Neuropsychologia 49 (2011) 878–887

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

A role for the right superior temporal sulcus in categorical perception of musical chords

Mike E. Klein[a,b,*], Robert J. Zatorre[a,b]

[a] Department of Neuropsychology, Montréal Neurological Institute, McGill University, Montréal, Québec H3A 2B4, Canada
[b] International Laboratory for Brain, Music and Sound Research, Montréal, Québec H3C 3J7, Canada

ARTICLE INFO

Article history:
Received 7 May 2010
Received in revised form
26 December 2010
Accepted 6 January 2011
Available online 12 January 2011

Keywords:
Categorical perception
fMRI
Superior temporal sulcus
Music perception
Hemispheric lateralization
Intraparietal sulcus

ABSTRACT

Categorical perception (CP) is a mechanism whereby non-identical stimuli that have the same underlying meaning become invariantly represented in the brain. Through behavioral identification and discrimination tasks, CP has been demonstrated to occur broadly across the auditory modality, including in perception of speech (e.g. phonemes) and music (e.g. chords) stimuli. Several functional imaging studies have linked CP of speech with activity in multiple regions of the left superior temporal sulcus (STS). As language processing is generally left-hemisphere dominant and, conversely, fine-grained spectral processing shows a right hemispheric bias, we hypothesized that CP of musical stimuli would be associated with right STS activity. Here, we used functional magnetic resonance imaging (fMRI) to test healthy, musically-trained volunteers as they (a) underwent a musical chord adaptation/habituation paradigm and (b) performed an active discrimination task on within- and between-category chord pairs, as well as an acoustically-matched, more continuously-perceived orthogonal sound set. As predicted, greater right STS activity was linked to categorical processing in both experimental paradigms. The results suggest that the left and right STS are functionally specialized and that the right STS may take on a key role in CP of spectrally complex sounds.

1. Introduction

Categorical perception (CP) is a phenomenon that occurs when signals that vary over a continuous physical scale are perceived as belonging to a small number of discrete groups. CP can be considered the converse of the default process of continuous perception, in which signals are perceived along a smooth continuum and are not lumped into categories. Two hallmarks of CP are (a) distinct categories with obvious boundaries that can be observed during labeling tasks and (b) a peak in discriminability between stimuli near a boundary, with complementary troughs far from boundaries (Liberman, Harris, Hoffman, & Griffith, 1957).

Formation and use of categories is thought to serve multiple related perceptual purposes. CP allows the perceptual system to quickly abstract complicated information – in the realm of speech, spectrally complex and rapidly changing acoustic signals – into "bins" for further downstream use. Put another way, the brain labels a speech sound as belonging to a certain phonemic category (e.g. /da/ or /ta/) and then can build words from these phonemes, as opposed to having to store and manipulate the much more complex auditory representation relayed from the brainstem.

Relatedly, this process provides a theoretically simple solution to the problem of acoustic variation between speech utterances. In the context of a particular phoneme, individual speech utterances vary considerably between speakers and, to a lesser extent, from act to act performed by the same speaker. Because no two voicing of a phoneme can be identical, though, linguistically, it makes sense to treat them as such, CP provides the means for a pre-conscious decision in favor of one of among a relatively small number of categories. CP was initially thought to be specific to speech processing (Mattingly, Liberman, Syrdal, & Halwes, 1971). Liberman et al. (1957) detailed the presence of non-linear features in subjects' identification and discrimination abilities, which show, respectively, how reliably a specific signal will be labeled as having membership in a certain category and the degree to which two neighboring signals along a certain portion of a continuous physical spectrum are differentiable. The theory that CP is a product of learning/familiarity was given traction by studies, beginning with Goto (1971), which showed that subjects perceived phonemes from their first language significantly more categorically than non-native speech contrasts (a well known example being the

* Corresponding author at: Montréal Neurological Institute, Cognitive Neuroscience Unit, 3801 University Street, Room 276, Montréal, Québec H3A 2B4, Canada. Tel.: +1 514 398 8519; fax: +1 514 398 1338.
E-mail addresses: michael.klein@mail.mcgill.ca, michaeleklein@gmail.com (M.E. Klein).

0028-3932/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2011.01.008

meaningful distinction between /l/ and /r/ in English, but not in Japanese).

Up until this point, the bulk of experiments looking at CP used stimuli that were exclusively linguistic and drew conclusions about the phenomenon that were specific to the speech domain. However, studies in the 1970s and '80s broadened the literature from the speech domain to the psychology of music, by looking at perception of musical intervals and chords with regard to category membership (with obvious examples being minor vs. major distinctions). Musically, the frequency ratio between a base note and its third defines the two-note interval (or chord if there are three or more notes) as being "minor" or "major." Burns and Ward (1978) showed categorical perception of intervals, as seen in identification and discrimination plots. Subjects showed troughs in discrimination ability in locations that correlated with interval category centres. While Burns and Ward's study focused on melodic (i.e. sequential) note presentation, Zatorre and Halpern (1979) showed that the same phenomenon occurred in harmonic (i.e. simultaneous) intervals. Additionally, the authors showed that CP of musical intervals was much more prevalent in trained musicians than in subjects who did not have significant musical training. Zatorre (1983) also addressed the putative existence of (and relationship between) "auditory" and "categorical" memory processing stages by selectively interfering with only the former. The experimental manipulation seemed to spare a "binary variable" that constituted the categorical memory.

In the past few years, numerous functional imaging studies have examined the neural correlates of CP in subjects performing linguistic tasks, with results generally implicating the left superior temporal sulcus (STS). The left and right STS each are large regions, spanning posteriorly-to-anteriorly from *y*-values of less than −40 (MNI space) to near the temporal pole, respectively, and encompassing large portions of Brodmann areas 21 (inferior STS/middle temporal gyrus (MTG)) and BA22 (superior STS/superior temporal gyrus (STG)) as well as smaller regions of BA38 and BA39 (temporal pole and angular gyrus, respectively). Here, we refer to STS regions most proximal to Heschl's gyrus as middle STS (mSTS) (*y*-values of approximately −25 to −5) and label the anterior STS (aSTS) and posterior STS (pSTS) accordingly. Liebenthal, Binder, Spitzer, Possing, and Medler (2005) compared blood–oxygen-level dependent (BOLD) responses in subjects who were discriminating phonemes in addition to a warped, non-phonemic continuum of comparably complex sounds that did not sound like English-language phonemes and could not be associated with pre-learned categories. Contrasting BOLD activity in the two conditions highlighted two peaks in the anterior/middle and posterior STS. An adaptation (i.e. short-interval habituation) paradigm (Joanisse, Zevin, & McCandliss, 2007) looked at BOLD activity contrasting conditions where oddball stimuli either did or did not cross a categorical boundary. The authors found greater BOLD activity for the between-category condition in the left STS, positioned between the peaks found by Liebenthal et al. The general correspondence of results between these two studies was notable, as the former utilized an active discrimination task and the latter a non-overt paradigm based upon a hypothesis of dishabituation/neural rebound, a design more common to ERP/MEG studies (Zevin & McCandliss, 2005) (Table 1).

Another recent study (Leech, Holt, Devlin, & Dick, 2009) showed that the left STS is likely involved more generally in CP and not merely limited to speech categorization. Subjects were trained on a video game, wherein certain fast-transforming complex sounds were indicative of an imminent game-play action. Study participants did not report these "acoustically-complex, artificial, and non-linguistic" stimuli as sounding speech-like. Because presentation of the sounds preceded (and predicted) specific upcoming events and required behavioral responses, acquisition of these new non-linguistic categories would be helpful with game performance. Participants who best learned these novel categories showed the greatest pre- to post-training change in BOLD response in the left pSTS, as observed during passive listening to these same stimuli. Thus, the authors concluded that CP correlated with left STS activity reflects auditory expertise in domains not limited to just language, and is susceptible to learning.

