Links could then be made between behavioral divergence and differing patterns of fMRI results. Three-category classification was chosen over more common 2-category experimental designs (which are often required in speech experiments due to the multidimensional nature of phoneme space) in order to: (1) generalize imaging results beyond a single pair of musical categories and (2) demonstrate behavioral CP that is clearly differentiable from anchoring/endpoint effects, mediated by short-term memory [see Hary and Massaro (1982) and Schouten (2003) for common criticisms of 2-category perceptual tasks].

Prescanning Behavioral Tasks

In our behavioral pretest, study participants were asked to perform a series of 4 tasks (2 identification tasks and 2 discrimination tasks). For each of these tasks, participants performed a practice run (2–5 min) to ensure that they were comfortable with the response interface and understood the instructions. For each type of task (e.g., identification, which was performed twice), participants heard the identical set of stimuli both times, but they were asked to attend to different qualities of the sounds (e.g., "listen for interval quality" or "listen for pitch of base note"). The experiment was counter-balanced, so that half of the participants performed tasks (1) and (2) prior to (3) and (4), with the other half first performing (3) and (4).

1. Identification of interval quality.

   Prior to performing the task, participants were asked to listen to a series of exemplars of each of the 3 true interval qualities. Ten examples of minor thirds were presented, all of which had 300-cent frequency ratios but varied randomly in pitch height, while the phrase "minor thirds" was displayed on the screen. This was immediately followed by 10 examples of major thirds and perfect fourths, respectively. For the task proper, participants were asked to simply assign each interval with a label by pressing a keyboard key: "j" for minor third; "k" for major third; and "l" for perfect fourth. Participants were asked to select whichever label an interval was closest to. Responses were not under time constraints, but participants were asked to make their selections as quickly as they could comfortably do so. After a response was logged, the next trial would begin after a delay of 2000 ms. For the practice run only, responses were followed by a visual displaying the participants' choice (e.g., "you selected major third") and the actual physical property of the interval (e.g., "the interval was closest to a major third"). No feedback was provided during post-practice runs. Nineteen intervals were presented in a pseudorandom order, with each interval type presented 4 times for a total of 78 trials. The pitch height for each interval was generated pseudorandomly.
2. Discrimination of interval quality.

   Participants were presented with pairs of intervals and asked to judge which of the 2 intervals was "wider" (i.e., whether the first- or second-presented interval had more separation between low and high notes). This instruction therefore does not constrain the listeners' judgment with respect to the categories that they may be familiar with. Participants were instructed to press "j" or "k" if they believed that the first- or second-presented interval met this criterion, respectively. The ratio between the 2 intervals of a trial always differed by 25 cents. Trials were balanced so that "j" and "k" were the correct responses an equal number of times, and so that the interval with the higher-pitched base note appeared first or second an equal number of times. As in (1), the intervals were presented in a pseudorandom order. The orthogonal dimension of pitch height for each interval was generated pseudorandomly, with an additional stipulation that the base notes of the 2 intervals in any one trial must differ by at least 37.5 cents in order to safeguard against the possibility of participants basing their judgments solely on the pitch of the intervals' top notes (in a situation where both intervals used identical or near-identical base notes). As in the identification task, participants first performed a practice run, where they were given visual feedback after each trial (e.g., "Incorrect: you selected the first interval and the second interval was wider"). No feedback was provided during the 5 post-practice runs (each run containing 17 trials, one for each discrimination pair, presented in a pseudorandom order).
3. Identification of pitch height.

   The stimuli used in this task were identical to those from (1). Participants were asked to attend not to quality of the intervals (minor, major, and perfect), but instead to the pitch of the base notes. (The 2-tone intervals were still used, but participants were instructed that they could ignore the top tone of each interval.) Prior to performing the task, participants were asked to listen to a series of exemplars of each of 3 base notes: C-natural, C-sharp, and D-natural. Ten examples of intervals with base notes of C-natural were presented, all of which had variable top notes, while the phrase "C-naturals" was displayed on the screen. This was then immediately followed by 10 examples of C-sharps and D-naturals, respectively. For the task proper, participants were asked to simply assign each presented base note with a label by pressing a keyboard key: "j" for C-natural; "k" for C-sharp; and "l" for D-natural. Participants were asked to select whichever label the presented sound was closest to. Feedback was given for a practice run (e.g., "you selected C-sharp, the presented sound was closest to D-natural"), but not the post-practice runs. All other methods were identical to those used in (1).
4. Discrimination of pitch height.

   Participants were presented with pairs of intervals and asked to judge which of the 2 intervals had a higher-pitched base note. As in (3), subjects were told that they could complete the task successfully without considering the top notes of the intervals, which were chosen pseudorandomly. As in all prior tasks, participants first performed a practice run, where they were given visual feedback after each trial (e.g., "correct: you selected the first interval and the first interval had the higher-pitched base note"). All other methods were identical to those used in (1–3).

Participants were chosen for the MRI experiment based on the degree of difference between peak and trough discrimination accuracy in task (2). Specifically, participants were screened to have an "M"-shaped interval quality discrimination function, with performance troughs near category centers (e.g., near 400 cents/"major third") and performance peaks far from these centers (e.g., near 450 cents/midway between "major third" and "perfect fourth"). This function shape, with discrimination accuracy peaks near hypothesized category boundaries, is characteristic of CP in speech and other domains (Liberman et al. 1957; Burns and Ward 1978). Performance peaks are thought to occur when the 2 stimuli in a discrimination task pair span such a boundary, with long-term memory systems assigning "all or nothing" labels to the sounds, which perceptually diverge.

fMRI Tasks and Data Acquisition

MRI volumes were acquired on a 3-T Siemens Magnetom Trio scanner. A high-resolution (voxel = 1 mm^3) T_1-weighted anatomical scan was obtained for each participant. For each functional trial, one whole-head frame of 39 contiguous T_2*-weighted images was acquired in an ascending, interleaved fashion (time repetition = 9.5 s, time echo = 30 ms, 64 × 64 matrix, voxel size = 3.5 mm isotropic), yielding a total of up to 351 BOLD volumes per subject (9 runs × 39 volumes/run). fMRI scanning was performed via a sparse temporal sampling protocol (Belin et al. 1999), where each trial consisted of 2000 ms of data acquisition that followed 7500 ms of relative quiet. In 90% of trials, a single melodic interval was presented 3 times for a total of 6 tones during this quiet time period, with each 750 ms tone followed by 500 ms of silence, and 250 ms of silence bookending the initial and final tones. Unlike the behavioral pretest, which utilized pitches that were mistuned between standard notes and ratios that were mistuned between semitones, the MRI protocol employed only intervals that started on 3 standard base notes ("middle" C natural, C sharp, and D natural) and used 3 standard interval ratios (300-cent minor thirds, 400-cent major thirds, and 500-cent perfect fourths). This 3 × 3 design yielded a set of 9 unique sound samples as stimuli. Subjects were not asked to explicitly or implicitly identify intervals according to the interval quality or base note. Instead, they performed an orthogonal task in which they were asked to listen attentively and to press a response button upon hearing a trial that contained only 5 tones instead of 6 (10% of trials). Such oddball/catch trials were used as a check on attention/alertness and these imaging data were discarded. This experimental protocol was chosen above an overt identification or discrimination task in order to look at processes that occur relatively automatically.

