Training humans to categorize monkey calls: Auditory feature- and category-selective neural tuning changes

Xiong Jiang[1], **Mark A. Chevillet**[1], **Josef P. Rauschecker**[1,2], **Maximilian Riesenhuber**[1,*]

[1]Department of Neuroscience, Georgetown University Medical Center, Washington DC, USA

[2]Institute for Advanced Study, Technische Universität München, Germany

SUMMARY

Grouping auditory stimuli into common categories is essential for a variety of auditory tasks, including speech recognition. We trained human participants to categorize auditory stimuli from a large novel set of morphed monkey vocalizations. Using fMRI-rapid adaptation (fMRI-RA) and multi-voxel pattern analysis (MVPA) techniques, we gained evidence that categorization training results in two distinct sets of changes: sharpened tuning to monkey-call features (without explicit category representation) in left auditory cortex, and category selectivity for different types of calls in lateral prefrontal cortex. In addition, the sharpness of neural selectivity in left auditory cortex, as estimated with both fMRI-RA and MVPA, predicted the steepness of the categorical boundary, whereas categorical judgment correlated with release from adaptation in the left inferior frontal gyrus. These results support the theory that auditory category learning follows a two-stage model analogous to the visual domain, suggesting general principles of perceptual category learning in the human brain.

INTRODUCTION

Object categorization is a crucial cognitive task occurring across all sensory modalities. At its core, it involves the learning and execution of a mapping of sensory inputs to behaviorally relevant labels (e.g., "friend" versus "foe"). A particular challenge in object recognition is that physically dissimilar stimuli can have the same label (e.g., the same phoneme spoken by different speakers, or the same face viewed under different lighting conditions), while physically similar stimuli might be labeled differently (e.g., based on the small differences in voice onset time that distinguish /d/ from /t/). In audition, higher mammals, including humans, have highly developed mechanisms to discriminate fine

*Correspondence should be addressed to the lead contact, M.R. (mr287@georgetown.edu). Department of Neuroscience, Georgetown University Medical Center, Research Building Room WP-12, 3970 Reservoir Rd. NW, Washington, DC 20007, USA.

acoustic differences among complex sounds and use them for auditory communication. This is one of the most remarkable achievements of auditory cortex and one that probably shapes its overall architecture. However, the underlying neural mechanisms are only poorly understood.

While there is now broad support among neurophysiologists and cognitive neuroscientists for the concept of a hierarchy of cortical areas subserving auditory processing, there is substantial disagreement about how this anatomical hierarchy supports the acquisition of auditory object categorization. For instance, several auditory categorization learning studies have revealed an involvement of the left superior temporal gyrus/sulcus (STG/STS), leading to the suggestion that the anterior part of auditory-sensory cortex is the key region in the identification and categorization of auditory "objects" (Binder et al., 2004; Griffiths and Warren, 2004; Kumar et al., 2007; Scott, 2005; Tsunada et al., 2011; Zatorre et al., 2004). In contrast, using functional magnetic resonance imaging (fMRI) techniques, other studies in which participants were trained to categorize non-speech sounds have reported learning effects in posterior STS (Leech et al., 2009), or category-selective tuning in "early" auditory areas (Ley et al., 2012). Yet, identifying the neural mechanisms that achieve the transformation from auditory representations based on acoustic similarity to representations reflecting category-based similarity necessary for perceptual categorization requires more informative fMRI imaging paradigms and analyses than the traditional univariate analysis of BOLD responses, which cannot quite measure neural selectivity and probe neural representations. By contrast, techniques like multivariate pattern analysis (MVPA) (Carp et al., 2010; Haxby et al., 2001; Norman et al., 2006) and fMRI rapid adaptation (fMRI-RA) (Grill-Spector et al., 2006; Jiang et al., 2007, 2013) have proven their ability to probe neural selectivity more directly than conventional fMRI techniques. Indeed, using MVPA (Ley et al., 2012) and fMRI rapid adaptation (fMRI-RA) (Glezer et al., 2015; Jiang et al., 2007) along with well-controlled pre-/post-training comparison scans, we and others have identified and isolated learning-induced neural changes in related brain regions that are less likely confounded by the tasks performed during the acquisition of fMRI data.

Motivated by a computational model of object recognition in the cerebral cortex (Riesenhuber and Poggio, 2002), we previously used fMRI-RA (Jiang et al., 2007) (later confirmed in a human electroencephalography-rapid adaptation (EEG-RA) study (Scholl et al., 2014)) to provide support for a two-stage model of perceptual category learning in vision (Riesenhuber and Poggio, 2002; Serre et al., 2007), involving a perceptual learning stage in visual cortex in which neurons come to acquire sharper tuning with a concomitant higher degree of selectivity for *physical* (e.g., visual shape) features of the training stimuli, with these stimulus-selective neurons then providing input to neural circuits located in higher cortical areas, such as prefrontal cortex (Freedman et al., 2003), which can learn to perform different functions on these stimuli, such as identification, discrimination, or categorization. A computationally appealing property of this hierarchical model is that the high-level perceptual representation in the sensory cortices can be used in support of other tasks involving the same stimuli (Riesenhuber and Poggio, 2002), permitting transfer of learning to novel tasks. Given that, from a computational point of view, categorization of both visual and auditory stimuli involves the same mapping problem from a sufficiently selective stimulus representation to a representation selective for the perceptual labels, we

hypothesized that the same two-stage model of perceptual category learning from vision might also apply to the auditory domain.

To test this hypothesis, we adopted an approach similar to our prior study of visual categorization learning (Jiang et al., 2007; Scholl et al., 2014), to study the learning of auditory categories by using an auditory morphing system that could precisely control for acoustic features. This allowed us to dissociate semantic (category) similarity from physical (acoustic) similarity (Chevillet et al., 2013). We combined two independent fMRI techniques, fMRI-rapid adaptation (fMRI-RA) and multi-voxel pattern analysis (MVPA). The results show that learning an auditory categorization task leads to the sharpening of stimulus representations in left auditory cortex. Both fMRI-RA and MVPA results support a direct link between the neural selectivity in this region and behavioral performance measured outside of the scanner. Moreover, both fMRI-RA and searchlight MVPA analyses revealed category-selective responses in lateral prefrontal cortex, but not in sensory (auditory) cortex. Our results provide strong support for the aforementioned model of two-stage perceptual categorization, indicating that similar learning strategies underlie perceptual categorization across sensory domains.

RESULTS

Stimuli, Category Training, and Categorical Judgment

Participants were trained to categorize stimuli from an auditory "morph space" continuum spanned by four auditory monkey call exemplars (two "coos" and two "harmonic arches", Figure 1A & 1B), using a two-alternative forced-choice (2AFC) task with feedback. The auditory stimuli within the (mathematical) space of morphed monkey calls were linear combinations of the four prototype exemplars (see STAR Methods). By blending differing prototype amounts from the two categories, we could continuously vary the acoustic features and precisely define the category boundary (with an equal 50% contribution from both categories). After an average of approximately six hours of training, participants (n=16) were able to reliably judge the membership of the morphed sound stimuli (see STAR Methods and Figure 2).

Following training, we fit each individual's categorization data on each morph line with a sigmoid function (solid lines in Figure 2),

$$f(x) = 1/(1 + \exp(-\alpha \times (x - \beta)); \tag{1}$$

to estimate the individual category boundary location (β) and steepness (α) (see STAR Methods for details). The steepness of individual morph lines was calculated for each subject individually, and used in the brain-behavioral correlation analyses described below. The average boundary steepness, a, and boundary location, b, for each of the four morph lines are shown in Figure S1.

FMRI-RA reveals sharpened neural selectivity in the left auditory cortex after categorization training

To probe neuronal selectivity using fMRI, we adopted an event-related fMRI-RA paradigm (Chevillet et al., 2013; Jiang et al., 2006, 2007) in which a pair of sounds of varying acoustic similarity was presented in each trial. The fMRI-RA approach is motivated by findings from monkey electrophysiology experiments reporting that when pairs of identical stimuli are presented sequentially, the second stimulus evokes a smaller neural response than the first (Miller et al., 1993). Additionally, it has been suggested that the degree of adaptation depends on stimulus similarity, with repetitions of the same stimulus causing the greatest suppression. In the fMRI version of this experiment, the BOLD-contrast response to a pair of stimuli presented in rapid succession is measured for pairs differing in specific perceptual aspects (e.g., shape or acoustic similarity), and the combined response level is assumed to predict the dissimilarity of the stimulus representations at the neural level (Grill-Spector et al., 2006). Indeed, we (Chevillet et al., 2013; Jiang et al., 2006, 2007, 2013) and others (Gilaie-Dotan and Malach, 2007; Murray and Wojciulik, 2004) have provided evidence that parametric variations of stimulus parameters can be linked to systematic modulations of the BOLD-contrast response in relevant sensory areas, and that fMRI-RA can be used as an indirect measure of neural population tuning. Furthermore, using the fMRI-RA technique, we have found that training-induced changes in behavioral performance can be linked to changes in neuronal tuning in corresponding brain regions (Glezer et al., 2015; Jiang et al., 2007). We hypothesized that following training, stimulus representations in auditory cortex should show increased selectivity, indicated by increased release from adaptation post-training relative to pre-training for similar stimulus pairs. For this study, we scanned participants using an event-related fMRI-RA paradigm presenting them with pairs of morphed auditory stimuli that varied in acoustic and categorical similarity (see Figure 1C). We created pairs of identical sounds (condition M0), pairs of sounds differing by 30% acoustic change (where 100% corresponded to the difference between two prototypes), with both sounds in a pair either belonging to the same category (i.e., 95%-65%, or 35%-5%), $M3_{within}$, or to different categories (i.e., 35%-65%, with the category boundary at the 50% mark), $M3_{between}$, and pairs of sounds differing by 60% acoustic change (i.e., 5%-65%, or 35%-95%), M6. In the scanner, subjects had to perform an attentionally demanding offset delay matching task (Chevillet et al., 2013) (see STAR Methods) for which the category labels of stimuli were irrelevant ("bottom-up" scan), thus avoiding potentially confounding influences due to difference in task difficulty across pairing conditions and other potential confounds caused by top-down effects of the task itself (Freedman et al., 2003; Grady et al., 1996; Sunaert et al., 2000). FMRI data using this paradigm were obtained both pre- and post-training. The contrast of $(M6 > M0) \times (Post > Pre)$, masked by Stimuli>Silence (p<0.001, uncorrected, see Figure S2A) was used to identify regions that showed increased neural selectivity (as evidenced by increased release from adaptation) to acoustic features post-training. This contrast revealed several clusters, including one cluster in the left, but not the right, superior temporal gyrus (termed "auditory cortex" here for brevity) (p<0.01, uncorrected, cluster size, 50 voxels, Table 1, also see Figure S2B). Similar results were observed with the contrast of $(M6\&M3_{between}\&M3_{within} > M0) \times (Post > Pre)$ (Table 1 and Figure S2C).