Table 1
Peak BOLD effects. All peaks are significant at the whole-brain level ($p < 0.05$, corrected), except for the second right STS peak.

Region	*x*	*y*	*z*	*t*	Contrast
Right STS	60	4	−8	5.66	Adapt1
Cerebellum	−44	−48	−40	4.75	Adapt2
Left occipital	−20	−92	30	4.62	Adapt1
Left IPS/inferior parietal lobule	−44	−56	50	4.60	Adapt2
Right STS*	44	−26	−4	3.39	Disc3

* Observation of statistical significance via anatomically-segmented right STS region-based analysis ($p < 0.001$ uncorrected).

The common thread between these imaging studies is the observation of significant BOLD activity in the left STS. The authors generally support the theory that the left STS is strategically positioned in the midst of the auditory "ventral stream" (Rauschecker & Tian, 2000), between more primary areas involved in the analysis of physical features of speech/other complex sounds and higher-order auditory cortex located in the left MTG and parts of the STS located more anteriorly. Liebenthal et al. suggest that phonemic recoding may be the earliest speech signal analysis that is lateralized to the left and that the STS is the actual "point of transition" – where sound starts to become speech. The implication here is that the category maps, themselves, reside within the left STS and that the observed BOLD signal, at least in part, reflects activity of the neurons that comprise the maps.

While the above imaging experiments of speech perception, as well as the study by Leech et al., make a very convincing case for a major role of the left STS in CP, they paint an incomplete picture of the phenomenon. The commonality between those studies is that they look for a BOLD response following categorization of rapidly-transforming, temporally-complex sounds. These findings cannot necessarily be taken as having highlighted the neural basis of *all* auditory categorical perception. Namely, they say little concerning acoustic stimuli lacking dynamic spectral variation, of which musical intervals are a prime example (and one that has already been shown to be perceived categorically). The idea of quickly- vs. slowly-varying auditory signals relates to theories of hemispheric specialization, in particular that the left hemisphere is tuned for perception of fast-changing signals (and thus is well-suited for speech) while the right hemisphere is tuned for higher spectral resolution. This theory – that left and right hemispheres, respectively, subserve these two parallel and complementary functions– was put forward by Zatorre, Belin, and Penhune (2002) as well as Poeppel (2003), whose argument was framed around putative "time integration windows" that are preferred by each of the two respective cortices. In this vein, numerous studies have shown that the right hemisphere is preferentially active for stimuli containing small variations in spectral energy (Boemio, Fromm, Braun, & Poeppel, 2005; Hyde, Peretz, & Zatorre, 2008; Schonwiesner, Rubsamen, & von Cramon, 2005; Zatorre & Belin, 2001). Thus, an imaging study that seeks to highlight brain areas involved in categorization of musical chords may implicate neural networks in the right temporal lobe responsible for a more inclusive concept of categorical perception. One can also make an alternate hypothesis that musical categories, such as minor and major, are mediated linguistically and thus rely heavily on the *left* STS for their percepts as categories. However, as any such linguistic labeling is predicated upon fine-tuned spectral analysis/extraction, it follows that some sort of

pre-categorical → categorical transformation must occur prior to associations with lexical elements, and that such a transformation is more likely to be primarily carried out by the right temporal lobe.

Here, we used fMRI to test the prediction that greater activity in or near the right STS of highly-trained musician subjects would be observed following presentation of stimuli comprised of chords from a larger number of musical categories. Such a finding would (a) suggest that there is something intrinsic to this brain region, bilaterally, that allows for transformations from nonspecific raw signal into pre-defined, cortically-based category and (b) lend credibility to theories that the relative strengths of the right and left temporal lobes are grounded in a differential sensitivity to slowly- and quickly-evolving sounds, respectively. While the specifics of any such findings (i.e. right STS activity associated with musical categories in musically-trained subjects) might not generalize to the population at large directly, observation of the predicted result would speak to a differential readiness/ability of the right vs. left STS to take on such a role in CP of spectrally-complex sounds. In addition to looking at differences between minor/major 2-category vs. single-category conditions, we created a set of acoustically-matched orthogonal sounds to serve as an additional experimental control (see Section 2.2). These orthogonal stimuli use absolute pitch cues and lack association with any learned musical categories. We predicted that, compared to the experimental triads, these orthogonal triads would be perceived in a less categorical manner, as measured by identification and discrimination scores. Finally, seeking converging evidence of functional localization, we employed two discrete experimental protocols: (1) an adaptation/oddball paradigm in which subjects were not asked to make explicit judgments related to category membership and (2) an ABX discrimination paradigm where overt, keyed responses were required.

2. Methods

2.1. Participants

We enrolled 35 participants in a behavioral pre-test. All subjects were right-handed, age 18–50, and did not claim to possess absolute pitch abilities. All were musicians with 4+ years of formal training on an instrument and claimed to be currently performing or practicing. All subjects gave informed consent to participate in this study, in accordance with procedures approved by the Research Ethics Committees of the McConnell Brain Imaging Centre and the Montreal Neurological Institute. Because we were interested in maximizing the likelihood of measuring the neural substrates of CP, following our pre-test, 19 of the 35 participants were excluded from further participation due to lack of sufficiently clear CP-like discrimination functions (see Section 2.3, for specifics of inclusion criteria). Additionally, two subjects who met these criteria chose not to participate in the imaging study and four more were eventually excluded due to failure to comply with instructions during scanner sessions. Thus, the imaging data are from a final cohort of 10 participants.

2.2. Stimuli

The behavioral pre-test involved two parallel sound sets, each containing 11 discrete triads (*see* Fig. 1). We generated an experimental and an orthogonal set, which shared one common triad. All of the triads were composed of three simultaneous 500 ms sine-wave tones (i.e. harmonic triads) that were generated using Audacity software and were derived from equally-tempered semitones (in which an octave lies 1200 cents above a starting frequency and each 100 cents signifies a 1/2 tone shift). Sound intensity was adjusted to each subject's comfort level and every triad was presented using a 50 ms linear ramp-up/down. The experimental sound set consisted of triads that ranged from true minor (middle note 300 cents above base note) to true major (middle note 400 cents above base note), in 10-cent increments (i.e. 300, 310, . . ., 390, 400). For all triads in the experimental set, the high note (musically, the 5th) was positioned 700 cents above the low/base note. Note that, for all triads in this set, the low and high notes were fixed at the same frequencies (G-natural at 392 Hz and the D-natural at 587.3 Hz) and only the middle note varied, from B-flat (300 cents above G-natural/466.2 Hz) to B-natural (400 cents above G-natural/496.8 Hz).

The orthogonal stimuli set was constructed in parallel to the experimental set. Our intent was to create a series of triads that did not span the categorical boundary between minor/major, while remaining as acoustically-related to the experimental stimuli as possible. As it is the ratio between the musical 1st and 3rd that determines the minor or major quality of the triad, we kept this ratio fixed at 350 cents (i.e. 1:∼1.22) for all triads in the orthogonal set. The 350-cent triad was chosen as it represents the midpoint on the minor/major continuum and does not clearly belong to either the former or latter category, as shown by identification ratings (*see* Section 3.1). As with the experimental set, the middle notes of these 11 triads ranged from B-flat (466.2 Hz) to B-natural (496.8 Hz). However, in order to keep consistent a 350-cent interval between low and middle tones, it was necessary to vary the frequency of the low tone from triad to triad. This is in direct contrast to the experimental set, where the frequency of the low tone was always fixed at 392 Hz. As the middle tone varied from 466.2 Hz to 496.8 Hz, the low tone varied from 380.8 Hz (between G-flat and G-natural) to 405.8 Hz (between G-natural and G-sharp). Likewise, the high tone (5th), which was always positioned 700 cents above the base tone, varied in the orthogonal sound set, from 570.6 Hz to 608 Hz. While all three tones vary in frequency from triad to triad within this sound set, the frequency ratio between the three tones is held constant. As a result, these orthogonal triads, unlike the experimental triads, do not differ from one another along the minor/major dimension, but instead differ on the basis of their absolute frequency. In order to keep a consistent naming scheme, individual triads from both sound sets will be referred to on a scale from 0 to 100 cents, which represents the distance above the low anchor triad from either set. However, it is important to note that this distance refers either to pitch-variation of the middle note (experimental triads) or of all three notes (orthogonal triads), depending on the sound set.