Each functional run consisted of 39 trials (and thus generated 39 BOLD volumes). After an initial silent trial, 4 pairs of silent baseline trials (9 silent trials in total) were interspersed between 3 sets of 10 experimental trials (one trial for each of the 9 unique sound samples, and one catch trial). These 10 trials were presented in a pseudorandom order, with the main constraint being that any one interval could not follow a trial using the same interval type or base note (e.g., a major third starting on D natural could not follow a major third starting on C sharp or C natural, and could not follow a minor third starting on D natural or a perfect fourth starting on D natural). This constraint was imposed to avoid potentially confounding adaptation effects.

Nine 39-trial runs were conducted, each of which contained sounds in a unique order of presentation. Each participant underwent each of the 9 runs, with half the participants performing the runs in the opposite order from the other half. Of the 10 participants enrolled in the MRI study, 6 completed the protocol exactly as planned. For 3 of the 10 participants, one run had to be discarded due to inattention (failure to press response button for at least 2 of the run's 3 catch trials). For 1 of those 3 participants, an additional run had to be discarded due to failure to comply with the instructions. The fourth participant's data had to be discarded due to an equipment malfunction.

GLM Analyses

A set of GLM analyses were performed in order to (1) determine cortical regions that were activated by sound (i.e., sound > silence contrast) and (2) to perform between-condition subtractions (e.g., major > minor) to compare with MVPA results. Standard GLM-based analyses were performed using FSL4's fMRI expert analysis tool (FEAT) (http://www.fmrib.ox.ac.uk/fsl/feat5/index.html). Preprocessing steps consisted of motion correction using MCFLIRT; nonbrain removal using brain extraction tool (BET); and spatial smoothing using a Gaussian kernel of full-width at half-maximum (FWHM) 7.0 mm. For each analysis (interval quality or pitch height), a design matrix was generated with one predictor for each category of stimulus (e.g., in the column for "minor," an "1" was assigned for all volumes following the presentation of minor intervals and a "0" for all other volumes). As part of FEAT, native space images were registered to the Montreal Neurological Institute (MNI) space using FNIRT. Following the first-level analysis, individual subjects' runs were combined using a second-level, fixed-effects analysis. Third-level between-subjects analyses were performed using FSL's FLAME mixed-effects model. Specific one-tailed contrasts were performed twice for each of 3 condition pairs in both the interval quality (e.g., minor > major) and the pitch height (e.g., C-natural > C-sharp) analyses. *Z*-(Gaussianised T/F) statistic images were thresholded using Gaussian Random Field theory-based maximum height thresholding with a (corrected) significance threshold of $P = 0.05$ (Worsley et al. 2002). [Note that these analyses were performed once using the entire cortical space, and a second time on a restricted region of interest (see next section) in order to provide the fairest possible comparison with MVPA results.]

MVPA Procedures

Prior to the main analyses, motion correction was performed by realigning all BOLD images with the first frame of the first run following the T_1-weighted scan (generally the fifth or sixth functional run) using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). An MVPA was performed on single-subject data in native space, prior to nonlinear registration using the MNI/ICBM152 template (performed with FSL4's FNIRT tool: http://www.fmrib.ox.ac.uk/fsl/fnirt/index.html), and a standard top-level between-subjects analysis, performed with SPM8.

The MVPAs were performed using the Python programming language's PyMVPA toolbox (Hanke et al. 2009) and LibSVM's linear support vector machine (SVM) implementation (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). Each participant's runs were concatenated to form a single long 4D time series (up to 351 3D volumes). Note that no spatial smoothing/blurring was performed on the functional data prior to MVPA. A text file was generated assigning each volume a run (1–9) and a condition (minor, major, or perfect for the interval quality analysis; C-natural, C-sharp, and D-natural for the pitch analysis). Within each run, we performed (a) linear detrending to remove signal changes due to slow drift and (b) *z*-scoring to place voxel values within a normal range (Pereira et al. 2009). As SVMs are pairwise classifiers, we ran individual analyses on pairs of 2 conditions (e.g., minor vs. major; C-sharp vs. D-natural). The final preprocessing step was to perform temporal averaging (Mourao-Miranda et al. 2006) on the BOLD data; we used 3 ->1 averaging, combining 3 images (e.g., all perfect fourths from the first 1/3 of a functional run) into a single image. SVMs for interval quality comparisons were both trained and tested using intervals from all 3 pitch height classes; the reverse is true for pitch-height analyses. Classification was performed using leave-one-out cross-validation, where a classifier was trained on data from 8 of the functional runs and tested on data from the 9th, and the procedure was then repeated 8 times testing on a novel run each time. SVM classification was performed using a searchlight procedure (Kriegeskorte et al. 2006), whereby the decoding algorithm considers only voxels from a small sphere of space (radius = 3 voxels, up to 123 voxels in a sphere). [While accuracy has been shown to generally increase along with the size of searchlight spheres (Oosterhof et al. 2011), we chose a radius of 3 voxels as a compromise between classifier performance and spatial specificity.] An accuracy score [percentage above chance (50%) that the classifier was able to successfully identify category] was calculated using an average of the 9 cross-validation folds, and this value was assigned to the center voxel of the sphere. This procedure was repeated using every brain voxel as a searchlight center (~35 000–45 000 spheres), yielding local accuracy maps for the entire brain. As the primary interest was in observing abstracted category representation (and not that of specific sound pairs), at this stage accuracy maps for each subject were averaged across the 2 pairwise classifications (i.e., minor third/major third maps were averaged with major third/perfect fourth maps). Minor third/perfect fourth classification was not performed, as these 2 stimuli sets differed more so from one another in physical and category distances (2 semitones) than the other 2 pairs (1 semitone) and would have added an additional confound to the analysis. A parallel averaging step was performed for the pitch height analysis: accuracy maps for C-natural/C-sharp discrimination were combined with those for C-sharp/D-natural discrimination. We note that certain MVPA studies that compare all possible decoding pairs (e.g., Formisano et al. 2008; Kilian-Hütten et al. 2011) often examine identity of auditory objects that have no inherent "ordinal" quality (i.e., whereas perfect fourths are larger than major thirds, which are larger than minor thirds, voice identities differ from one another in myriad ways that are difficult to rank), and thus do not need to consider this particular confound.