We then investigated the relationship between the fMRI-RA-estimated neural selectivity for acoustic features and the behavioral performance measured by the steepness, α, of subjects' categorization curves obtained outside the scanner. First, we identified an ROI in the left auditory cortex (MNI: −56 −36 2) using the contrast of M6>M0 (see Table 1 and Figure 3A) from the post-training scans, and extracted fMRI responses from this ROI (Figure 3B). Repeated-measures ANOVA with one within-subject factor, experimental condition (M0, $M3_{within}$, $M3_{between}$, and M6), revealed a significant difference between conditions, $F(3,45)$ = 5.224, $p<0.01$. Post-hoc paired t-tests revealed a significant difference between M6 and the other three conditions, M0, $p<0.004$, $M3_{within}$, $p<0.044$, and $M3_{between}$, $p<0.010$ (two-tailed), but not amongst the other three conditions. Furthermore, there was a significant correlation between fMRI adaptation in this ROI (measured as M6-M0) and the steepness (α) of subjects' categorization curves obtained outside the scanner, suggesting a direct link between behavioral performance and the acoustic selectivity in the left auditory cortex, $r=0.554$, $p<0.026$ (Figure 3C).

These results are in agreement with our hypothesis of training-induced neural sharpening of stimulus- but not explicitly category-selective representations in sensory cortex. The relatively weak effect in the whole-brain analysis might be due to the small sample size together with the variability associated with between-group comparison (Pre- vs. Post-training, see STAR Methods), and/or the lack of separate localizer scans for an ROI-based analysis, which is expected to increase the sensitivity in detecting differences (Saxe et al., 2006). To further validate the finding of sharpened neural tuning in left auditory cortex after training and the link between neural selectivity in this ROI and behavioral performance, we therefore conducted an independent analysis on the same data set using a different technique, multi-voxel pattern analysis, MVPA (Haxby et al., 2001).

MVPA reveals an increase in the distinctiveness of response patterns in the left auditory cortex to morphed monkey calls that correlates with behavioral performance

Compared to the univariate approach in traditional fMRI data analyses, multivariate analyses offer the potential for increased sensitivity (Norman et al., 2006). We adopted a method originally proposed by Haxby and colleagues (Haxby et al., 2001) and used by that group and others to estimate neural tuning (Carp et al., 2010; Haxby et al., 2001; Park et al., 2004). Briefly, we first estimated the fMRI responses to each distinct stimulus pair (n=14, see STAR Methods) using the General Linear Model for even and odd runs separately. We then calculated the correlations across activation maps between all 14 distinct stimulus-pairs of the odd and even runs. The correlation coefficients were Z-transformed, and the sharpness of neural tuning was estimated via the "distinctiveness" of activation patterns, which was defined as the difference between the mean same- and different-pairs' Z-transformed correlation coefficients, averaged over all such pairwise comparisons (Carp et al., 2010; Park et al., 2004). We limited our analysis to left and right auditory cortex – using the auditory ROIs defined by the contrast of stimuli > silence from the group whole-brain analysis (see STAR Methods and Figure S3A). The left and right auditory cortices were analyzed independently. For each individual subject, we used the N most strongly activated voxels (selected by the t-map of stimuli > silence, N=10, 20, 50, 100, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, 2000) within each ROI (left or right auditory cortex, Figure S3A and

S3B), as in previous studies (Park et al., 2004). The result with N=200 is shown here and used for the correlation analysis (Figure 4), results for other choices of N are shown in Supplementary Materials (Figure S3C and S3D). Similar results were obtained when using the anatomically defined superior temporal cortex (including Heschl's gyrus) (Figure S3E and S3F).

Two-sample t-tests revealed an increase in the distinctiveness of activation patterns after training, defined as the difference between the mean same- and different-pairs correlation coefficients, in the left ($p < 0.039$), but not the right ($p > 0.360$), auditory cortex (Figure 4A). Repeated-measure ANOVA with one within-subject factor, left vs. right hemisphere, and one between-subject factor, pre- vs. post-training scans, revealed a significant difference between the left and right hemisphere, $F(1,28) = 9.859$, $p < 0.005$, a marginal effect between the pre- and post-training scans ($p < 0.084$), and, critically, a significant interaction between the two factors, $F(2,28) = 4.085$, $p < 0.029$, suggesting that training sharpened neural tuning in the left, but not the right, auditory cortex, in line with the fMRI-RA data. Furthermore, similar to the fMRI-RA data, there was a significant correlation between MVPA-derived distinctiveness of activation patterns in the left auditory cortex and the steepness, α, of subjects' categorization curves obtained outside the scanner, $r=0.56$, $p < 0.025$ (Figure 4B), providing additional support for a direct correlation between behavioral performance and neuronal tuning selectivity as estimated by the distinctiveness of BOLD activation patterns (Williams et al., 2007).

To probe the consistency and location of training effects in auditory cortex, we computed a heat map of the number of times each individual voxel in the left auditory cortex was chosen for the N=200 analysis, shown in Figure 4C, ranging from n=1 (i.e., this specific voxel was only included in one subject) to n=10 (i.e., this specific voxel was included in a total of ten subjects). We then conducted additional analyses to compare the categorization training-induced changes in the distinctiveness of activation patterns in primary (A1) and non-primary (non-A1) auditory cortices in the left hemisphere; left A1 was defined anatomically (Tzourio-Mazoyer et al., 2002), and the left non-primary auditory cortex (non-A1) was defined as the region in Figure S3A minus A1. The result with N=200 is shown here, and the results for other choices of N are shown in Supplementary Materials (Figure S3G and S3H). Two-sample t-tests revealed an increase in the distinctiveness of activation patterns after training, defined as the difference between the mean same- and different-pairs correlation coefficients, in left non-A1 ($p < 0.035$), but not left A1 ($p > 0.781$)(Figure 4D). Repeated-measures ANOVA with one within-subject factor, A1 vs. non-A1, and one between-subject factor, pre- vs. post-training scans, revealed a marginal effect of interaction between the two factors, $F(2,28) = 2.795$, $p = 0.105$, suggesting that training might have sharpened neural tuning more in non-primary than in primary auditory cortex. In addition, there was a significant correlation between MVPA-derived distinctiveness of activation patterns in the left non-A1 region and the steepness (α) of categorization curves, $r=0.55$, $p < 0.026$ (Figure 4E). These results are in line with the presumed hierarchical structure of auditory cortical processing (Chevillet et al., 2011; Cohen et al., 2009, 2016; Tsunada and Cohen, 2014). The lack of a training effect in primary auditory cortex, as had been found in some previous studies (Hui et al., 2009; Ley et al., 2012; Yin et al., 2014), is likely due to the complex auditory stimuli used in the present study.

To further probe the specificity of training effects, we applied the MVPA technique to the five clusters shown in Figure 3A, using all voxels from each of the ROIs. Out of the five ROIs, only the one in the left auditory cortex (MNI: −56 −36 2) showed an increase in the distinctiveness of activation patterns after training (p=0.043), but not any of the other four ROIs (at least p>0.497) (Figure S3I), providing additional evidence that training refined neural tuning in the left auditory cortex, but not any other regions.

Training-induced category selectivity in task circuits as revealed by fMRI-RA

To probe which brain regions exhibited category-related selectivity and thus might include category-selective neurons, we scanned participants following training using the same fMRI-RA paradigm and stimuli, this time while they were performing a categorization task requiring them to judge whether the two auditory stimuli in each trial belonged to the same or different categories. Thus, the pair of auditory stimuli in the M0 and $M3_{within}$ conditions belonged to the same category, while the pairs of auditory stimuli in the $M3_{between}$ and M6 conditions belonged to different categories. We predicted that brain regions containing category-selective neurons should show stronger responses in the $M3_{between}$ and M6 trials than in the $M3_{within}$ and M0 trials, as the stimuli in each pair in the former two conditions should activate different neuronal populations while they would activate the same group of neurons in the latter two conditions (Jiang et al., 2007). The average behavioral performance is shown in Figure 5A.

We first conducted a whole-brain analysis (see STAR Methods) to examine the brain regions that were involved in the categorization task, using the contrast of Stimuli > Silence. This contrast revealed a broad network of brain regions, including auditory cortex, superior/middle temporal pole, supplementary motor area (SMA), dorsal and ventral premotor cortex, motor cortex, cerebellum, lateral prefrontal cortex, inferior parietal cortex, angular/supramarginal gyrus, precuneus, occipital cortex, hippocampus, insular cortex, putamen, and thalamus (see Figure S4A for more detailed information).