For the pre-test, sounds presentation and data collection were conducted using Max/MSP software (Cycling '74 Inc., http://www.cycling74.com) and Sennheiser HD280 Pro headphones. In-scanner tasks were administered with Presentation software (Neurobehavioral Systems, http://www.neurobs.com) and MR-Confon Peltor Optimex magnetic resonance-compatible headphones.

2.3. Pre-test tasks

Subjects performed identification and discrimination tasks of both sound sets as part of a behavioral pre-test, conducted inside a sound booth. Prior to performing the identification task, subjects listened to repeating and alternating presentations of the two endpoint-triads. These endpoint (a.k.a "anchor") triads were the true minor and major triads for experimental set identification, or the two analogous triads if the subjects were performing the task on the orthogonal set. The order of presentation was counter-balanced so that half of the subjects first heard the experimental triads and half the orthogonal triads. During the fMRI portion of the experiment, subjects performed a similar discrimination task, and also underwent an adaptation/oddball paradigm.

2.3.1. Identification

After familiarization with the anchor triads, subjects were presented with trials that contained a single triad that could come randomly from anywhere in the set. They were then asked to rate that triad on a scale of 1–6: (1) subject is sure triad is closer to low anchor, (2) subject thinks the triad is closer to the low anchor, but is not positive, (3) subject is fairly unsure, but if pressed to guess, would place the triad closer to the low anchor, (with (4), (5), and (6) the complementary choices for the high anchor). Subjects had unlimited time to make their selections and, following each choice, were given a 2-s silent period prior to presentation of the next triad. Each of the 11 possible triads from a given set was presented 12 times in a pseudo-random order.

2.3.2. Discrimination

Following the identification task, subjects performed an ABX discrimination task on the same triad set. For each trial in this task, subjects heard three triads, each separated from the next by 500 ms of silence. In this task, "A" could be any one of the 11 possible triads; "B" would be a triad, 2 steps away from "A" (either up or down) on the continuum; and "X" would be a repetition of either "A" or "B." An example from the experimental set would be presentation of a 30-cent triad ("A"), followed by a 10-cent triad ("B") and another 10-cent triad ("X"). After each presentation, subjects were asked to click "1" if they believed X matched A or to click "2" if they believed X matched B. In the above example, a response of "2" is correct. Following each response, there was a 2-s silent period prior to the next trial. Subjects were not provided with correct/incorrect feedback. There were an even number of X = A and X = B trials as well as an even number of trials where A > B or B > A, in terms of frequency/position in the stimuli set. Each of the 9 possible complementary triad pairs from a given set was presented 12 times in a pseudo-random order. Each subject performed two identification tasks and two discrimination tasks for each triad set.

2.3.3. Inclusion criteria

In order to qualify for the fMRI portion of the study, a subject had to show a (a) discrimination performance peak for the minor/major triad set that was 25%+ better than the average of their within-category endpoints and (b) 50/70 cent discrimination rate that was not significantly lower than their peak performance (whether that peak was found at 40/60, 60/80, etc.). The second criterion was included because, as the large majority of subjects' performance peaks were found at 50/70, this pair was selected to become the between-category condition used in-scanner. 16 of the initial 35 subjects met both of the above criteria. Of these 16, two subjects

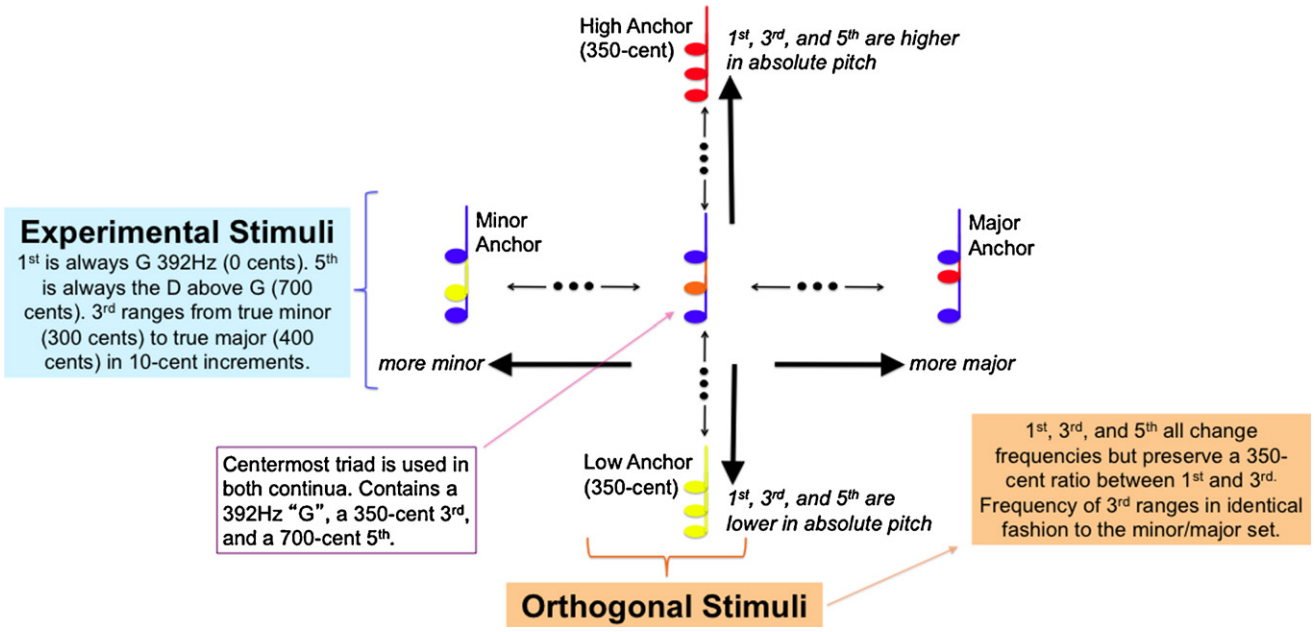

Fig. 1. Two triad sets. Experimental stimuli are represented horizontally. Moving from left to right, the triads become progressively more major (from 300 cents to 400 cents, in 10 cent increments). This was done by varying the frequency of the middle note, while the frequencies of the bottom and top notes remain constant. The mid-most triad (350 cents) is shared with the 2nd stimuli set. Orthogonal stimuli are represented vertically. Moving from bottom to top, triads become progressively higher in frequency; this is true for all three notes of the triads (as opposed to only the middle note, as in the experimental set). Because the frequency ratio for the three notes of each triad is held constant, these orthogonal triads do not differ from one-another in the minor/major dimension.

declined to participate in the fMRI section. Data from four further subjects who were scanned were excluded from the imaging analyses due to subjects' failure to comply with in-scanner instructions (i.e. required behavioral responses that were absent or inconsistent). Thus, our imaging data come from a final cohort of 10.

2.4. In-scanner procedure

Each participant underwent an anatomical scan and two functional imaging runs. Each run consisted of eight blocks of triads: four each for the adaptation (ADPT) and discrimination (DISC) protocols (*see below for details of each protocol*). For each protocol, two blocks contained triads from only the experimental sound set (EXP) and two contained triads from only the orthogonal sound set (ORT). Run "A" was ordered DISCexp > ADPTexp > DISCort > ADPTort > ADPTexp > DISCexp > ADPTort > DISCort. Run "B" was ordered ADPTort > DISCort > ADPTexp > DISCexp > DISCort > ADPTort > DISCexp > ADPTexp. We used a counterbalanced design so that half the subjects underwent run "A" then "B" and half "B" then "A."