Prior to performing group-level analyses, participants' brain masks were generated with FSL4. The averaged accuracy values, which served as effect sizes for the group-level analysis, were then linearly transformed into a subject's native anatomical space before being nonlinearly transformed into standard space using FSL4's linear and nonlinear registration tools (FLIRT and FNIRT). While there is an inherent smoothness to the searchlight MVPA procedure, at this stage we explicitly smoothed each subject's accuracy maps (a 7-mm FWHM isotropic Gaussian kernel) in order to best account for intersubject brain variability and to perform and interpret group-level statistics. The registered and smoothed accuracy maps were then input into SPM8, which output group-level *t*-statistics for each voxel.

The threshold for statistical significance was set voxelwise at $t > 7.98$ (corrected for multiple comparisons, family-wise error (FWE) < 0.05, $n = 9$). While data were collected and are presented for the entire cortex, significance testing was performed on a restricted volume in line with the a priori hypothesis of involvement of the right STS/left IPS, based on results from an earlier study (Klein and Zatorre 2011). This mask was created off a standard MNI152 anatomical image by delimiting the full extent of the gray matter in these regions. This approach was used due to a lack of consensus in the MVPA/searchlight literature about a methodology for setting accurate group significance thresholds that are not extremely conservative [a full-brain, purely between-subjects ($n = 9$) analysis using random field theory thresholding yields a *t*-statistic cutoff above $t = 16$]. Stated another way, there is no set method outlined for determining a "smoothed variance ratio" (Worsley et al. 2002) for these data, as is often implemented in standard GLM analyses. Because of this ambiguity and for completeness we have also reported all peaks comprised of voxels significant at $P < 0.001$ (uncorrected), with at least 10 contiguous voxels meeting this criterion. All of these peaks would generally be considered statistically significant in a standard GLM analysis (*t*-statistics > ~5) and, while some do not meet the very conservative threshold used here, we believe that the nearly symmetric positioning and large spatial extents of the parietal and temporal peaks (see Results section) lend weight to the argument that a substantial portion of these $t < 7.98$ results are not merely false positives.

Separately, as a check on searchlight statistical procedures, the voxels within the previously described ROI mask (left IPS and right STS, anatomically defined based on the results found in Klein and Zatorre 2011) were also used within an "Monte Carlo" permutation test. For each subject, the identity labels for training examples were randomly scrambled and tested 1000× in order to generate subject-by-subject null distributions. These analyses yielded a single ROI decoding value for each subject (determined without label scrambling), which could be (a) compared with the subjects' null distributions to generate subject-wise *P*-values and (b) input into a group-level *t*-test. Three-category MVPAs (m3/M3/P4, 33.3% chance accuracy) were performed in order to generate and report a single *P*-value per subject. While the experiment was not specifically designed to test for single-subject significance (and complex feature elimination procedures were not employed), these permutation tests were performed to provide converging evidence for categorical decoding using markedly different procedures than with the primary searchlight analyses.

Results

Identification

Figure 2 shows identification functions for 3 representative subjects for both interval quality (Fig. 2*b*) and pitch height (Fig. 2*c*). Graphs are shown for individuals, in addition to the $n = 10$ group data (Fig. 2*a*), as averaging necessarily obscures the sharp boundaries of the functions due to individual variability in the location of boundaries and category centers. Three obvious labeling plateaus are evident in the plot for interval quality, but not for pitch height. To quantify the degree to which participants' identification task responses were "categorical," we first generated a "triple-plateau" function, which served as a model "perfectly" categorical response. Identification responses were recoded as 1, 2, and 3 for minor third, major third, and perfect fourth, respectively (likewise for the pitch height task). The model function was created by labeling intervals from 287.5 to 337.5 cents as "1," 350 cents as "1.5," 362.5 to 437.5 cents as "2," 450 cents as "2.5," and 462.5 to 512.5 cents as "3." The 1.5 and 2.5 values were chosen as these sound stimuli were physically exactly half way between the exemplar sound tokens. For each participant, we calculated difference scores between that participant's response function and the ideal function (one each for interval quality and pitch identification). Participants performed the interval quality identification in a significantly more categorical manner than the pitch task, as judged by proximity to the model function ($df = 36$, paired-sample 1-tailed *t*-test, $P = 0.00018$). While we ultimately selected our 10 MRI participants based on their discrimination task responses, this group also performed the interval quality task in a significantly more categorical manner than the screened-out cohort of 27 participants ($df = 35$, unpaired-sample 1-tailed *t*-test, $P = 0.0035$).

Discrimination

For the interval quality discrimination task, we screened for subjects showing an M-shaped function with peaks at or near theoretical categorical boundaries (e.g., 337.5 vs. 362.5 cent discrimination) and troughs at or near the categorical centers (e.g., 387.5 vs. 412.5 cent discrimination) (see Fig. 2*d*, dotted gray line). This task proved very difficult for participants due to the variability in pitch space (i.e., while the 2 intervals in a given trial were always 25 cents apart, the base notes of those intervals could be separated by as much as 2 semitones, with an average spacing of about 1 semitone). However, a subset of our sample did show this M-shaped function [and to a significantly greater degree than for pitch height discrimination ($df = 9$, paired-sample 1-tailed *t*-test, $P = 0.0058$)] (see Fig. 2*e*,*f* for discrimination functions of 3 individuals, presented for identical reasons as stated above). These same subjects discriminated all interval pairs near category boundaries ("between-categories") with significantly greater success than those near category centers ("within-categories") (80% correct vs. 67% correct, paired-sample 1-tailed *t*-test, $df = 9$, $P = 0.0016$). The same effect was not present for pitch height discrimination (78% correct vs. 72% correct, paired-sample 1-tailed *t*-test, $df = 9$, $P = 0.25$), with lower accuracies occurring only near the ends of the function (i.e., an "inverted U," not an M-shaped function), indicative of short-term memory/attentional-based anchoring effects. These 10 subjects were then enrolled in the functional imaging experiment.

We did not expect the majority of musicians to show a clear categorical discrimination function, due to prior research (Zatorre and Halpern 1979), indicating a very high degree of task difficulty when using intervals with roving pitches. Our sample of 37 was screened not with an intention to generalize results to a larger population, but instead to select for those individuals demonstrating clearest evidence for CP of musical categories. While it is highly likely that the use of a nonroving lower pitch would have greatly enhanced the observable CP qualities of task performance in more subjects, it would not have allowed for simple abstraction beyond specific frequencies and note pairs.