To probe the brain regions that were sensitive to category differences, we first compared the activation of M6 versus M0, since participants could very reliably judge the category memberships of the pair of auditory stimuli in the M0 and M6 conditions. As listed in Table 2, several brain regions, including prefrontal and parietal cortices, showed stronger activations to M6 than to M0 (also see Figure S4B). To further examine the differential activations to trials in which the two calls belonged to the same ($M3_{within}$ and M0) versus different categories (M6 and $M3_{between}$), a comparison of M6 and $M3_{between}$ versus $M3_{within}$ and M0 was conducted, and similar brain regions were found (Table 2 and Figure S4C), further supporting the involvement of these brain areas in the representation of learned stimulus categories. We then extracted the fMRI responses in the left inferior frontal gyrus (IFG) ROI (MNI: −48 8 18). There were significant differences between M0 and the three other conditions (at least p<0.0003), but not between the three conditions, $M3_{within}$, $M3_{between}$, and M6 (at least p>0.33) (Figure 5B). Yet, this lack of a difference between the $M3_{within}$ and $M3_{between}$ conditions might have been due to the low performance on these conditions (Figure 5A), suggesting the fMRI response in this region might correlate with the subjects' categorical decision. To test this hypothesis, i.e., that a link exists between

activations in the regions identified in the category-selective whole-brain analysis and behavioral performance on the categorization task, we followed the approach used in our earlier visual study (Jiang et al., 2007) and examined the correlation of the difference between the fMRI response for the $M3_{between}$ and $M3_{within}$ conditions in this ROI (as an index of how sharply neurons in this area differentiated between the two categories) and the average of the behavioral categorization accuracy on those trials within the scanner (as a measure of behavioral performance), predicting a positive correlation between the two variables. It is worth noting that the ROI definition, M6 versus M0, was independent of the conditions involved in the correlation analysis, $M3_{between}$ versus $M3_{within}$. This analysis revealed a significant correlation between behavioral performance in the $M3_{within}$ and $M3_{between}$ conditions and the fMRI adaptation effects ($M3_{between}$ - $M3_{within}$) in the left lateral prefrontal cluster (r=0.562, p<0.024) (Figure 5C), suggesting a direct link between category judgment decisions and category selectivity in these areas. By contrast, there were no significant correlations between behavioral performance and responses in the other brain regions identified in Table 2 (M6>M0), with the lowest p value for the right IFG ROI (r=0.33, p<0.207, see Figure S4D–H).

Previously we (Jiang et al., 2007; Roy et al., 2010) and others (van der Linden et al., 2014) have shown that category tuning in prefrontal cortices was task-dependent. Specifically, responses indicated category selectivity when subjects were performing a categorization task, but no such selectivity was found when subjects were performing a spatial localization task (Jiang et al., 2007). Hence, we examined the responses in these ROIs when subjects were performing the offset detection task post-training for which category membership of stimuli was irrelevant, by extracting the fMRI signal from this left IFG ROI using the data from fMRI Experiment 2 (Figure 5D). Repeated-measure ANOVA with two within-subject factors, task (offset detection vs. categorization) and experimental conditions (M0 vs. $M3_{within}$ vs. $M3_{between}$ vs. M6), revealed a significant difference between experimental conditions, $F(3,45)=17.347$, p<0.001, but not between tasks, $F(1,15)=0.438$, p=0.518, and, critically, a significant interaction between the two factors, $F(3,45) = 8.999$, p=0.001, suggesting the adaptation at this IFG ROI is task-dependent. For the offset detection task (Figure 5D), post-hoc paired t-tests revealed no significant difference between the M0 and M6 conditions or any other comparisons, except $M3_{between}$ vs M6 (p<0.001).

Category-selective circuits revealed via MVPA

To further investigate brain regions that were selective for the trained categories, we employed a searchlight approach using the MVPA methodology described above. Briefly, the fMRI responses to each distinct pair were estimated, but analysis was limited to the four distinct M0 pairs (each consisting of two identical stimuli that are close to one of the four prototypes, see Figure 1C), with two distinct stimulus pairs for each category: Call-1-1, Call-2-2, Call-3-3, and Call-4-4, with the first two sets belonging to one category, and the last two belonging to the other category. We then calculated the correlations across activation maps for M0 trials that belonged to the same- and different-categories between odd and even runs (see STAR Methods). Category selectivity was defined as the difference in Z-transformed correlation coefficients between same- (e.g., Call-1-1 vs Call-2-2) and different-category (e.g., Call-1-1vs Call-3-3) pairs. A 7x7x7 voxel cubic searchlight

approach (see STAR Methods) revealed four category-selective clusters (Figure 6A) (p<0.005, uncorrected, at least 50 contiguous voxels), masked by fMRI-RA (M6>M0) and functional connectivity analysis (see below) (a conjunction of p<0.05, uncorrected). It might be worth noting that all four ROIs were in frontal regions of the brain, including one in the left IFG, two in the right IFG, and one in the SMA, whereas none were found in superior temporal cortex (STC) nor in more posterior brain regions, suggesting a central role of frontal cortex in making category decisions, providing converging evidence in support of the aforementioned two-stage model of categorization. The fMRI response profiles in these ROIs (during the top-down scans) are shown in Figure 6B.

To investigate whether these category-selective regions (identified via fMRI-RA and MVPA) were indeed part of the brain regions involved in the categorization of the morphed monkey calls, we further investigated the overlap between the searchlight MVPA-revealed category-selective regions (Figure 6C, shown in red), the fMRI-RA-revealed category-selective regions (M6>M0) (in green), and the regions functionally connected to left auditory cortex (in blue) identified using functional connectivity analysis (see STAR Methods) (Figure 6C). For illustration purposes, a reduced threshold was used (p<0.01 for fMRI-RA and MVPA, and p<0.001 for functional connectivity, uncorrected, at least 20 contiguous voxels). A strong overlap was found to exist between the brain maps revealed by the three techniques, suggesting category-selective regions (identified from both fMRI-RA and MVPA analyses) were part of a broader network of brain regions connected to auditory cortex when subjects were categorizing the auditory stimuli.

DISCUSSION

The present study tackles an important problem of auditory perception: How do we form categories of sounds? This task requires generalizing over dissimilar sounds that belong to the same category, while remaining sensitive to small differences that distinguish categories. Work on category learning in visual perception has established that the process consists of two essentially separate sub-processes: an improvement in tuning of neurons in sensory regions, and a change in top-down projections from frontal regions to sensory regions. In this study, we adopted an experimental paradigm used previously in related studies of the visual system (Jiang et al., 2007; Scholl et al., 2014) and trained human participants to categorize auditory stimuli from a large novel set of morphed monkey vocalizations and measured tuning before and after training. Training-induced changes in neuronal selectivity were examined using two independent fMRI-techniques: fMRI-rapid adaptation (fMRI-RA) and multi-voxel pattern analysis (MVPA). Both of these techniques have been used previously to measure neuronal tuning. The two approaches revealed independently that categorization training with the novel monkey calls sharpened the neural tuning to acoustic features of the auditory stimuli in the left auditory cortex while inducing task-dependent category tuning in lateral prefrontal cortex. Furthermore, functional connectivity analysis revealed that task-dependent category-selective regions were part of a broad network of brain regions connected to auditory cortex, suggesting they are situated at a higher level in the auditory categorization process. Similar to the findings in the visual domain (Jiang et al., 2007), the present study also finds a significant correlation between neural selectivity in the lateral prefrontal cortex with behavioral categorization ability. It appears, therefore, that

perceptual category learning makes use of identical neural mechanisms in the visual and auditory domains.

Compared to the visual system, much less is known about category learning in the auditory system. This is surprising, because learning to distinguish and produce speech sounds in one's own language depends heavily on categorical perception (Repp, 1984). Early accounts have emphasized the importance of motor representations of what were thought to be the elements of speech ("phonemes") (Liberman et al., 1967). Categorical perception is invariant against changes in motor representations, so perceptual invariance was thought to be based on these motor representations. Although the theory of direct perception through the motor system (the "motor theory of speech perception") has long been discredited, modern terminology has turned to the concept of sensorimotor integration and control as a closed-loop system or internal model, in which speech perception and production are closely intertwined (Chevillet et al., 2011; Rauschecker, 2011; Rauschecker and Scott, 2009). It has been argued that the conversion of perceptual into articulatory representations (and vice versa) occurs in inferior frontal regions ("Broca's area"; (Rauschecker and Scott, 2009)).

In the dual-stream model of auditory perception, as originally formulated in the rhesus macaque (Rauschecker, 1998; Rauschecker et al., 1995), auditory perception, including the recognition of "auditory objects", is achieved in the auditory ventral stream, which originates in the anterolateral belt areas of auditory cortex and projects to ventrolateral prefrontal cortex (Romanski et al., 1999). The hierarchical structure of auditory cortical processing in the ventral stream of the monkey has led to similar models of auditory and speech perception in man (Hickok and Poeppel, 2007; Leaver and Rauschecker, 2010, 2016; Norman-Haignere et al., 2015; Rauschecker and Scott, 2009). It makes sense, therefore, that exposure to unfamiliar sounds in the present study leads, first of all, to tuning changes at the level of the auditory cortex. The fact that the changes were strongest on the left, even though the stimuli were not human speech, may be seen as an indication that communication sounds almost universally use information-bearing elements that are in a similar spectro-temporal space across species and speaks for the gradual evolution of human speech from more primitive animal communication systems. Even though human articulators are vastly more diverse and refined than those in monkeys, some of the most basic elements of speech, such as frequency-modulated (FM) sweeps and band-passed noise bursts, are also contained in monkey calls and have a rich representation in rhesus monkey auditory cortex (Rauschecker and Tian, 2004; Tian and Rauschecker, 2004; Tian et al., 2001). Lateralization of learning to the left hemisphere could also have been influenced by the use of verbal category labels, as has been hypothesized for the visual domain (van der Linden et al., 2014). Indeed, our prior visual study that did not use verbal labels found right-lateralized learning (Jiang et al., 2007), while our previous visual learning study using the same stimuli with verbal labels (Scholl et al., 2014) found learning effects predominantly in the left hemisphere. It will be interesting in future studies to investigate how verbal vs. non-verbal category labels influence the lateralization of learning, and whether these effects are again similar in visual and auditory learning.

There are some obvious limitations of the present study. First, while training duration in the present study was comparable to studies of category learning in the visual domain (Gillebert

et al., 2009; Jiang et al., 2007; van der Linden et al., 2014; Scholl et al., 2014), this amount of exposure is small compared to the brain's accumulated exposure to environmental or speech sounds such as phonemes or words, and it will be interesting to probe in future studies if extensive experience with novel stimuli is associated with different learning effects. Second, while our study focused on stimulus categorization, a cognitive task found across sensory domains, the auditory domain offers the additional opportunity to study the interaction of perception and production, and future studies might investigate how training to categorize not just acoustic-phonetic stimuli but also their articulatory counterparts leads to additional changes in other brain areas (Rauschecker et al., 2008).