Blocks were separated from one another by two silent trials where no sounds were played, followed by a "cue" trial, where subjects were told which protocol to follow in the upcoming block. Each run contained a total of 166 10-s trials: 76 from the adaptation experiment (19 per block × 4 blocks); 64 from the discrimination experimental (16 per block × 4 blocks); 18 silent; and 8 cue. Trials using the middle-frequency triad pair of each stimuli set (50/70) were presented twice as often as those from either the low- or high-frequency pairs (0/20 and 80/100, respectively). Triad pairs from the pre-test, other than 0/20, 50/70, and 80/100, were not used for the imaging experiment as we sought to contrast the most boundary-spanning (50/70) and least boundary-spanning (0/20 and 80/100) conditions.

2.4.1. Adaptation paradigm

A single ADPT block contained 19 trials and used triads from only one of the two sound sets. Each trial (*see* Fig. 2) was one of two types. Repeating type (REP) was presented as A–A–A–A–A, where the same triad was presented 5X, with 500 ms silent gaps between sounds. Changing ("oddball") type (CHG) was presented as A–A–A–A–B, where one triad was presented four times followed by a second triad that was presented once. As with REP, there were 500 ms silent gaps between sounds. In any given trial, A and B were complementary triads from a pair (ex: if A = 70, B = 50). REP and CHG trials were presented with equal frequency and in a random order. Of each block's 19 trials, 4 contained triads from the 0/20 pair, 4 from the 80/100 pair, and 8 from the 50/70 pair.

The remaining 3 trials per block were employed for a separate purpose. The adaptation paradigm, itself, required no overt responses from subjects. However, in order to ensure that they remained alert and were attentive to the sounds, we had subjects undergo each ADPT block under the guise of an overt "loudness" task. Subjects were requested to make a key-press if a trial's final triad was heard as being quieter than the preceding 4. Thus, in addition to the 16 trials mentioned above (in which all 5 triads were of equal intensity), 3 trials contained final triads that were of 1/4 the amplitude of the first 4. While behavioral responses were checked for compliance with the loudness task, we did not analyze fMRI data collected from these trials. For this paradigm, subjects were not specifically instructed to listen for whether the final triad was of different pitch quality than the first 4.

2.4.2. Discrimination paradigm

The in-scanner ABX discrimination task (*see* Fig. 2) was similar to that described for the pre-test. As mentioned above, one difference was that subjects heard and discriminated only the 0/20, 50/70, and 80/100 pairs from each set. A second difference was that, where the pre-test allowed for a response period of unlimited duration, the fMRI task required a response before the onset of BOLD volume acquisition. This period of relative quiet ranged between 3.8 and 4.8 s in duration and, following presentation of triad X, subjects were asked to respond as "quickly as possible" by pressing one of two buttons on an MRI-compatible controller, depending on whether they heard X as matching A (choice 1) or X as matching B (choice 2). Of each DISC block's 16 trials, 4 contained triads from the 0/20 pair, 4 from the 80/100 pair, and 8 from the 50/70 pair.

2.4.3. Image collection and analysis

Images were acquired on a 1.5 T Siemens Sonata scanner. A high-resolution (voxel = 1 mm^3) T1-weighted scan was obtained for anatomical localization. Dur-

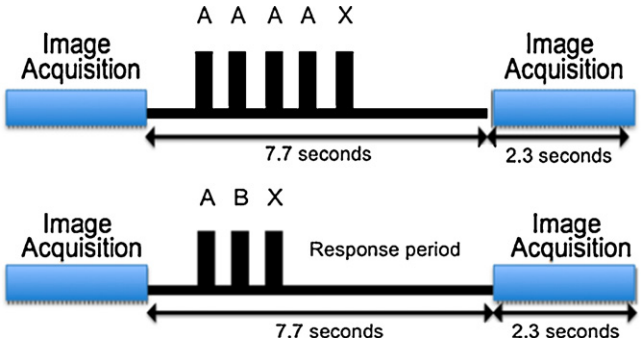

Fig. 2. Single trials from adaptation (top) and discrimination (bottom) experiments. Each 10-s trial was comprised of 2.3 s for image acquisition following 7.7 s for sound presentation and behavioral responses. Longer durations of stimuli during adaptation trials were offset by the lack of a need for a response period. Trials occurred in blocks containing only those of same type (e.g. discrimination of experimental triads, adaptation using orthogonal triads, etc.).

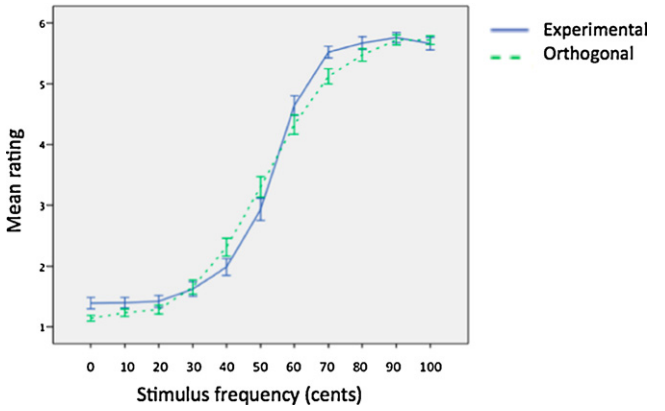

Fig. 3. Identification performance. Mean ratings are from a scale of 1 to 6, as presented triads are perceived as resembling the low to high anchor triads of each sound set, respectively. *X*-axis values represent cents above a minor triad as determined by the middle note (experimental) or cents of each of the three notes above the lowest-frequency triad (orthogonal). Error bars show 95% confidence intervals.

ing two functional runs, one whole-head frame of 36 contiguous T2*-weighted images was acquired in ascending, interleaved fashion (TR = 10 s, 64X64 matrix, voxel size = 8 mm^3 (2 mm × 2 mm × 2 mm)). We used a sparse-sampling procedure (Belin, Zatorre, Hoge, Evans, & Pike, 1999): tasks were performed between the 2.3-s acquisitions to prevent scanner noise from interfering with the auditory stimuli. Sound samples were presented near the beginning of the 7.7-s non-acquisition window. Relative timings between scan acquisitions and tasks were systematically varied or "jittered" by up to ±500 ms to maximize the likelihood of obtaining the peak of the hemodynamic response for each task.

All BOLD images were realigned with the third frame of the first run to correct for motion artifacts. To increase the signal-to-noise ratio, images were smoothed with a 6-mm full-width at half-maximum (FWHM) isotropic Gaussian kernel. Image analyses were conducted utilizing the general linear model via fMRISTAT as outlined by Worsley et al. (2002). Motion-correction parameters were used as covariates in fMRISTAT to further account for motion artifacts in the imaging results. In-house software was used to non-linearly transform each subject's images into standardized space using the MNI/ICBM 152 template, prior to conducting the group analyses (Collins, Neelin, Peters, & Evans, 1994; Mazziotta et al., 2001). Peaks from the full-brain analysis were considered significant if above a threshold of $t > 4.57$, which was corrected for multiple comparisons ($p = 0.05$). The program *stat_summary* assessed the threshold for significance by selecting the minimum among the values given by a Bonferroni correction, random field theory, and the discrete local maximum (Worsley, 2005). We report peaks of neural activity if their voxel or cluster *p*-values are <0.05. For a portion of our fMRI analysis, we pre-defined a region spanning the right STS. Within this predicted area we report any peaks that were significant above an uncorrected threshold of $p = 0.001$. We performed the location-based analysis because our primary prediction, based upon multiple streams of prior research, focused on this specific right temporal region. As the speech/language literature has highlighted activity peaks over multiple areas in the left STS, we delineated the entire right STS, spanning from the most posterior to most anterior regions of the sulcus. The STS was manually segmented based upon anatomical landmarks: (a) from posterior to anterior for as long as the sulcus was clearly visible (near angular gyrus ($Y = -46$) to near temporal pole ($Y = 6$)); (b) dorsal/ventral from the most central/superficial point of the STG to that of the MTG; and (c) encompassing the entire sulcus (superficial to white matter).