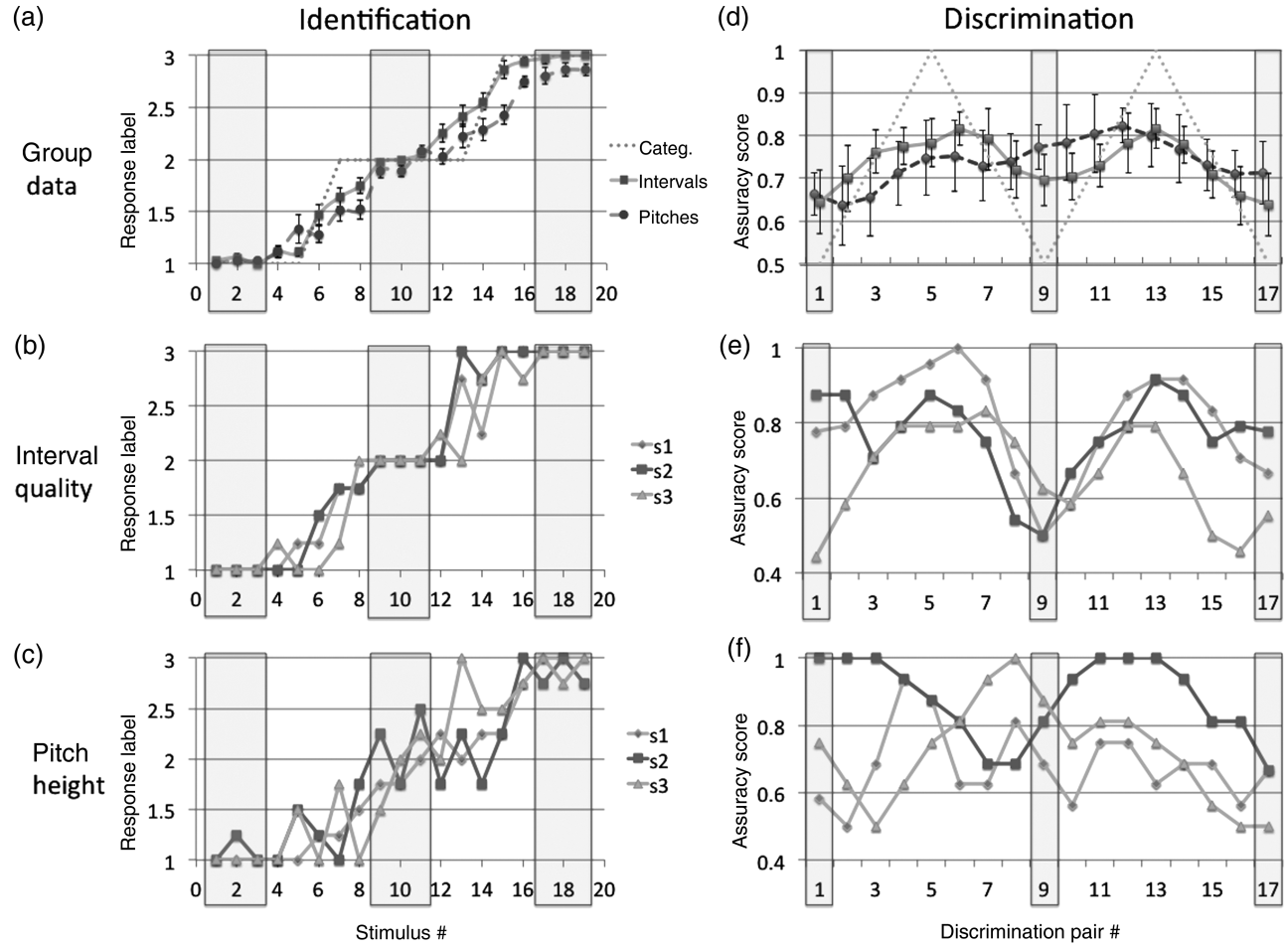

Figure 2. (behavioral results) Behavioral results for identification (*a–c*, left column) and discrimination (*d–f*, right column). The top row depicts $n = 10$ group data for the subjects enrolled in the imaging experiment; the middle row depicts interval quality identification and discrimination for 3 representative subjects (s1, s2, and s3); and the bottom row depicts pitch height identification and discrimination for those same 3 subjects. For the "identification" charts, the *x*-axis represent the stimulus number, corresponding to musical interval size (frequency ratio) and the *y*-axis, the participants' averaged responses (where 1, 2, and 3 = identification as low, mid, and high tokens, respectively). Theoretical category centers for the stimuli were at *x*-axis positions 2, 10, and 18, which represent canonical intervals or pitch heights. As in (*b*), which depicts results solely from the interval quality analysis, the *x*-axis positions 2, 10, and 18 correspond to the interval qualities of minor third (300 cents), major third (400 cents), and perfect fourth (500 cents), respectively. In (*c*), which depicts results from the pitch height analysis, those same 3 *x*-axis positions correspond to base note pitch heights of C-natural, C-sharp, and D-natural, respectively. (*a*) contains results from both analyses. Approximate theoretical category centers, defined for simplicity as the standard/token intervals or pitches ±1 mistuning from the standard, are indicated in gray boxes (e.g., stimulus 17, 18, and 19, corresponding to intervals of 487.5, 500, and 512.5 cents, respectively). For the group data, theoretically perfect categorical functions (gray dotted lines) are shown alongside perceptual functions for interval quality (blue solid lines) and pitch height (red dashed lines), with error bars showing SEMs. For the "discrimination" charts, the *x*-axes also represent stimuli number, although these stimuli are now "pairs" rather than single intervals (e.g., stimulus 9 for the interval quality analysis depicts trials for discrimination of 387.5 vs. 412.5 cents). The *y*-axes are averaged accuracies of subjects' responses, where 1 indicates 100% correct and 0.5 indicates chance-level performance. For both the identification and discrimination tasks, group pitch height data deviated from the ideal categorical functions to a significantly greater degree than the interval quality data (see Results section). For the individuals, clear identification labeling plateaus were observed for interval quality (*b*) but not for pitch height (*c*) and, likewise, clear M-shaped discrimination functions were observed for interval quality (*e*) but not for pitch height (*f*).

GLM Analysis

A contrast of all sound > silence (excluding volumes following presentation of the rare/oddball 5-tone stimuli) revealed 3 large significant clusters: (1) the right STG/STS (3242 voxels); (2) the left STG (2290 voxels); and (3) the left/right supplementary motor area (974 voxels, cluster spans the interhemispheric fissure, but contains more voxels in the left hemisphere). No statistically significant group activation peaks were observed anywhere in the brain for any pairwise contrast in either the interval quality or pitch height analyses, which was predicted due to the high degree of physical similarity between all 9 sound samples used in the imaging experiment.