Taken together, the very similar visual and auditory category learning results further enhance the concept that sensory systems function according to similar principles, once sensory information reaches the level of the cortex. This begins with receptive field organization of simple- and complex-like cells in primary areas (Tian et al., 2013) as an early indicator of hierarchical processing; it is further evident from the existence of dual anatomical and functional processing streams (Rauschecker and Tian, 2000; Ungerleider and Haxby, 1994); and culminates in the role of prefrontal cortex, which translates sensory into task-based representations (Goldman-Rakic et al., 1996).

STAR*METHODS

GENERAL

Please see below.

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Maximilian Riesenhuber (mr287@georgetown.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects—A total of twenty-eight subjects were enrolled in the study (16 females, 18–32 years of age). There were three fMRI-scans, a pre-training scan with a dichotic listening task (Experiment 1), a post-training scan with the same task (Experiment 2), and a post-training scan with a categorization task (Experiment 3). Out of the 28 subjects, ten subjects only participated in Experiment 1 (dropping out of the study before finishing up training, mostly due to boredom), eleven subjects (enrolled to compensate for the dropped-out subjects – budget constraints did not allow pre-training scans in those replacement subjects) only participated in Experiments 2 and 3, one subject only participated in Experiments 1 and 2, and six subjects participated in all three scans (Experiment 1, 2, and 3). Data from two subjects were excluded, including the subject who only participated in Experiment 1 and 2, due to chance-level categorization performance after training, and another subject who participated in all three scans due to missing behavioral data. In the end, for pre-training dichotic, post-training dichotic, and post-training categorization scans, data from 15, 16, and 16 subjects were included, respectively. Among them, five subjects finished all three scans, nine subjects only finished the pre-training scan (Experiment 1), and 11 subjects only finished the two post-training scans (Experiment 2 and 3). The Georgetown University

Institutional Review Board approved all experimental procedures, and all subjects gave written informed consent before participating. All subjects were right-handed, and reported no history of hearing problems or neurological illness.

METHOD DETAILS

Stimuli—To probe neuronal tuning for a novel (and yet natural) class of sounds, we generated acoustic continua between natural recordings of monkey calls using the MATLAB toolbox STRAIGHT. STRAIGHT allows for finely and parametrically manipulating the acoustic structure of high-fidelity natural voice recordings (Kawahara and Matsui, 2003). Using this toolbox, it is possible to create an acoustic continuum of stimuli that gradually "morph" from one prototype stimulus to another. Monkey call stimuli were recordings of natural monkey calls from an existing digital library (Hauser, 1998; Shannon et al., 1999). As shown in Figure 1, we created a novel acoustic stimulus space, defined by two "harmonic arch" calls and two "coo" calls. To do this, we extended the STRAIGHT toolbox to accommodate four prototype stimuli, and to assign each prototype a weight in the morphing calculation, creating a four-dimensional monkey call "morph space". By morphing different amounts of the prototypes we could generate thousands of unique morphed monkey calls, and precisely define a category boundary (see Figure 1). The category of a stimulus was defined by whether "harmonic arch" or "coo" prototypes contributed more (>50%) to a given morph. Thus, stimuli that were close to, but on opposite sides of, the boundary could be similar, whereas stimuli that belonged to the same category could be dissimilar. This careful control of acoustic similarity within and across categories allowed us to disentangle the neural signals for acoustic vs. category selectivity explicitly. For training, stimuli were selected from the whole morph space (see below), whereas for testing, we created four cross-category "morph lines" by generating stimuli at 5% intervals between all pairs of prototypes from different categories (where 100% is the total acoustic difference between prototypes). Morphed stimuli were generated up to 25% beyond each prototype, for a total of 301 stimuli per morph. All stimuli were then resampled to 48 kHz, trimmed to 300 ms duration and root-mean-square (RMS) normalized in amplitude. A linear amplitude ramp of 10 ms duration was applied to sound offsets to avoid auditory artifacts. Amplitude ramps were not applied to the onsets, however, to avoid interfering with the natural onset features of the monkey calls.

Categorization Training and Testing—Subjects were trained to categorize sounds from the monkey call auditory morph space using a 2AFC (two alternative-forced choice) task with feedback. Training utilized a web-based training paradigm that allowed subjects to train themselves on the auditory categorization task from home (Scholl et al., 2014). Training progress was monitored remotely, ensuring high compliance and increasing efficiency. In each trial during training, a single sound was presented as well as two possible labels ("joizu" and "eheto") presented side by side ("joizu + eheto", or "eheto + joizu"), randomly chosen to avoid a fixed link between category label and motor response.

When a subject chose an incorrect label, the subject received visual feedback that an error was made. The sound was then repeated once, along with the correct label. Similar to our previous studies in monkeys and humans (Freedman et al., 2003; Jiang et al., 2007), subjects

were initially trained on the regions of the morph space farthest from the category boundary. Stimuli were randomly selected from the regions where stimuli contained 90–100% contribution from one category. The task was then gradually made more difficult by extending the training region closer to the category boundary in 5% increments per training block. Subjects were required to correctly categorize > 90% of the 200 stimuli presented from each region before moving on to the next level of difficulty. There were 200 trials per online training session. After participants reached the highest level of task difficulty, their categorization performance along the four morph lines was measured at a morph step discretization of 20 steps (in increments of 5% morph difference) between the two prototypes using the same 2AFC paradigm as in the training period but without feedback. Note that, as in our visual studies (Jiang et al., 2007; Scholl et al., 2014) different stimuli were used during training (where auditory stimuli were randomly chosen from the morph space) and testing (where auditory stimuli were constrained to lie on the relevant morph lines).

We then fit the resulting data with a sigmoid function to estimate the boundary location as well as boundary sharpness for each subject (see Equation 1 and Figure 2).

Event Related Adaptation Experiments 1 & 2 ("Bottom Up", Dichotic Listening Task)—To observe responses largely independent of category processing, we scanned subjects while they performed an attention-demanding distractor task using the same experimental design as in our previous fMRI-RA study of phoneme processing (Chevillet et al., 2013). In each trial, subjects heard pairs of sounds, each of which was 300 ms in duration, separated by 50 ms (as used in previous studies (Chevillet et al., 2013; Joanisse et al., 2007; Myers et al., 2009)). Each of these sounds persisted slightly longer in one ear than in the other (~30 ms between channels). The subject was asked to listen for these offsets, and to report whether the two sequentially played sounds persisted longer in either the same or different ears. Subjects held two response buttons, and an "S" and a "D," indicating "same" and "different," were presented on opposing sides of the screen to indicate which button to press. Their order was alternated on each run, to disentangle activation due to decisions from motor activity (i.e., to average out the motor responses). The nature of this task required that subjects listen closely to all sounds presented, but was independent of the category nature of the sounds. The average performance across subjects was 69.8±4.2% in pre-training and 70.9±4.8% in post-training scans (no significant difference between the four conditions in both pre- and post-training scans, nor between the pre- and post-training scans, at least p>0.1), indicating that the task was attention-demanding, minimizing the chance that subjects covertly categorized the stimuli in addition to doing the dichotic listening task. Images were collected for 6 runs, each run lasting 669 s. Trials lasted 12 s each, yielding 4 volumes per trial, and there were two silent trials (i.e., 8 images) at the beginning and end of each run. The first 4 volumes of each run were discarded. Analyses were performed on the 52 trials – 10 to 11 each of the five different conditions defined by the change of acoustics and category between the two stimuli presented in each trial: M0, same category and no acoustic difference; $M3_{within}$, same category and 30% acoustic difference; $M3_{between}$, different category and 30% acoustic difference; M6, different category and 60% acoustic difference; and null trials (Fig. 1C). All stimuli were from morph lines 1 and 4. Trial order

was randomized and counterbalanced using M-sequences (Buracas and Boynton, 2002), and the number of presentations was equalized for all stimuli in each experiment.

Event Related Adaptation Experiment 3 ("Top Down", Categorization Task)—
To assess the neural mechanisms underlying the auditory categorization process, participants also participated in one more fMRI-RA experiment following Experiment 2 after categorization training, using the same design as in Experiments 1 and 2, except that participants now were asked to judge whether the two auditory stimuli within each trial belonged to the same or different categories, with "same category" and "different category" buttons switched across runs, to unconfound motor responses from category selectivity, following a similar rationale as in the dichotic listening task (see the previous section).

fMRI Data Acquisition—All MRI data were acquired at the Center for Functional and Molecular Imaging at Georgetown University on a 3.0-Tesla Siemens Trio Scanner using whole-head echo-planar imaging (EPI) sequences (Flip Angle = 90°, TE = 30 ms, FOV = 205, 64x64 matrix) with a 12-channel head coil. In both fMRI-adaptation experiments, a clustered acquisition paradigm (TR = 3000 ms, TA = 1500 ms) was used such that each image was followed by an equal duration of silence before the next image was acquired. Stimuli were presented after every fourth volume, yielding a trial time of 12.0 s. In all functional scans, 28 axial slices were acquired in descending order (thickness = 3.5 mm, 0.5 mm gap; in-plane resolution = 3.0 x 3.0 mm^2). Following functional scans, high-resolution (1x1x1 mm^3) anatomical images (MPRAGE) were acquired. Auditory stimuli were presented using Presentation (Neurobehavioral Systems) via customized STAX electrostatic earphones at a comfortable listening volume (~65–70 dB) worn inside ear protectors (Bilsom Thunder T1) giving ~26 dB attenuation.