3. Results

3.1. Behavioral results

A two-way ANOVA performed on identification ratings from all 35 subjects during the pre-test (*see* Fig. 3) showed a significant interaction effect between sound set and stimulus frequency ($F = 8.848$, $p < 0.001$). Tukey's honestly significant difference (HSD) *post hoc* tests showed a significant difference in mean rating of the experimental vs. orthogonal triads at frequencies of 40, 50, 60, and 70 cents ($p < 0.05$ for all), but not at the left-most (0, 10, 20, and 30 cents) and right-most (80, 90, and 100 cents) ends of the functions.

For discrimination performance from all 35 subjects during the pre-test (*see* Fig. 4), a two-way ANOVA showed a significant interaction effect between sound set and stimulus frequency ($F = 5.154$, $p < 0.001$). Tukey's honestly significant difference (HSD) *post hoc* tests showed a significant difference in discrimination performance of the experimental vs. orthogonal triads at frequency pairs of 0/20, 10/30, 20/40, 30/50, 70/90, and 80/100 cents ($p < 0.05$ for all), but not at the centre of the functions (40/60, 50/70, and 60/80 cents). To confirm that the experimental triads were being perceived in a categorical-like manner, further Tukey's HSD *post hoc* tests showed that peak discrimination performance of this sound set (50/70 comparison, 84% accuracy) was significantly better than at the 0/20 (56% accuracy, $p < 0.05$) and 80/100 (56% accuracy, $p < 0.05$) endpoints. Performance at the two endpoints did not differ significantly from one another. The orthogonal triads were discriminated with a peak accuracy of 85% (50/70) and endpoint accuracies of 69% (0/20 and 80/100).

In-scanner discrimination data (*see* Fig. 4) are from the final cohort of 10 subjects. A two-way ANOVA showed a significant interaction effect between sound set and stimulus frequency ($F = 29.385$, $p < 0.001$). Tukey's honestly significant difference (HSD) *post hoc* tests showed a significant difference in discrimination performance of the experimental vs. orthogonal triads at frequency pairs of 0/20 as well as 80/100 cents ($p < 0.05$ for both), but not at 50/70 cents. Once again, to confirm that the experimental triads were being perceived in a categorical-like manner, Tukey's HSD *post hoc* tests showed that peak discrimination performance of this sound set (50/70 comparison, 91% accuracy) was significantly higher than at the 0/20 (48% accuracy, $p < 0.05$) and 80/100 (66% accuracy, $p < 0.05$) endpoints. Unlike in the pre-test, in-scanner performance at the 0/20 endpoint was significantly below performance at the 80/100 endpoint ($p < 0.05$). These 10 subjects did not show a similar performance pattern during the pre-test, discriminating experimental triads at 91% (50/70), 59% (0/20), and 51% (80/100). The orthogonal triads were discriminated in-scanner with a peak accuracy of 93% (50/70) and endpoint accuracies of 85% and 90% (0/20 and 80/100, respectively).

3.2. fMRI results

We analyzed a total of six contrasts: three for each experimental paradigm. The contrasts were chosen to employ as much parallelism as possible between the two paradigms. However, it is important to note that certain elements do not exactly translate across the experiments. The adaptation paradigm primarily looked at oddball-related habituation effects across the two sound sets. The discrimination paradigm was more closely tied to an active behavior. Relatedly, because the observed in-scanner discrimination of major-category (but not minor-category) triads was better than what was expected based upon pre-test behavioral results, we chose to focus on the minor- and between-category discrimination pairs for this second paradigm. This was done with the intent of maximizing the chances of observing BOLD activity related to CP, which was the primary goal of the study (*also see* Section 4). Separately, in order to complement the right STS sub-analysis described in the Section 2, a similar region-specific analysis was conducted in the left STS, although we did not predict activity in the latter area. No significant peaks were observed using the same threshold criteria ($p < 0.001$ uncorrected).

3.2.1. Adaptation paradigm

The first contrast from our adaptation paradigm (*Adapt1*) compared BOLD activity from all oddball experimental trials with repeating experimental trials, after subtraction of the analogous orthogonal volumes: $[[EXP_{CHG} - EXP_{REP}] - [ORT_{CHG} - ORT_{REP}]]$. A significant peak was found in the right aSTS ($x = 60$, $y = 4$, $z = -8$; $t = 5.66$, *see* Fig. 5).

The second contrast (*Adapt2*) looked at BOLD activity following EXP_{CHG} stimuli that crossed the minor/major categorical boundary

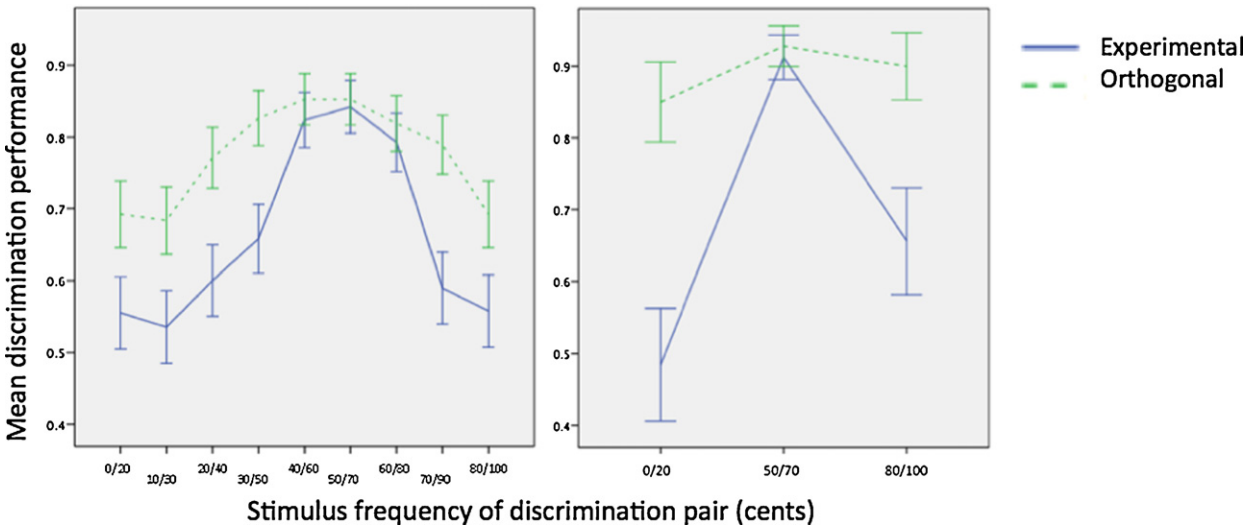

Fig. 4. Discrimination performance from pre-test (left) and scanner session (right). Discrimination scores are out of 1 (100% accuracy). *X*-axis shows position in frequency space of triads within a given sound set. *X*-axis values represent cents above a minor triad as determined by the middle note (experimental) or cents of each of the three notes above the lowest-frequency triad (orthogonal). Error bars show 95% confidence intervals.

(i.e. the 50/70 pair, between-category: "BW") minus the analogous ORT_{CHG} trials: $[EXP_{CHG\text{-}50/70} - ORT_{CHG\text{-}50/70}]$. This contrast showed a peak that was significant at the whole-brain level in the left intra-parietal sulcus (IPS)/inferior parietal lobule ($x = -44$, $y = -56$, $z = 50$; $t = 4.60$). A sub-threshold peak in a similar *right*-hemispheric region was also observed ($x = 52$, $y = -46$, $z = 44$; $t = 3.34$).