Searchlight fMRI

Group-level searchlight results showed significant accuracy peaks in the right STS ($t = 9.34$; $x, y, z = 48, -14, -14$) and left IPS ($t = 9.93$; $x, y, z = -30, -50, 46$; see Fig. 3). No other brain regions contained voxels that surpassed $t = 7.98$. The number of contiguous voxels that passed a $P < 0.001$ uncorrected threshold were similar in these 2 regions: 66 voxels (left IPS) and 53 voxels (right STS). We also note, both because of spatial extent and approximately symmetrical locations to the 2 significant peaks, a region in the right IPS ($t_{max} = 5.24$; $x, y, z = 36, -54, 46$; 57 contiguous voxels at $P < 0.001$ uncorrected) and the left STS ($t_{max} = 5.09$; $x, y, z = -50, -14, -16$; 15 contiguous voxels at $P < 0.001$ uncorrected). No other cortical

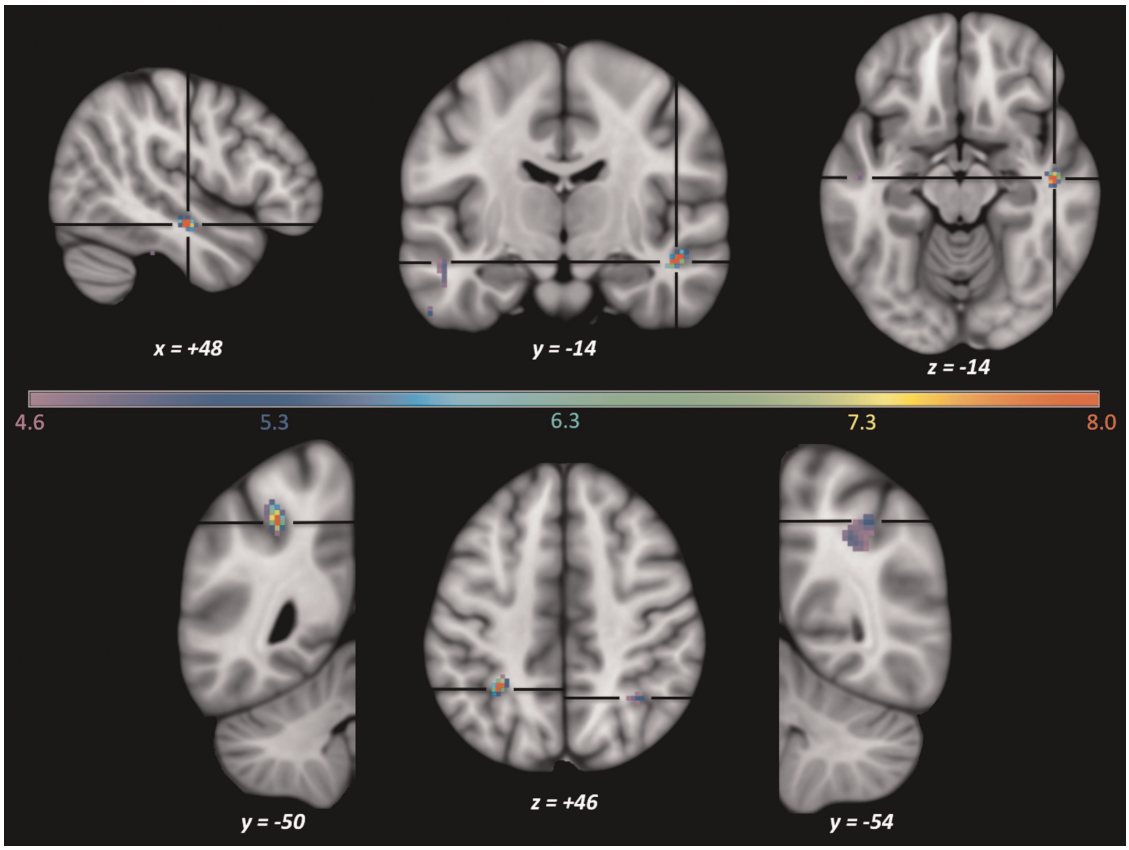

Figure 3. (searchlight imaging results) Group-level ($n = 9$) statistical peaks for the searchlight decoding analysis for interval quality, overlaid on an MNI152 0.5-mm T_1 anatomical image. Colored voxels indicate t-statistics ranging from $t = 4.6$ (violet, $P = 0.001$ uncorrected) to $t = 8.0$ (red, $P = 0.05$ corrected for multiple comparisons). The top panel shows results from the right (and left) STS. The bottom panel shows results from the left and right IPS. All voxels depicted in deep red (situated in the left IPS and right STS) are statistically significant ($t > 7.98$).

regions contained 10 or more contiguous voxels surpassing the $P < 0.001$ uncorrected threshold.

No significant group-level accuracies were observed anywhere in the brain for the pitch height discrimination pairings.

We next examined raw classification accuracies (i.e., effect sizes) in the peak voxels of these 4 regions. Looking at 9-subject averages (chance-level accuracy = 50%), we observed accuracies of 55.8% in the right STS (individuals ranged between 53.1% and 59.1%), 56.1% in the left STS (range 49.0–59.7%), 57.1% in the right IPS (range 49.9–63.0%), and 55.2% in the left IPS (range 53.2–58.0%; see Fig. 4). These individual values are presented for description only: Statistical significance testing was assessed solely via group analyses (performed naively over the entire cortical space).

We note that the larger t-values in the right STS/left IPS appear to be driven by smaller variability (rather than larger effect sizes) compared with the analogous peaks in the opposite hemispheres. The overall average decoding accuracy of all cortical searchlight spheres in all 9 subjects for the minor/major and major/perfect discriminations was near chance at 50.5%, which suggests a combination of a chance distribution (centered at 50% correct, underlying the vast majority of spheres) and the smaller number of information-containing spheres (accuracy > 50% correct). This indicates no consistent brain-wide over-fitting in the decoding analyses, which would have led to artificially high "null" decoding averages.