QUANTIFICATION AND STATISTICAL ANALYSIS

fMRI Data Analysis—Data were analyzed using the software package SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). After discarding images from the first twelve seconds of each functional run, EPI images were temporally corrected to the middle slice, spatially realigned, resliced to 2 x 2 x 2 mm^3, and normalized to a standard MNI reference brain in Talairach space. Images were then smoothed using an isotropic 6 mm Gaussian kernel. For whole-brain analyses, a high-pass filter (1/128 Hz) was applied to the data. We then modeled fMRI responses with a design matrix comprising the onset of predefined non-null trial types (M0, M3$_{within}$, M3$_{between}$ and M6) as regressors of interest using a standard canonical hemodynamic response function (HRF), as well as six movement parameters and the global mean signal (average over all voxels at each time point) as regressors of no interest. The parameter estimates of the HRF for each regressor were calculated for each voxel. The contrasts for each trial type against baseline at the single-subject level were computed and entered into a second-level model (ANOVA) in SPM8 (participants as random effects) with additional smoothing (8 mm). For all whole-brain analyses, thresholds of at least p < 0.001 (uncorrected) and at least 20 contiguous voxels were used unless specified otherwise.

Multi-Voxel Pattern Analysis (MVPA)—To further investigate the change in neural encoding in auditory cortex after categorization training, we used multi-voxel pattern analysis (MVPA) to examine the pattern of neural activations to different stimulus pairs, following the method developed by Haxby and colleagues (Haxby et al., 2001) and used by others (Carp et al., 2010; Park et al., 2004). There were a total of 14 unique stimulus pairs, 7 from each of two morph lines (including one of each from the M3$_{between}$ condition, and 2 of each from M0, M3$_{within}$, and M6 conditions, see Fig. 1C). Previous studies have provided evidence that overlap in BOLD response patterns to difference stimuli is related to neural tuning (Carp et al., 2010; Park et al., 2004). For this analysis, we first re-modeled fMRI responses with a design matrix comprising the onset of the fourteen unique stimulus pairs (regardless of trial types) as regressors of interest using a standard canonical hemodynamic response function (HRF), as well as six movement parameters and the global mean signal (average over all voxels at each time point) as regressors of no interest. We then obtained the neural responses to each unique stimulus pair (n=14) relative to the null trials from the even and odd runs. Next, we obtained the correlations across activation maps for all pairs of stimulus pairs for both left and right auditory cortex (using activation maps (stimuli > null trials) from the group analysis) – there were 196 (14$_{odd}$ x 14$_{even}$) sets of correlation coefficients, with 14 sets from the correlation of the same stimulus pairs (e.g., even run stimulus pair 1, 5%Call1_95%Call3 (prime) & 35%Call1_65%Call3 (target), correlated with odd run stimulus pair 2, 5%Call1_95%Call3 (prime) & 35%Call1_65%Call3 (target)) of even and odd runs, and 182 from that of different stimulus pairs (e.g., even run stimulus pair 1, 5%Call1_95%Call3 (prime) & 35%Call1_65%Call3 (target), correlated with odd run stimulus pair 2, 35%Call2_65%Call4 (prime) & 65%Call2_35%Call4 (target)). Correlation coefficients were Fisher Z-transformed and the difference between the mean same- and different-stimulus pair correlations was calculated as an indirect measure of neural tuning. The difference in neural tuning in pre- and post-training data was analyzed with two-sample t-tests for left and right auditory cortex separately, using different numbers of voxels (10, 20, 50, 100, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, and 2000) to assess the robustness of the observed effects. Voxels were chosen based on their response amplitude relative to null trials (selecting the voxels with the highest responses, as in previous studies (Carp et al., 2010; Park et al., 2004)).

MVPA Searchlight Analysis for Categorical Representations—To further investigate the brain regions selective to the trained categories, we employed a searchlight approach using the MVPA methodology described above. First, the fMRI responses to each distinct pair (n=14) were estimated for the top-down scans, but unlike the MVPA analysis for stimulus selectivity above, here we limited the analysis to the four distinct M0 pairs (each consisting of two identical stimuli close to one of the four prototypes, see Figure 1C) – with two distinct pairs in each category: Call-1-1, Call-2-2, Call-3-3, and Call-4-4, with the first two belonged to one category, and the last two belonged to the other category. Subjects were able to make reliable categorical judgments on the four M0 pairs (see Fig. 5A). We then calculated the correlations across activation maps for the four distinct pairs (with two pairs from each category) between odd and even runs, which led to 16 (4x4) sets of correlation coefficients, with 8 sets from two pairs of stimuli that belonged to the same category and 8 sets from two pairs of stimuli that belonged to the different categories. The

correlation coefficients were Z-transformed, and the categorical selectivity was then estimated via the "distinctiveness" of categorical activation patterns, which was defined as the difference between the mean same-(e.g., Call-1-1 vs Call-2-2) and different-category pairs (e.g., Call-1-1 vs Call-3-3) Z-transformed correlation coefficients.

A 7x7x7 voxel cubic searchlight was used, and the difference between the mean same- and different-category pairs Z-transformed correlation coefficients was assigned to the center voxel. The resulting "categorical" map from each subject was then smoothed with an additional isotropic 8 mm Gaussian kernel, and entered into a second-level whole brain analysis (one-sample t-test with a gray matter mask). To exclude voxels that were less likely to be involved in the categorization judgment, a mask with the contrast of stimuli > silence was applied (using the t-map with a threshold at 0.85, approximately corresponding to $p < 0.2$, uncorrected), and only voxels that survived this threshold were used to calculate the correlation coefficients. Similar results were obtained using different thresholds and different searchlight sizes.

Functional Connectivity Analysis—To examine a network of brain regions involving in the categorization task, a functional connectivity analysis was conducted using the left auditory cortex ROI (Figure S3A) as the seed region (Whitfield-Gabrieli and Nieto-Castanon, 2012). Briefly, we first extracted the time series from the seed region, then calculated the pairwise correlations between the seed region and each individual voxel from the whole brain, after controlling for nuisance factors, including head motion, signal from white matter and CSF, and whole brain global signal. The correlation coefficients were then z-transformed and further smoothed with an additional isotropic 8 mm Gaussian kernel before entering into second-level group analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

Binder JR, Liebenthal E, Possing ET, Medler DA, and Ward BD (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci 7, 295–301. [PubMed: 14966525]

Buracas GT, and Boynton GM (2002). Efficient design of event-related fMRI experiments using M-sequences. NeuroImage 16, 801–813. [PubMed: 12169264]

Carp J, Park J, Polk TA, and Park DC (2010). Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. NeuroImage 56, 736–743. [PubMed: 20451629]

Chevillet M, Riesenhuber M, and Rauschecker JP (2011). Functional correlates of the anterolateral processing hierarchy in human auditory cortex. J. Neurosci. Off. J. Soc. Neurosci 31, 9345–9352.

Chevillet MA, Jiang X, Rauschecker JP, and Riesenhuber M (2013). Automatic phoneme category selectivity in the dorsal auditory stream. J. Neurosci. Off. J. Soc. Neurosci 33, 5208–5215.

Cohen YE, Russ BE, Davis SJ, Baker AE, Ackelson AL, and Nitecki R (2009). A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. Proc. Natl. Acad. Sci. U. S. A 106, 20045–20050. [PubMed: 19897723]

Cohen YE, Bennur S, Christison-Lagay K, Gifford AM, and Tsunada J (2016). Functional Organization of the Ventral Auditory Pathway. Adv. Exp. Med. Biol 894, 381–388. [PubMed: 27080679]

Freedman DJ, Riesenhuber M, Poggio T, and Miller EK (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. Off. J. Soc. Neurosci 23, 5235–5246.

Gilaie-Dotan S, and Malach R (2007). Sub-exemplar shape tuning in human face-related areas. Cereb. Cortex N. Y. N 1991 17, 325–338.

Gillebert CR, Op de Beeck HP, Panis S, and Wagemans J (2009). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. J. Cogn. Neurosci 21, 1054–1064. [PubMed: 18752400]

Glezer LS, Kim J, Rule J, Jiang X, and Riesenhuber M (2015). Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. J. Neurosci. Off. J. Soc. Neurosci 35, 4965–4972.

Goldman-Rakic PS, Cools AR, and Srivastava K (1996). The Prefrontal Landscape: Implications of Functional Architecture for Understanding Human Mentation and the Central Executive [and Discussion]. Philos. Trans. R. Soc. B Biol. Sci 351, 1445–1453.

Grady CL, Horwitz B, Pietrini P, Mentis MJ, Ungerleider LG, Rapoport SI, and Haxby JV (1996). Effect of task difficulty on cerebral blood flow during perceptual matching of faces. Hum. Brain Mapp 4, 227–239. [PubMed: 20408201]

Griffiths TD, and Warren JD (2004). What is an auditory object? Nat. Rev. Neurosci 5, 887–892. [PubMed: 15496866]

Grill-Spector K, Henson R, and Martin A (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci 10, 14–23. [PubMed: 16321563]

Hauser (1998). Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. Anim. Behav 55, 1647–1658. [PubMed: 9642008]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, and Pietrini P (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430. [PubMed: 11577229]

Hickok G, and Poeppel D (2007). The cortical organization of speech processing. Nat. Rev. Neurosci 8, 393–402. [PubMed: 17431404]

Hui GK, Wong KL, Chavez CM, Leon MI, Robin KM, and Weinberger NM (2009). Conditioned tone control of brain reward behavior produces highly specific representational gain in the primary auditory cortex. Neurobiol. Learn. Mem 92, 27–34. [PubMed: 19249380]

Jiang X, Rosen E, Zeffiro T, VanMeter J, Blanz V, and Riesenhuber M (2006). Evaluation of a Shape-Based Model of Human Face Discrimination Using fMRI and Behavioral Techniques. Neuron 50, 159–172. [PubMed: 16600863]

Jiang X, Bradley E, Rini RARRA, Zeffiro T, Vanmeter J, and Riesenhuber M (2007). Categorization training results in shape- and category-selective human neural plasticity. Neuron 53, 891–903. [PubMed: 17359923]

Jiang X, Bollich A, Cox P, Hyder E, James J, Gowani SA, Hadjikhani N, Blanz V, Manoach DS, Barton JJS, et al. (2013). A quantitative link between face discrimination deficits and neuronal selectivity for faces in autism. NeuroImage Clin 2, 320–331. [PubMed: 24179786]

Joanisse MF, Zevin JD, and McCandliss BD (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. Cereb. Cortex N. Y. N 1991 17, 2084–2093.