The third adaptation paradigm contrast (*Adapt3*) looked at between- and within-category experimental oddball conditions:

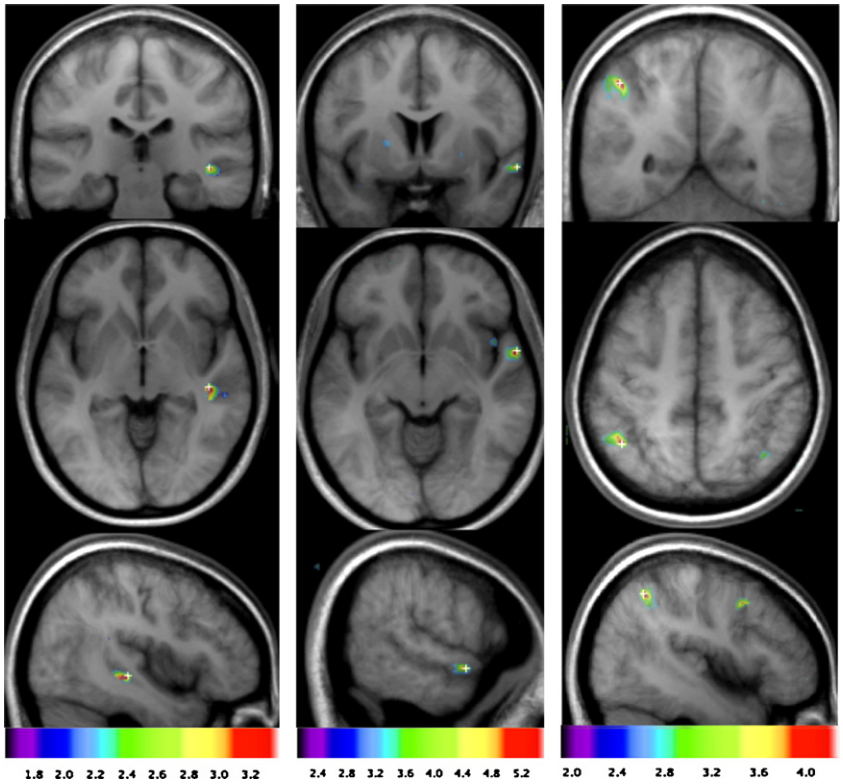

Fig. 5. BOLD peaks. Contrast Disc3 (left) from our discrimination protocol (right STS sub-analysis) compares BOLD activity following discrimination of between-category experimental triads minus discrimination of within-category (minor) triads ($EXP_{50/70} - EXP_{0/20}$). This comparison is meant to isolate activity arising following presentation of multiple categories (i.e. minor and major) vs. a single category. A peak ($t = 3.39$, right STS sub-analysis) was observed in the right middle/posterior STS ($x = 44$, $y = -26$, $z = -4$). Contrast Adapt1 (centre) from our adaptation protocol compared BOLD activity following presentation of all experimental oddball trials (EXP-CHG) with non-oddball trials (EXP-REP), after subtraction of similar volumes from the orthogonal sound set ((ORT-CHG) – (ORT-REP)). This comparison is meant to isolate a rebound from adaptation, but only when such a rebound taps into neural substrates that contain category information. A peak ($t = 5.66$) was observed in the right aSTS ($x = 60$, $y = 4$, $z = -8$). Contrast Adapt2 (right) compared BOLD activity following presentation of boundary-spanning experimental oddball trials ($EXP\text{-}CHG_{50/70}$) with the analogous orthogonal volumes ($ORT\text{-}CHG_{50/70}$). This comparison is also meant to isolate a rebound from adaptation, but only when associated with a second and distinct musical category. A peak ($t = 4.60$) was observed in the left IPS/inferior parietal lobule ($x = -44$, $y = -56$, $z = 50$). All anatomical underlays are from the nonlinearly-registered average of the 10 subjects tested.

[$EXP_{CHG\text{-}50/70} - EXP_{CHG\text{-}0/20,\ 80/100}$]. No significant peaks were observed. This contrast was primarily conducted for congruence with Disc3 (below), a main contrast from the discrimination experiment.

3.2.2. Discrimination paradigm

Disc1, a discrimination paradigm contrast that was employed to parallel Adapt1, did not show any significant peaks. This contrast compared activity following discrimination of all experimental triads with that of activity following discrimination of all orthogonal triads: [EXP − ORT].

Disc2, which was constructed to parallel Adapt2, did not yield any significant peaks. While Adapt2 compared 50/70 oddballs across the two sound sets, Disc2 simply compared BOLD activity following discrimination of the experimental and orthogonal 50/70 pairs: [$EXP_{50/70} - ORT_{50/70}$].

The primary discrimination contrast, *Disc3*, compared between-category and within-category (minor) conditions ([$EXP_{50/70} - EXP_{0/20}$]) and showed a significant peak within the right middle/posterior STS ($x = 44$, $y = -26$, $z = -4$; $t = 3.39$, significant via right STS sub-analysis, *see* Fig. 5). We also note the presence of a large, though sub-threshold, peak nearby in the right STG ($x = 50$, $y = -26$, $z = 14$; $t = 4.31$).

4. Discussion

4.1. Behavioral performance

Overall behavioral performance of subjects, observed both during the pre-test and in the scanner, yielded data that show all the signs of classic CP functions. This categorical effect was much stronger for the experimental than the orthogonal triads, suggesting that the latter successfully functioned as an appropriate control. Identification functions for the experimental and orthogonal triad sets showed a significant interaction effect, with subsequent *post hoc* tests indicating that the differences came primarily from the centre of the plots. Mean identification ratings at 40 and 50 cents were significantly closer to the low anchor for the experimental vs. the orthogonal triads. The opposite was true at 60 and 70 cents, suggesting the experimental function showed more of the "quick transition" that is hallmark of a boundary region between categories. Subjects were required to respond in terms of a triad's "closeness" to one anchor vs. the other based on a rating scale. The orthogonal identification ratings, while less categorical than the experimental, did not take the form of a perfectly linear function as triads increased in frequency. We believe that this finding reflects anchoring effects, which likely are due either to a response bias (i.e. subjects' tendency not to respond as "unsure") and/or perceptual factors involving auditory memory or volatility of the mental representations of the anchor sounds (Acker, Pastore, & Hall, 1995). Regardless of any such effects, CP was demonstrably stronger in the experimental identification function, thus providing evidence that we were using a proper orthogonal control.

Discrimination data confirmed the findings from identification. Although we used $n = 10$ for our in-scanner task, data were reported from all 35 pre-test subjects in order to show that observed CP effects were general to our entire sample of musicians. In order to best distinguish the neural substrates of CP, the 10 best-performing subjects were scanned and analyzed, and in-scanner discrimination data from these subjects were also reported. Both data sets showed a peaked, CP-like function for the experimental sounds and less CP-like functions for the orthogonal sounds: a result that echoed our identification findings. The experimental pre-test function showed within-category performances slightly above chance (56% accuracy), with the performance peak at the 50/70-cent comparison (84% accuracy). This peak accuracy was almost identical to that from the orthogonal stimulus function (85%), which also occurred at the 50/70-cent comparison. As with identification, the orthogonal plot does not appear as a purely continuous perceptual function, which in this case would be a flat line. Instead, it contains endpoint troughs, which likely are due to the same anchoring effects spoken about above. It is of note that the discrimination peaks of the two sound sets (91% and 93% for experimental and orthogonal, respectively) are almost identical, suggesting that any BOLD differences observed when contrasting these two conditions are likely not a performance effect of the behavioral task.