Intersubject variability gave rise to small spatial dissociations between maximum average accuracy peaks (i.e., group beta values) and statistical peaks (i.e., group t-values). In the same local neighborhoods as the statistical peaks, we observed

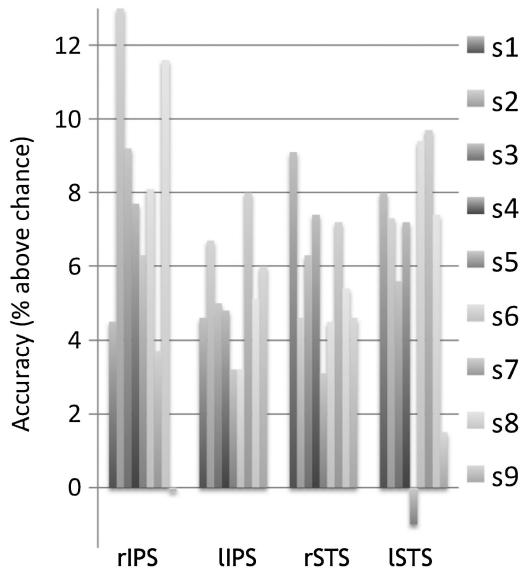

Figure 4. (individual searchlight results) Single-sphere decoding accuracies (% above chance) for each of the 9 fMRI study participants at locations determined by group statistical peaks. (Individuals' maximally decodable spheres have higher accuracies, but variable locations.) The locations of the sphere centers in the MNI space are $x, y, z = 36, -54, 46$ (right IPS); $x, y, z = -30, -50, 46$ (left IPS); $x, y, z = 48, -14, -14$ (right STS); and $x, y, z = -50, -14, -16$ (left STS).

local peaks in average classification of 57.6% in the right STS ($x, y, z = 52, -2, -14$); 56.9% in the left STS ($x, y, z = -50, -10, -16$), 58.5% in the right IPS ($x, y, z = 32, -54, 42$); and 57.7% in the left IPS ($-34, -48, 50$). For the left STS, right IPS, and left IPS, the spatial distances between beta and *t*-statistic peaks are very small (4, <6, and <7 mm, respectively). While this distance is somewhat larger for the right STS (12–13 mm), the maximum group average peak is still clearly positioned in the sulcus (in a more anterior position).

Permutation Test

We observed decoding accuracies above chance for the 9 individuals at +2.9% ($P = 0.139$), +3.8% ($P = 0.099$), +4.2% ($P = 0.075$), +3.4% ($P = 0.135$), +6.6% ($P = 0.020$), +3.8% ($P = 0.097$), +2.4% ($P = 0.166$), −4.2% ($P = 0.682$), and +4.1% ($P = 0.085$). Even though the experimental protocol was not designed to test significance at the single-subject level (and additional feature elimination was not performed within the ROI), 8 of 9 subjects showed positive trends ($P < 0.17$). Furthermore, just as with the searchlight analyses, permutation testing suggested small yet consistent effects in individuals, which reached statistical significance when considered as a group. Inputting the 9 decoding accuracies into a 1-tailed single-sample *t*-test, we observed a significant group-level effect at $P = 0.008$ (degrees of freedom = 8).

Discussion

CP has repeatedly been demonstrated to be a robust behavioral phenomenon using both speech and certain nonspeech auditory stimuli. However, a combination of the limitations of available experimental protocols and analytic methods, as well as a general focus on speech-specific process, has left ambiguous the identification of its full neural correlates. The sample of trained musicians presented here demonstrates behavioral CP functions for musical intervals, while MVPA of their functional brain data implicates local information-containing regions in the superior temporal and intraparietal sulci in the representation of abstract musical interval categories. The right STS and left IPS were also highlighted in an earlier study (Klein and Zatorre 2011), despite the use of dramatically different experimental designs (active discrimination vs. a more passive orthogonal task) and analysis strategies (magnitude-based contrast analysis vs. multivariate classification algorithms). These regions thus demonstrate locally distributed response patterns linked to specific musical categories and theoretically comprise important regions in a cortical network for sound categorization. These results argue that such a network is recruited automatically for some types of nonspeech auditory processing.

The STS may serve a critical early role in a ventral stream of information processing, with particular links having been made between left STS and phoneme perception (Liebenthal et al. 2005; Joanisse et al. 2006). The present results suggest bilateral STS processing for musical intervals, with a right hemispheric bias, thus generalizing the role of the STS beyond the speech modality. The right STS may subserve an early "post-auditory" stage of processing (Pisoni 1975; Zatorre 1983), where continuous acoustic signals are converted to invariant "all-or-nothing" codes. These invariant, over-learned categorical memories may be mediated by Hebbian neural population codes (Hebb 1949) distributed over many of voxels in a region. Triggering of these population codes may result in robust invariant BOLD responses, visible above noise to classification algorithms. While these analyses do not allow a full review of the spatial extent of these putative population codes, a sufficient portion (as defined by decoding success) of the circuits appears to exist at scales similar to the size of the searchlight spheres (~123 3.5 mm^3 isotropic voxels, about 5 ml). The left STS, less implicated here, could be performing a parallel stream of categorical processing, tuned for different features of the signal. Alternatively, left STS response patterns could be representative of (a) interhemispheric communication or (b) access to the verbal lexicon, as these musical categories cannot be completely dissociated from their names (e.g., "minor"). We do not believe that the STS is the exclusive mediator of musical CP, but instead plays a dominant role in the ventral stream component of categorical processing.

The use of multiple intervals with variable pitch height ensured that these putative category maps represented abstract features beyond specific sound samples, instead reflecting learned relative pitch relationships between musical notes. Classifiers were trained and tested blind to the specific pitch classes (pitch height) of the musical intervals and thus were only able to utilize information related to category membership, rather than absolute pitch information. In fact, as specific tones were not repeated within interval categories, but were reused across categories, the SVMs had to learn to largely "disregard" absolute frequency-driven features. While categorical distinctions are not requisite for successful MVPA, null imaging results from the pitch height analysis suggest that, here, categorical quality is the distinguishing feature detectable by the classifier. The absence of MVPA results in the pitch height analysis could be due to the use of sound stimuli that (a) were highly physically similar to one another and (b) exhibited considerable overlap in the frequencies of their note pairs, and suggests that top-down, memory-based processing may be the critical component in eliciting a robust stable BOLD response pattern in the interval quality analysis. While subjects did show some ability to "identify" sounds based on pitch height, they did not demonstrate the clear labeling plateaus consistently found for interval identification (see Fig. 2). This finding, as well as the lack of an M-shaped function for pitch height discrimination, is highly indicative of short-term anchoring effects, as opposed to access to an over-learned long-term memory store. Likewise, the lack of significant orthogonal MVPA results suggests that fMRI classification success may rely heavily on the degree of perceptual differentiability, which may, in turn, originate from either bottom-up or top-down processes. [Although, at least in the visual domain, BOLD data have been used to decode certain physical stimuli even in the absence of conscious perceptual differences (Haynes and Rees 2005; Kamitani and Tong 2005).] A recent MVPA study (Lee et al. 2011) of "nonmusician" subjects using melodic musical stimuli did not yield significant decoding for minor vs. major sounds, suggesting that these categorical processes are highly experience-driven, in accordance with previous behavioral studies demonstrating little or no categorical musical perception in nonmusicians (Burns and Ward 1978; Zatorre and Halpern 1979).