Kawahara H, and Matsui H (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., (IEEE), p. I-256–I-259.

Kumar S, Stephan KE, Warren JD, Friston KJ, and Griffiths TD (2007). Hierarchical processing of auditory objects in humans. PLoS Comput. Biol 3, e100. [PubMed: 17542641]

Leaver AM, and Rauschecker JP (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. Off. J. Soc. Neurosci 30, 7604–7612.

Leaver AM, and Rauschecker JP (2016). Functional Topography of Human Auditory Cortex. J. Neurosci. Off. J. Soc. Neurosci 36, 1416–1428.

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

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

Liberman AM, Cooper FS, Shankweiler DP, and Studdert-Kennedy M (1967). Perception of the speech code. Psychol. Rev 74, 431–461. [PubMed: 4170865]

van der Linden M, Wegman J, and Fernández G (2014). Task- and experience-dependent cortical selectivity to features informative for categorization. J. Cogn. Neurosci 26, 319–333. [PubMed: 24047386]

Miller EK, Li L, and Desimone R (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J. Neurosci. Off. J. Soc. Neurosci 13, 1460–1478.

Murray SO, and Wojciulik E (2004). Attention increases neural selectivity in the human lateral occipital complex. Nat. Neurosci 7, 70–74. [PubMed: 14647291]

Myers EB, Blumstein SE, Walsh E, and Eliassen J (2009). Inferior frontal regions underlie the perception of phonetic category invariance. Psychol. Sci 20, 895–903. [PubMed: 19515116]

Norman KA, Polyn SM, Detre GJ, and Haxby JV (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci 10, 424–430. [PubMed: 16899397]

Norman-Haignere S, Kanwisher NG, and McDermott JH (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. Neuron 88, 1281–1296. [PubMed: 26687225]

Park DC, Polk TA, Park R, Minear M, Savage A, and Smith MR (2004). Aging reduces neural specialization in ventral visual cortex. Proc. Natl. Acad. Sci. U. S. A 101, 13091–13095. [PubMed: 15322270]

Rauschecker JP (1998). Cortical processing of complex sounds. Curr. Opin. Neurobiol 8, 516–521. [PubMed: 9751652]

Rauschecker JP (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear. Res 271, 16–25. [PubMed: 20850511]

Rauschecker JP, and Scott SK (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci 12, 718–724. [PubMed: 19471271]

Rauschecker JP, and Tian B (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A 97, 11800–11806. [PubMed: 11050212]

Rauschecker JP, and Tian B (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol 91, 2578–2589. [PubMed: 15136602]

Rauschecker AM, Pringle A, and Watkins KE (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. Hum. Brain Mapp 29, 1231–1242. [PubMed: 17948887]

Rauschecker JP, Tian B, and Hauser M (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268, 111–114. [PubMed: 7701330]

Repp BH (1984). Categorical Perception: Issues, Methods, Findings In Speech and Language, Advances in Basic Research and Practice, Vol. 10, Lass NJ, ed. (Academic Press), pp. 244–335.

Riesenhuber M, and Poggio T (2002). Neural mechanisms of object recognition. Curr. Opin. Neurobiol 12, 162–168. [PubMed: 12015232]

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, and Rauschecker JP (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci 2, 1131–1136. [PubMed: 10570492]

Roy JE, Riesenhuber M, Poggio T, and Miller EK (2010). Prefrontal Cortex Activity during Flexible Categorization. J. Neurosci 30, 8519–8528. [PubMed: 20573899]

Saxe R, Brett M, and Kanwisher N (2006). Divide and conquer: a defense of functional localizers. NeuroImage 30, 1088–1099. [PubMed: 16635578]

Scholl CA, Jiang X, Martin JG, and Riesenhuber M (2014). Time course of shape and category selectivity revealed by EEG rapid adaptation. J. Cogn. Neurosci 26, 408–421. [PubMed: 24001003]

Scott SK (2005). Auditory processing--speech, space and auditory objects. Curr. Opin. Neurobiol 15, 197–201. [PubMed: 15831402]

Serre T, Wolf L, Bileschi S, Riesenhuber M, and Poggio T (2007). Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell 29, 411–426. [PubMed: 17224612]

Shannon RV, Jensvold A, Padilla M, Robert ME, and Wang X (1999). Consonant recordings for speech testing. J. Acoust. Soc. Am 106, L71–4. [PubMed: 10615713]

Sunaert S, Van Hecke P, Marchal G, and Orban GA (2000). Attention to speed of motion, speed discrimination, and task difficulty: an fMRI study. NeuroImage 11, 612–623. [PubMed: 10860790]

Tian B, and Rauschecker JP (2004). Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol 92, 2993–3013. [PubMed: 15486426]

Tian B, Reser D, Durham A, Kustov A, and Rauschecker JP (2001). Functional specialization in rhesus monkey auditory cortex. Science 292, 290–293. [PubMed: 11303104]

Tian B, Ku mierek P, and Rauschecker JP (2013). Analogues of simple and complex cells in rhesus monkey auditory cortex. Proc. Natl. Acad. Sci. U. S. A 110, 7892–7897. [PubMed: 23610391]

Tsunada J, and Cohen YE (2014). Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. Front. Neurosci 8, 161. [PubMed: 24987324]

Tsunada J, Lee JH, and Cohen YE (2011). Representation of speech categories in the primate auditory cortex. J. Neurophysiol 105, 2634–2646. [PubMed: 21346209]

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, and Joliot M (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–289. [PubMed: 11771995]

Ungerleider LG, and Haxby JV (1994). "What" and "where" in the human brain. Curr. Opin. Neurobiol 4, 157–165. [PubMed: 8038571]

Whitfield-Gabrieli S, and Nieto-Castanon A (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect 2, 125–141. [PubMed: 22642651]

Williams MA, Dang S, and Kanwisher NG (2007). Only some spatial patterns of fMRI response are read out in task performance. Nat. Neurosci 10, 685–686. [PubMed: 17486103]

Yin P, Fritz JB, and Shamma SA (2014). Rapid spectrotemporal plasticity in primary auditory cortex during behavior. J. Neurosci. Off. J. Soc. Neurosci 34, 4396–4408.

Zatorre RJ, Bouffard M, and Belin P (2004). Sensitivity to auditory object features in human temporal neocortex. J. Neurosci. Off. J. Soc. Neurosci 24, 3637–3642.
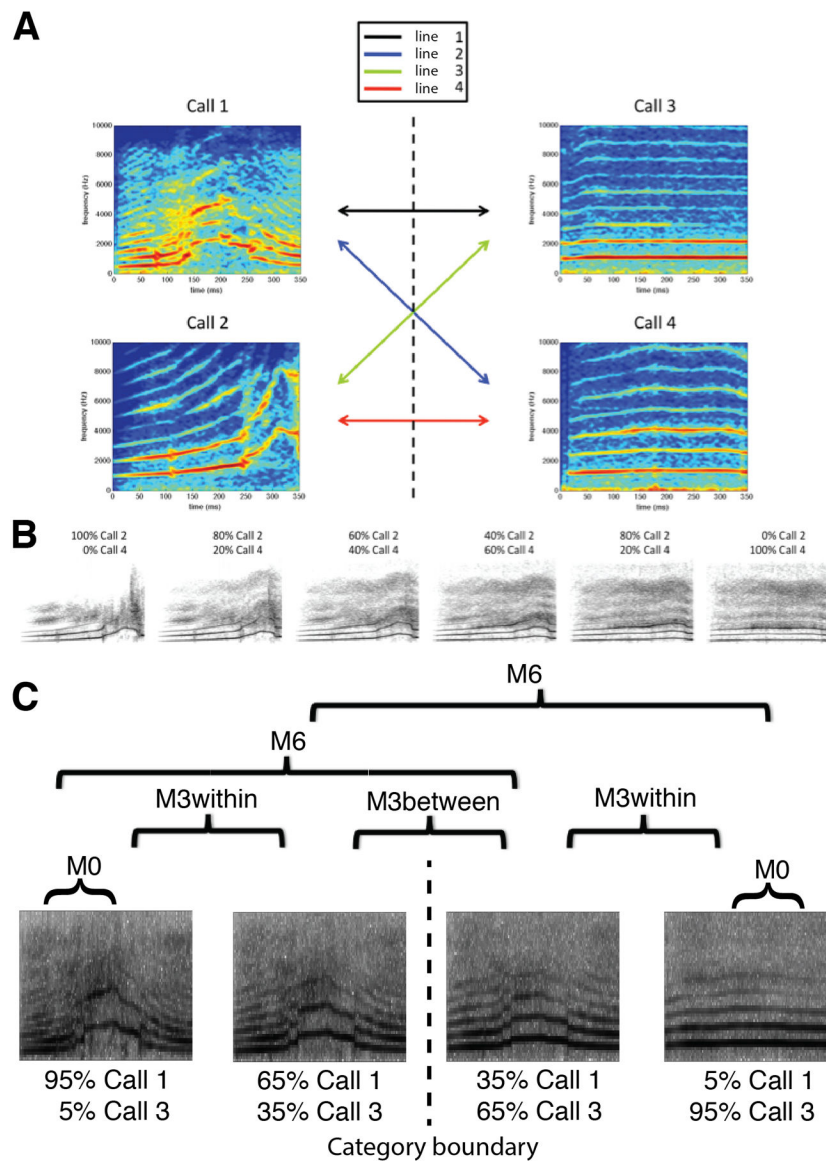
Figure 1. Prototype stimuli and morph space.