The in-scanner behavioral functions follow the same general pattern as those from the pre-test, with certain differences. First, the three orthogonal triad pairs were discriminated with more consistent (and higher) accuracy than during the pre-test, which is likely an effect of practice/exposure. This same flattening of the function was not observed for the experimental stimuli, which appear to have been perceived even *more* categorically during the scanner session. Both of these points speak to a likely dominance of category-based processing: in other words, task-based short-term practice effects could not compete with over-learned CP, which has been acquired throughout participants' entire lifetimes. While some degree of the performance increase from pre-test to scanner may be due to subjects being tested on only 6 triad pairs in the latter sessions (a subset of the 18 pre-test pairs), this alone cannot fully explain the differential changes observed between the experimental vs. orthogonal sound sets. A final difference was a performance imbalance between discrimination of triads taken from the minor and major ends of the continuum (48% and 66%, respectively), which had been discriminated at essentially identical rates by the $n = 35$ population at pre-test (56% for both). As stated in the results section, this was not due to an issue with the $n = 10$ subsample, which actually showed the reverse performance trend during the pre-test (59% for minor, 51% for major). Because this last finding was both unexpected and difficult to explain, we felt it appropriate to use only minor-category fMRI trials for discrimination protocol contrasts, as our main intent was to measure the neural correlates of CP by comparing clear within- vs. between-category conditions. Despite these small differences between pre-test and scanner session data, we feel, as with identification, that the discrimination results as a whole confirm that CP effects for the experimental triads were demonstrably stronger, providing additional evidence that the orthogonal triads functioned as a proper control for use in imaging contrasts.

4.2. Right temporal activity

In the present study, our goal was to test whether regions in the right STS are preferentially active for stimuli containing more musical category information. As predicted, the right STS showed such BOLD responses, which were present across both of our experimental paradigms.

The first adaptation paradigm contrast (Adapt1) elicited a large BOLD peak in the aSTS and Disc3 showed a significant peak in the middle/posterior right STS (*see* Fig. 5; latter peak assessed via the location-based analysis). Taken together, the peaks elicited across both experimental paradigms suggest that observed activity in this right temporal region is a real effect. The large anterior peak is located in a position that is roughly symmetrical to the more anterior of two *left* STS peaks from Liebenthal et al. (2005) ($x = -60$, $y = -8$, $z = -3$). Liebenthal et al. compared BOLD activity following discrimination judgments of phonemes against a warped, acoustically-matched set of non-speech-like sounds. Like the contrast used by Liebenthal, et al., Adapt1 compared both within- and between-category experimental stimuli against stimuli from an orthogonal control condition. Likewise, the more posterior right

STS peak shows general correspondence with those of Liebenthal et al. ($x = -56, y = -31, z = 3$), as well as Joanisse et al. (2007) ($x = -66, y = -26, z = 7$ and $x = -64, y = -25, z = -7$) (n.b. Peak locations listed for Liebenthal et al. and Joanisse et al. are in Talairach coordinates, though discrepancy from MNI coordinates are minor).

Liebenthal et al. have proposed that phonemic recoding may be the earliest kind of speech processing that is truly lateralized to the left temporal lobe. Liebenthal et al. and Joanisse et al.'s phonemic CP results provide evidence that the middle/anterior left superior temporal region is where this recoding takes place, a conclusion that has been supported by other imaging studies of phonemic perception (Hutchison, Blumstein, & Myers, 2008; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007). The left pSTS training effect observed by Leech et al. (2009) ($x = -54, y = -37, z = -1$), which dealt exclusively with temporally-complex *non-speech* sounds, indicates that this left hemispheric specialization may be more general in nature. Looking to the more anterior STS, studies have implicated both the left and bilateral aSTS in higher-order speech processes that contribute to phrase- or sentence-level comprehension (e.g. phonetic, semantic, syntactic) (Davis & Johnsrude, 2003; Humphries, Willard, Buchsbaum, & Hickok, 2001; Narain et al., 2003). These ultra-phonemic processes, which lie farther down the putative "ventral stream," are also likely making use of certain types of speech category information (e.g. noun vs. verb). Our results, which contrast (a) category-containing stimuli against stimuli perceived significantly less categorically, as well as (b) between-category stimuli against within-category stimuli, show analogous right hemispheric activity to the left temporal peaks of the speech literature. As our control stimuli were selected to be well-matched for spectral complexity, we believe that the observed right STS BOLD signals are truly reflective of pitch-based *categorical* processing, which extends prior findings that show a more general right auditory cortex bias for fine-grained spectral processing (Hyde et al., 2008; Zatorre & Belin, 2001).

The ventral and dorsal streams make up the individual components of the "two-stream hypothesis" that was originally put forward by Mishkin and Ungerleider (1982). The theory was initially formulated with respect to the visual system and argued for a ventral "what" pathway that handles identification of objects, as well as a dorsal "where" pathway that deals with objects' locations in space. As part of the hypothesis, the ventral and dorsal streams are thought to be primarily-mediated by the temporal and parietal lobes, respectively, with more abstract representations of objects existing further from primary sensory areas. This theory has been extended to the auditory domain (Rauschecker & Tian, 2000), with more recent two-pathway models involving abstraction beyond simple what vs. where components to encompass sensory-motor aspects of processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Sensitivity to features of auditory objects has been linked to antero-ventral areas of right temporal cortex (i.e. ventral stream) (Zatorre, Bouffard, & Belin, 2004) and, generally, the category-centric exploration of phoneme identification/discrimination and resultant left STS findings fall under the broad heading of "ventral stream."

The right STS activity observed in our discrimination paradigm may to some degree represent higher neural processing demands following exposure to a greater number of categories, as it was observed following discrimination of boundary-spanning triad pairs (2 categories), after contrasting with within-category minor pairs (1 category). However, employing an active discrimination task raises the possibility that the observed STS activity may reflect task-related *use* of any categorical information, as opposed to "pure" category percepts, themselves. This issue was addressed via our adaptation paradigm, where subjects were not instructed to judge sounds for category/pitch quality. The Adapt1 contrast, which yielded the large right aSTS peak, grouped together 1- *and* 2-category experimental triad pairs, which were then compared with all orthogonal pairs. We note that the two paradigms each have different degrees of memory load and attentional requirements. In the discrimination task, subjects paid more explicit attention to the experimentally-relevant features of the triads, though they were not instructed to listen specifically for the "quality" of sounds (merely to compare/choose among them). While the orthogonal AAAAX task (related to loudness) was easier and required different and likely fewer attentional processes, it ensured that subjects' focus was still on the auditory modality. Regarding memory load, performance of both tasks likely utilized working memory as well as echoic memory. If there were no musical categories, the ABX task could be performed via echoic memory, without any need to remember A (i.e. B either matches X or does not match X). For within-category comparisons, the most successful strategy likely involves a shift in focus toward sensory memory as soon as B is heard (with the opposite being true of between-category comparisons). While the discrimination task is the more demanding of the two, both tasks, in a sense, really only require one triad to be "kept in mind" prior to presentation of X, with such tracking likely involving a blend of memory-types.

It is of note that the Adapt1 adaptation contrast compared oddball and repeating trials, after subtraction of the orthogonal from experimental volumes. Based on the behavioral data, the orthogonal triad pairs were even more discriminable than the experimental pairs, so it is improbable that participants simply could not perceive the orthogonal oddball ("change") trials as sounding different from repeating trials. It may be the case that observed anterior activity follows equally from single- and multi-category stimuli, but is less related to non-categorizable stimuli. This hypothesis could explain the lack of such an anterior peak in the Disc3 discrimination contrast, which did not use a control from the less-categorically-perceived sound set. It is of note that Liebenthal et al.'s results, which include both middle/posterior and middle/anterior STS peaks, were also from a contrast of both 1- and 2-category experimental stimuli against category-free orthogonal stimuli. A second possibility is that the aSTS may be involved in combining category information relayed from the middle/posterior STS with pre-categorical auditory information, thus making it most sensitive to changes that are specific to already-binned objects.