A recent speech study (Kilian-Hütten et al. 2011), meanwhile, used a categorical "midpoint" approach to demonstrate CP via auditory recalibration in the absence of acoustic

differences between stimuli. We considered a related approach for this study: demonstration of MVPA differences following the presentation of stimuli that varied by a single continuous physical parameter, yet were perceived as members of 2 discrete categories. This alternative approach is powerful in that it minimizes acoustically driven confounds, but does not easily generalize beyond the examined category pair. Thus, in order to drive generalizability, we chose to demonstrate categorical versus noncategorical processing via an orthogonal, absolute pitch roving dimension. Unlike other auditory "objects," where absolute frequency may be largely unrelated to the dimension of interest [e.g., sound identity (Staeren et al. 2009)], simple pitch values, critically, define the category identity of musical intervals and thus can be manipulated to form the basis for both the experimental and orthogonal stimuli dimensions. Thus, to make the MVPA results as generalizable as possible, we chose to test the categorical component of the analysis across 3 categories, and to dissociate the acoustically driven, noncategorical component by way of a second tone frequency-based dimension.

The decoding results, which provide evidence that such categorical information is present in the STS (but do not show any such evidence for the STG), stand in contrast with recent studies, suggesting that early auditory areas mediate complex, object-based processing (Staeren et al. 2009; Kilian-Hütten et al. 2011; Ley et al. 2012). These recent studies are all excellent demonstrations of MVPA's ability to reveal that auditory cortex is involved in classification of sounds, but say less about true categorical—"perception"—as classically defined, as none reported results from behavioral discrimination tasks. The STS results presented here support a hierarchical auditory ventral stream processing model (Hickok and Poeppel 2007), which is not necessarily contradictory to architectures that may also contain myriad feedback/forward connections and parallel processing stages. The null results in and around Heschl's Gyrus (HG) may be due in part to the use of standard BOLD voxel size (3.5 mm isotrophic, as opposed to <2 mm voxels used in certain studies) or a high degree of variance in the shape/location of tonotopic maps in individuals. (Voxel size here was chosen as a compromise between relatively small size and full-brain coverage.) We therefore do not dismiss the idea that early auditory areas play a nontrivial role in categorical sound processing, as we report only null evidence in this study.

However, some of the differences between our relatively focal STS results and those of other auditory MVPA studies mentioned (relatively distributed over large portions of HG/STG/STS) could be due to the strict categorical nature of the utilized sound stimuli. While stimuli such as cats/guitars (Staeren et al. 2009) or syllables spoken by different voices (Formisano et al. 2008) are clearly differentiable and "identifiable" with near-perfect accuracy, they have not been shown to display all the hallmarks of CP as originally defined (Liberman et al. 1957), where discrimination is limited by identification [or, at least "partially" limited, according to revisions of the theory (Pisoni 1971; Zatorre 1983)]. It is therefore plausible that these multidimensional "cognitive categories" rely heavily on supraperceptual processes, which, in turn, use more widely distributed neural networks ("categorical cognition" as opposed to "categorical perception"). The behavioral data presented here (clear 3-category identification functions with aligned M-shaped discrimination functions) reflect the fixed, specific nature of over-learned musical categories and not a more general configuration of features. These 2 processing models—distributed versus hierarchical—may not be mutually exclusive, with the former putatively more applicable in situations where categories are less well-established or more like natural semantic categories (as demonstrated by Staeren et al. 2009), and the latter for more purely "perceptual" categorization.

The IPS results suggest the involvement of a dorsal stream of information processing. Unlike those in the STS, the IPS peaks fall well outside brain regions highlighted in the all sounds > silence contrast, suggesting that decoding in these parietal regions is performed on "supraauditory" information (and, separately, argues against ubiquitous use of activation masks as a first step in feature elimination methodologies, in accordance with findings from Jimura and Poldrack (2011)). The dorsal stream, originally postulated as the spatial processing system (Mishkin and Ungerleider 1982), is now more often considered to underlie the transformation and combination of information between sensory modalities (e.g., Culham and Kanwisher 2001) and into motor and execution codes (Hickok and Poeppel 2007; Rauschecker and Scott 2009). The IPS specifically has also been implicated in high-level sound transformations that require relative pitch processing (Foster and Zatorre 2010). The IPS peaks may thus be reflecting information that is still sensory, but no longer strictly auditory and on route to interfacing with the motor system. The motor theory of perception (Galantucci et al. 2006) is particularly relevant here, as our subjects all had extensive instrumental musical training. It follows that these individuals have formed strong associations between categories of musical sounds and the sets of movements required to make such sounds (Zatorre et al. 2007). A recent fMRI repetition suppression paradigm of expert pianists (Brown et al. 2013) demonstrated the involvement of the IPS in auditory-motor transformations for correct positioning of fingers, which, in combination with the presented results, implicates the IPS as a crucial "audio"motor interface (in addition to its more well-established role in visuomotor processing).

The location of the parietal peaks, particularly the left IPS, invites comparison with the more ventral area "spt." Spt is believed to form part of the auditory dorsal stream (Hickok and Poeppel 2007), is considered a "sensorimotor interface," and has been implicated in both speech production and perception (Hickok et al. 2008). Dorsally streaming music- versus speech-related information is likely destined for shared yet distinct frontal regions, with these results suggesting that spatially distinct processes emerge early. Furthermore, with the exception of few instruments (notably voice), music production relies heavily on the hands: this is notable as the parietal peaks observed here lie in the IPS, which is believed to underpin transformations between vision and limb and hand/grip movements (Cavina-Pratesi et al. 2010). However, in opposition to speech perception/production (with its strong one-to-one correspondence between sound/movement), a musical interval can be played using a variety of gestures requiring myriad sets of fingers/notes/instruments. It follows that frontal lobe perceptual decoding, such as the phonemic decoding reported by Lee et al. (2012), may require motor specificity beyond that provided for by abstract musical categories.