(A) Multi-exemplar audio morphing. Stimulus morphing using the software STRAIGHT (Kawahara and Matsui, 2003) was previously restricted to those points along a single vector between two prototype sounds. We have extended this method to allow generation of stimuli throughout an arbitrary morphing "space" defined by a given set of sounds. Here, a morph space is defined by four monkey calls (two harmonic arches, and two coos). By manipulating the degree to which each stimulus contributes to a morph, stimuli can be generated at any location throughout this auditory morph space. Individual morph lines crossing the defined category boundary are generated for post-training identification testing. (B) Sample morph line. This is an example morph line between one harmonic arch and one coo. The category boundary that subjects learn is located between the third and fourth sound. There were four "cross-category" morph lines, which were used in post-training psychophysics experiments. Two of them, lines 1 and 4, with best performance (based on

piloting data) were used in fMRI experiments. (C) Stimuli for fMRI conditions. Stimuli were generated along two independent morph lines (Line 1 and 4, Fig. 1A) using sound "morphing" software (STRAIGHT toolbox for MATLAB), and a total of 14 stimulus pairs were constructed to probe acoustic and category selectivity. Pairs were arranged to align with the category boundary at the 50% morph-point between the two prototypes. Four different conditions were tested: M0, $M3_{within}$, $M3_{between}$, and M6.
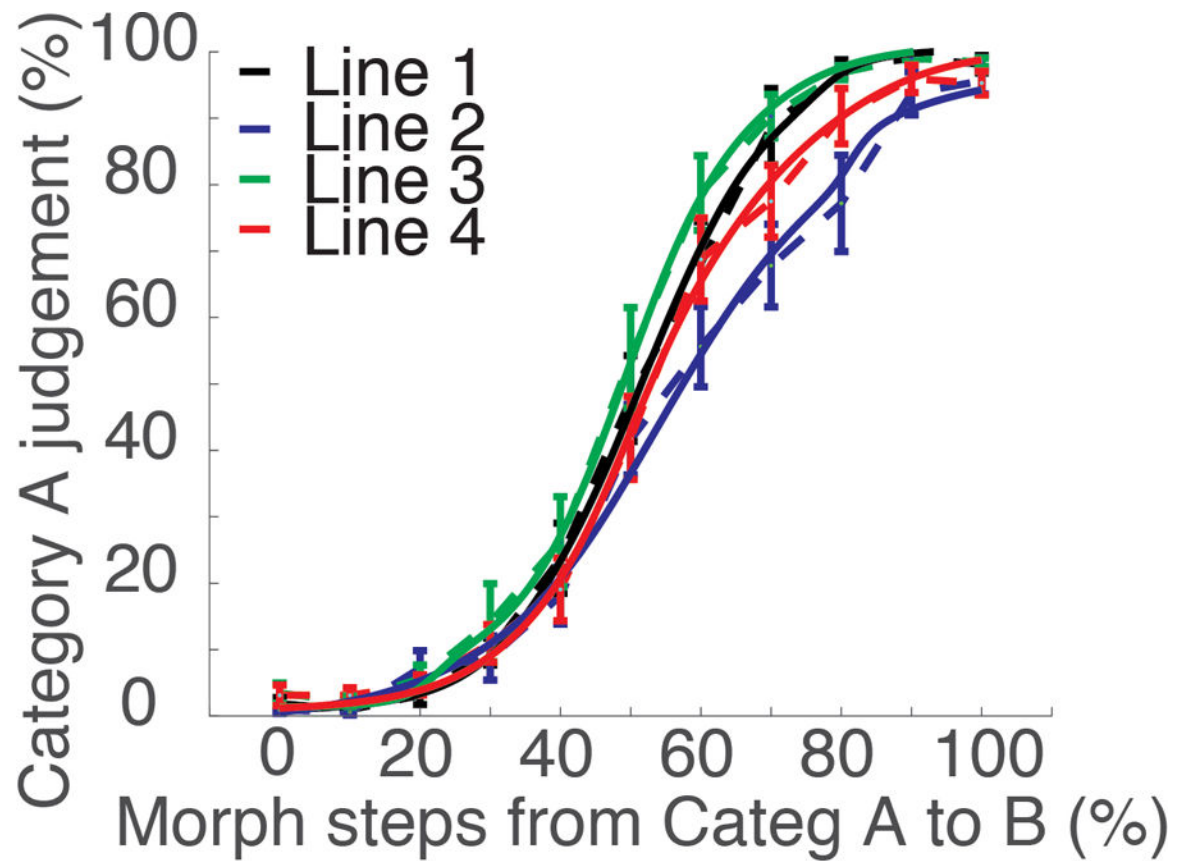
Figure 2. Behavioral performance.
Behavioral categorization performance outside the scanner (n=16). After training and prior
to post-training MRI scanning, we measured subjects' ability to categorize the morph stimuli
along four morph lines using stimuli varying by 10% steps (shown in dashed lines). We then
fitted the categorization performance to a sigmoid function (Eq. 1). The fitting curves of all
morph lines and all subjects are shown as solid lines. Error bars represent SEM.
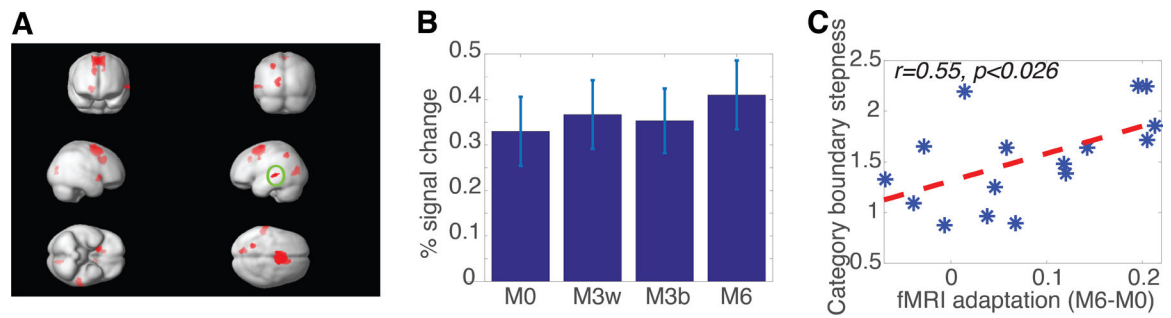
Figure 3. fMRI-RA reveals increased neural selectivity for acoustic features in the left auditory cortex.

(A) fMRI adaptation results using the contrast of M6 > M0, at a threshold of p<0.01, uncorrected, at least 50 contiguous voxels, masked by Auditory Stimuli > Silence (p<0.001, uncorrected). The left auditory ROI (MNI: −56 −36 2, circled as O in the figure) was used for the analyses in (B) and (C). (B) fMRI-responses in the left auditory ROI during the post-training bottom-up scans. (C) The steepness of the category boundary in behavioral testing conducted outside of the scanner (measured as the α values in Equation 1) correlated with fMRI adaptation (measured as the difference between M6 and M0 condition), an indirect measure of neural tuning. Error bars represent SEM.
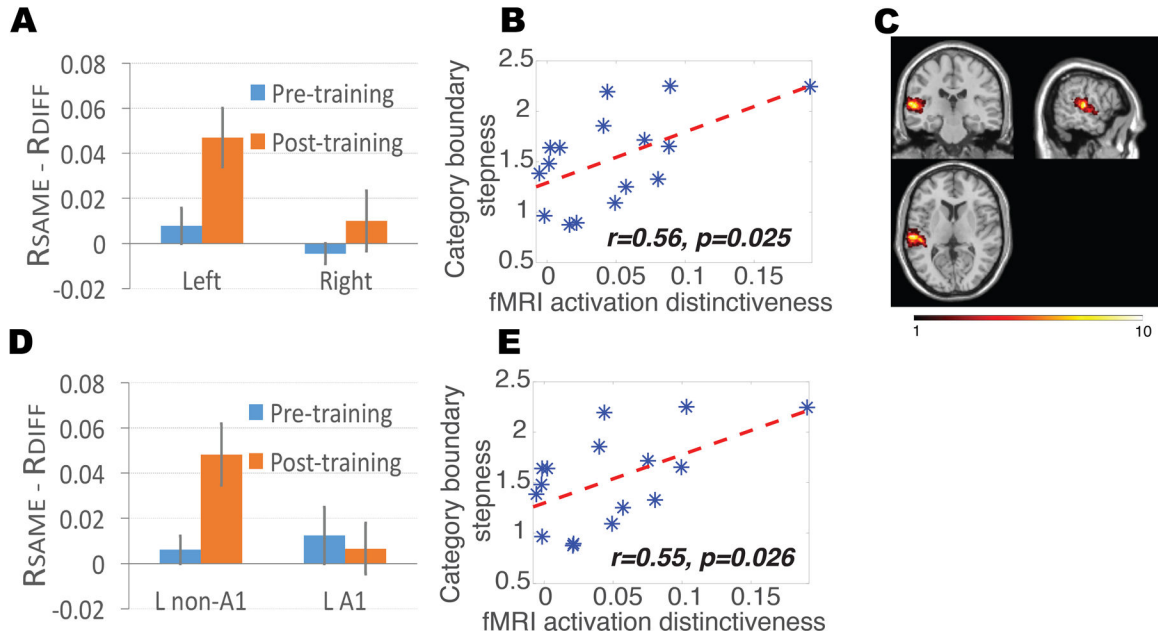
Figure 4. MVPA reveals increased neural selectivity for acoustic features in the left auditory cortex.