We believe that the sum of these results provide evidence for a role of the right STS in perception of spectrally-complex auditory categories. As mentioned in Section 1, while these specific results do not generalize beyond subjects with musical training who show strong behavioral CP traits, they do suggest a predisposition of the right STS to take on a larger role than the left. We feel that, most likely, the functional results presented here arise via a combination of a specialization of right temporal lobe, present in a large proportion of the general population, and a specific sort of training/learning that capitalizes on this hemispheric bias. Questions remain, including the degree to which temporal regions respond to single vs. multiple categories, as well as the degree to which category representations are distinct or overlap with one another. Taken as a whole, the body of literature strongly suggests that bilateral ventral streams, and more specifically the left and right STS, underlie auditory categorical perception. However, observation of auditory category-related BOLD activity seems to be a subtle phenomenon, with some studies yielding significant peaks only via a large number of participants and a subset of contrasts (e.g. Liebenthal et al. scanned 25 subjects and observed significant STS activity for a phonemic vs. non-phonemic contrast, but not for a between-category vs. within-category contrast). Additionally, many auditory CP studies employ temporal lobe-ROI analyses in addition to looking at whole-brain activity (Hutchison et al., 2008; Joanisse et al., 2007). Likewise, while we observed one very clear BOLD peak in the right aSTS, the more posterior right STS peak was detected using a

relatively liberal threshold for significance. However, our STS peaks show general right/left location correspondence to those from the speech literature. It may be the case that traditional "A minus B" univariate analyses of BOLD signal will often lack the sensitivity needed to differentiate between certain closely-related auditory categories, whether they are specific to music (e.g. minor vs. major), speech (e.g. /ta/ vs. /da/), voice (male vs. female), etc. Recently, there has been a movement toward using multivariate information-based approaches to the localization of brain function. By looking at multiple neighboring voxels simultaneously, a "searchlight" of the brain may determine whether regionally-specific activity patterns can successfully predict and classify future events (Kriegeskorte, Goebel, & Bandettini, 2006). It follows that categorical maps, while distributed beyond individual voxels, may still be localizable to anatomically distinct regions (Staeren, Renvall, De Martino, Goebel, & Formisano, 2009). The study by Staeren et al. showed that activity in bilateral STS regions could be used as an effective predictor of both auditory object category (e.g. cat vs. guitar sounds) and fundamental frequency, with a significant degree of regional overlap between these two independent variables. These classifier-based results provide further evidence for a pivotal role of the STS in perception of category, while also suggesting that observation of distributed *patterns* of activity, though still regionally local, may be critical to the identification of more detailed and precise category maps.

4.3. Intraparietal sulcus

Bilateral activity in the IPS was observed in the second adaptation paradigm contrast, Adapt2, which compared oddball stimuli that crossed the minor/major boundary and the analogous oddballs from the orthogonal set. This was not a result that we had predicted: neither Liebenthal et al.'s nor Joanisse et al.'s phoneme studies had reported significant BOLD activity in either IPS. This region deserves additional examination with regard to what role it may be playing in CP of musical stimuli. The IPS is part of what has classically been considered the "dorsal stream" (Culham & Kanwisher, 2001). Some recent studies have suggested that the IPS may play a large role in dealing with the frequency relationships between stimuli. Rinne et al. (2007) observed IPS recruitment to large pitch shifts in sound discrimination tasks. Zarate and Zatorre (2008) and Zarate, Wood, and Zatorre (2010) showed that the IPS may play a major role in auditory feedback monitoring for vocal regulation following pitch-shifts and, additionally, may interact with the right pSTS to extract the directionality of such a pitch-shift. Another recent study (Foster & Zatorre, 2009) showed that performance of a task that involved transposition of melodies correlated with BOLD activity in the right IPS. This latter finding points to a role of the IPS in the cognition of *relative* pitch. Since interval categories are based upon frequency ratio relationships (and not the absolute frequency distance between two notes), it would follow that CP for chords may preferentially recruit neural networks that make use of interval "quality." In other words, the IPS may be recruited when comparing stimuli that differ in interval type (minor vs. major), but may not be utilized to as great an extent when such a quality is missing (e.g. in our orthogonal triads that differ in terms of absolute pitch space, but not in terms of "minor-" or "major-ness"). The above contrast from the adaptation protocol, which compares major/minor and orthogonal triad pairs of approximately equal discriminability (based on behavioral data), provides evidence for such recruitment. It is of note that the discrimination paradigm contrast, *Disc3*, which does not compare relative vs. absolute pitch conditions, lacks significant BOLD activity in either IPS. Thus it may be the case that the IPS is preferentially recruited to help manipulate musical category information, but is relatively less sensitive to which particular category or categories are present at any given time. Musical categories, including chords (minor, major, etc.) and intervals (3rd, 4th, 5th, etc.), differ along a spectrum that has a dimension of perceptual "size" (e.g. a 5th is perceived as being a larger interval than a 3rd). On the contrary, phonemes are not intuitively thought of in terms of size, or any other linear dimension (i.e. /ta/ cannot be thought of as larger than /da/) and hence lack inherent underlying ordering. The absence of analogy, in this particular dimension, between musical and phonemic categories may explain the lack of observed IPS activity in prior studies of speech categorization.

5. Conclusion

The present data provide evidence for the involvement of the right STS in CP of spectrally-complex auditory stimuli. The results support models of hemispheric specialization for differential spectral resolution, as well as the role of a ventral stream as the basis of CP of numerous stimulus types.

Acknowledgements

We thank Marc Bouffard for technical assistance, and Dr. Jeffrey Binder for ideas about the control condition, as well as the staff of the McConnell Brain Imaging Centre for help in acquiring the data. This work was supported by funding from the Canadian Institutes of Health Research.

References

Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception and Psychophysics*, *57*, 863–874.
Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, *10*(4), 417–429.
Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395.
Burns, E. M., & Ward, W. D. (1978). Categorical perception-phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456–468.
Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*(2), 192–205.
Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11*(2), 157–163.
Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423–3431.
Foster, N. E., & Zatorre, R. J. (2009). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*,
Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317–323.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393–402.
Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *Neuroreport*, *12*(8), 1749–1752.
Hutchison, E. R., Blumstein, S. E., & Myers, E. B. (2008). An event-related fMRI investigation of voice-onset time discrimination. *Neuroimage*, *40*(1), 342–352.
Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*(2), 632–639.
Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*(9), 2084–2093.
Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.
Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience*, *29*(16), 5234–5239.
Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology. Human Perception and Performance*, *54*, 358–368.
Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebal Cortex*, *15*(10), 1621–1631.
Mattingly, I. G., Liberman, A., Syrdal, A., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, *2*, 131–157.
Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., et al. (2001). A probabilistic atlas and reference system for the human brain: International Con-

sortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1412), 1293–1322.
Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*(1), 57–77.
Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebal Cortex*, *13*(12), 1362–1368.
Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebal Cortex*, *17*(10), 2251–2257.
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*, 245–255.
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724.
Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Science of the United States of America*, *97*(22), 11800–11806.
Rinne, T., Kirjavainen, S., Salonen, O., Degerman, A., Kang, X., Woods, D. L., et al. (2007). Distributed cortical networks for focused auditory attention and distraction. *Neuroscience Letters*, *416*(3), 247–251.
Schonwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*(6), 1521–1528.
Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, *19*(6), 498–502.
Worsley, K. J. (2005). An improved theoretical *P* value for SPMs based on discrete local maxima. *Neuroimage*, *28*(4), 1056–1062.
Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., et al. (2002). A general statistical analysis for fMRI data. *Neuroimage*, *15*(1), 1–15.
Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, *48*(2), 607–618.
Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, *40*(4), 1871–1887.
Zatorre, R. (1983). Category-boundary effects and speeded sorting with a harmonic musical-interval continuum: Evidence for dual processing. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(5), 739–752.
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebal Cortex*, *11*(10), 946–953.
Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Science*, *6*(1), 37–46.
Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *Journal of Neuroscience*, *24*(14), 3637–3642.
Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception and Psychophysics*, *26*, 384–395.
Zevin, J. D., & McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioural Brain Function*, *1*(1), 4.