In summary, the STS and IPS results presented here, along with earlier fMRI data for musical interval categorization (Klein and Zatorre 2011) and multiple speech studies, indicate the likely presence of 2 streams of auditory information processing

for CP. The right STS, a critical component of the putative ventral stream, may underlie successful identification and recognition of simple musical categories, with the presented bilateral (but right lateralized) pattern of results complementing the speech phoneme CP literature (Wolmetz et al. 2011). In contrast, the dorsal IPS nodes may reflect a transformation stage between unimodal auditory and motoric information. These current analyses do not indicate the degree to which these streams remain separate entities or interact (and, if so, how). Finally, these results demonstrate the power of MVPA to enable mapping of highly automatic cognitive/perceptual processes, even in the absence of demanding behavioral tasks, which generally require larger working memory loads and complex control conditions, both of which may confound imaging results.

Funding

This work was supported by the Canadian Institutes of Health Research (CIHR) (grant nos MOP14995 and MOP11541) and by the Canada Fund for Innovation (grant no. 12246), and by infrastructure support from the Fonds de Recherche du Québec Nature et Technologies via the Centre for Research in Brain, Language and Music.

Notes

We thank the staff of the McConnell Brain Imaging Centre for help in acquiring imaging data, as well as various members of the PyMVPA mailing list for assistance with analysis scripts. *Conflict of Interest*: None declared.

References

Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B. 1999. Event-related fMRI of the auditory cortex. NeuroImage. 10:417–429.

Bornstein M. 1984. Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. Psychol Res. 46:207–222.

Brown RM, Chen JL, Hollinger A, Penhune VB. 2013. Repetition suppression in auditory–motor regions to pitch and temporal structure in music. J Cogn Neurosci. 25(2):313–328.

Burns EM, Ward WD. 1978. Categorical perception—phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. J Acoust Soc Am. 63:456.

Cavina-Pratesi C, Monaco S, Fattori P, Galletti C, McAdam TD, Quinlan DJ, Goodale MA, Culham JC. 2010. Functional magnetic resonance imaging reveals the neural substrates of arm transport and grip formation in reach-to-grasp actions in humans. J Neurosci. 30:10306–10323.

Culham JC, Kanwisher NG. 2001. Neuroimaging of cognitive functions in human parietal cortex. Curr Opin Neurobiol. 11:157–163.

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

Foster NEV, Zatorre RJ. 2010. A role for the intraparietal sulcus in transforming musical pitch information. Cereb Cortex. 20:1350–1359.

Galantucci B, Fowler CA, Turvey MT. 2006. The motor theory of speech perception reviewed. Psychon Bull Rev. 13:361–377.

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S. 2009. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics. 7:37–53.

Hary JM, Massaro DW. 1982. Categorical results do not imply categorical perception. Atten Percept Psychophys. 32:409–418.

Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. Nat Rev Neurosci. 7:523–534.

Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci. 8:686–691.

Hebb DO. 1949. The organization of behavior: a neuropsychological theory. 1st ed. New York: Wiley Publications in the Mental Health Sciences.

Hickok G, Okada K, Serences JT. 2008. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 101:2725–2732.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.

Hutchison ER, Blumstein SE, Myers EB. 2008. An event-related fMRI investigation of voice-onset time discrimination. NeuroImage. 40:342–352.

Jimura K, Poldrack RA. 2011. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia. 50:544–552.

Joanisse MF, Zevin JD, McCandliss BD. 2006. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. Cereb Cortex. 17:2084–2093.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 8:679–685.

Kilian-Hütten N, Valente G, Vroomen J, Formisano E. 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. J Neurosci. 31:1715–1720.

Klein ME, Zatorre RJ. 2011. A role for the right superior temporal sulcus in categorical perception of musical chords. Neuropsychologia. 49:878–887.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci. 103:3863.

Lee YS, Janata P, Frost C, Hanke M, Granger R. 2011. Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. NeuroImage. 57:293–300.

Lee YS, Turkeltaub P, Granger R, Raizada RDS. 2012. Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci. 32:3942–3948.

Leech R, Holt LL, Devlin JT, Dick F. 2009. Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci. 29:5234–5239.

Ley A, Vroomen J, Hausfeld L, Valente G, De Weerd P, Formisano E. 2012. Learning of new sound categories shapes neural response patterns in human auditory cortex. J Neurosci. 32:13273–13280.

Liberman AM, Harris KS, Hoffman HS, Griffith BC. 1957. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol Hum Percept Perform. 54:358–368.

Liebenthal E, Binder J, Spitzer S, Possing E, Medler D. 2005. Neural substrates of phonemic perception. Cereb Cortex. 15:1621–1631.

Mishkin M, Ungerleider LG. 1982. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav Brain Res. 6:57–77.

Mourao-Miranda J, Reynaud E, McGlone F, Calvert G, Brammer M. 2006. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. NeuroImage. 33:1055–1065.

Mur M, Bandettini PA, Kriegeskorte N. 2008. Revealing representational content with pattern-information fMRI—an introductory guide. Soc Cogn Affect Neurosci. 4:101–109.

Myers EB, Blumstein SE, Walsh E, Eliassen J. 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci. 20:895–903.

Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J. 2011. A comparison of volume-based and surface-based multi-voxel pattern analysis. NeuroImage. 56:593–600.

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage. 45:S199–S209.

Pisoni DB. 1975. Auditory short-term memory and vowel perception. Mem Cogn. 3:7–18.

Pisoni DB. 1971. On the nature of categorical perception of speech sounds [Doctoral thesis]. Michigan Univ Ann Arbor.

Raizada RDS, Poldrack RA. 2007. Selective amplification of stimulus differences during categorical processing of speech. Neuron. 56:726–740.

Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 12:718–724.
Repp BH. 1984. Categorical perception: issues, methods, findings. In: Lass NJ, editors. Speech and language: advances in basic research and practice. Vol. 10. New York: Academic Press. pp. 243–335.
Schouten B. 2003. The end of categorical perception as we know it. Speech Commun. 41:71–80.
Simon HJ, Studdert-Kennedy M. 1978. Selective anchoring and adaptation of phonetic and nonphonetic continua. J Acoust Soc Am. 64:1338–1357.
Staeren N, Renvall H, De Martino F, Goebel R, Formisano E. 2009. Sound categories are represented as distributed patterns in the human auditory cortex. Curr Biol. 19:498–502.
Wolmetz M, Poeppel D, Rapp B. 2011. What does the right hemisphere know about phoneme categories? J Cogn Neurosci. 23:552–569.
Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans AC. 2002. A general statistical analysis for fMRI data. NeuroImage. 15:1–15.
Zatorre RJ. 1983. Category-boundary effects and speeded sorting with a harmonic musical-interval continuum: evidence for dual processing. J Exp Psychol Hum Percept Perform. 9:739.
Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory–motor interactions in music perception and production. Nat Rev Neurosci. 8:547–558.
Zatorre RJ, Halpern AR. 1979. Identification, discrimination, and selective adaptation of simultaneous musical intervals. Percept Psychophys. 26:384–395.