(A) MVPA revealed an increase in BOLD response pattern distinctiveness in the left, but not the right STC (corresponding to auditory cortex), following training ($p<0.039$ and $p>0.360$, respectively, paired t-test). (B) The steepness of the category boundary in behavioral testing (measured as the α values in Equation 1) correlated with MVPA-estimated neural tuning in the left auditory cortex, in line with fMRI-RA results in Figure 3C. (C) Heat map of the voxels used in the MVPA analyses shown in Fig. 4A and 4B. Voxel color indicates number of participants whose set of most significant voxels included that particular voxel. (D) MVPA revealed an increase in BOLD response pattern distinctiveness in the left non-primary (L non-A1), but not the left primary (L A1) auditory cortex, following training ($p<0.035$ and $p>0.781$, respectively, paired t-test). (E) The steepness of the category boundary in behavioral testing (measured as the α values in Equation 1) correlated with MVPA-estimated neural tuning selectivity in the left auditory cortex after excluding voxels in the primary auditory cortex (L non-A1). Error bars represent SEM.
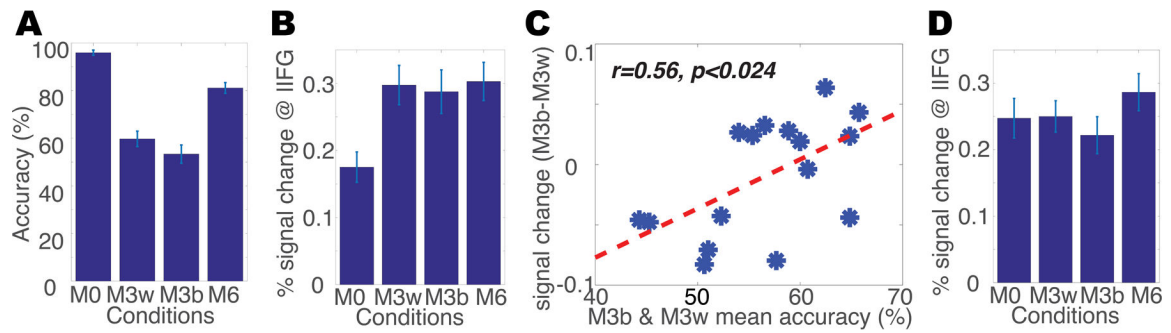
Figure 5. FMRI-RA reveals training-induced neural plasticity for category selectivity.
(A) Behavioral performance of category judgment inside scanner. (B) fMRI responses in left dorsolateral prefrontal cortex (MNI: −48 8 18, defined via the contrast of M6>M0, see Table 2) during the top-down scans. (C) The significant correlation between fMRI adaptation (M3$_{between}$ *minus* M3$_{within}$) in the same left prefrontal ROI as in 5B and the average performance of M3$_{between}$ and M3$_{within}$. (D) fMRI responses in the same left prefrontal ROI during the bottom-up scans. Error bars represent SEM.
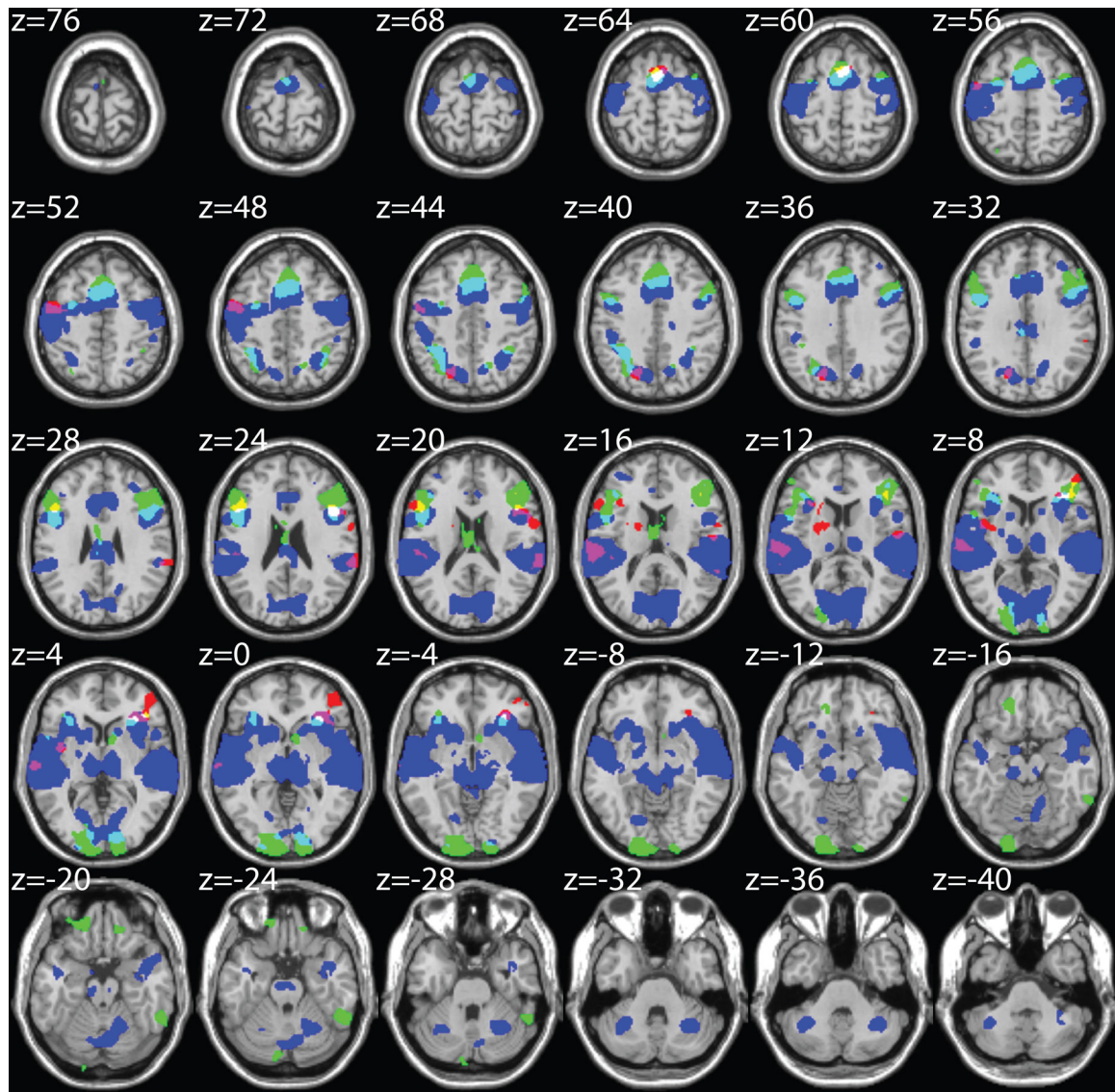
Figure 6. MVPA reveals training-induced neural plasticity for category selectivity.
(A) Category-selective brain regions revealed by search-light MVPA analysis. (B) fMRI response profiles during the top-down categorization scans in the four ROIs in (A). (C) Overlap between regions showing category selectivity (from fMRI-RA and MVPA analyses) and regions functionally connected to the left auditory cortex. (Red) MVPA-identified category selective regions using a searchlight approach (7x7x7 voxels), p<0.01 (uncorrected, 20 contiguous voxels; for illustration purpose, less stringent thresholds are used here). (Green) Category-selective regions as defined by fMRI-RA with the contrast of M6 > M0 (p<0.01, uncorrected, at least 20 contiguous voxels), masked by Stimuli > Silence (p<0.005) from fMRI Scan 3. (Blue) Brain regions functionally connected to left auditory cortex (p<0.001, uncorrected, at least 20 contiguous voxels). Regions of overlap between the three analyses. (Yellow) MVPA & fMRI-RA. (Violet) MVPA & functional connectivity. (Cyan) fMRI-RA & functional connectivity. (White) MVPA & fMRI-RA & functional connectivity.

The MNI coordinates of the peak of each cluster are shown next to the clusters. Error bars represent SEM.
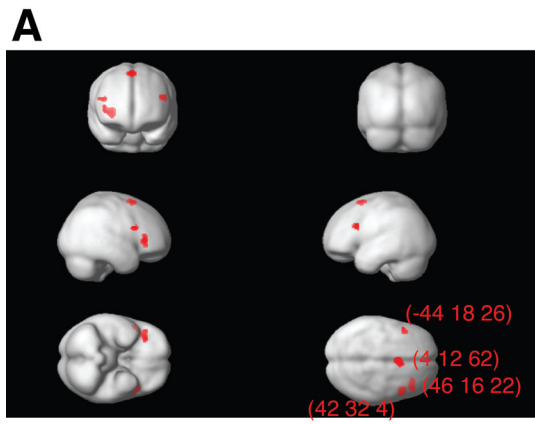
A

(-44 18 26)

(4 12 62)

(46 16 22)

(42 32 4)

Figure 7.

B

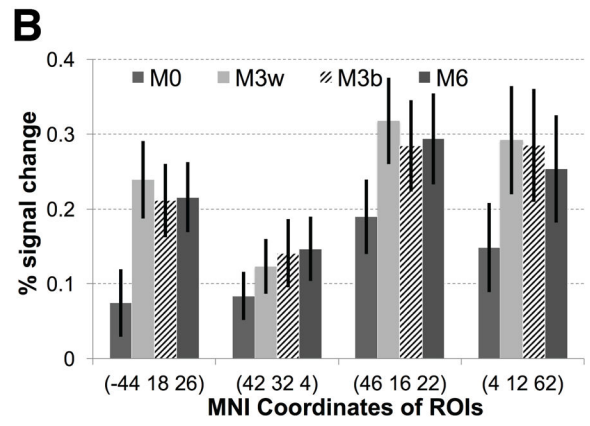**Table 1.**

FMRI-RA results of Experiment 1 and 2. Brain regions with increased release from adaptation after training, using a threshold of p<0.01, uncorrected, 50 contiguous voxels, masked by Stimuli > silence (p<0.001, uncorrected).

Region	mm³	Z_max	MNI Coordinates		
			X	Y	Z
M6 > M0 x Post > Pre					
L/R SMA / L Mid Cingulum/	3472	3.78	8	14	54
		2.98	−8	−6	52
L Supramarginal/L Sup Temporal	936	3.34	−62	−16	16
L Mid/Sup Temporal	**864**	**3.14**	**−56**	**−36**	**2**
L Sup Parietal	560	3.03	−20	−60	46
		2.63	−18	−62	56
L Postcentral	480	2.78	−54	−18	58
M6/M3b/M3w > M0 & Post > Pre					
L Supramarginal/L Sup Temporal	1096	3.36	−64	−18	16
L Mid/Sup Temporal *	**192**	**2.91**	**−70**	**−24**	**−4**
		2.49	**−64**	**−32**	**0**
M6 > M0 & Pre > Post					
None					
M6/M3b/M3w > M0 & Pre > Post					
None					
M6 > M0 (Post)					
L/R SMA / LR Sup Frontal/L Mid Cingulum/	11640	4.81	6	18	56
		3.89	−2	16	56
		3.83	10	−4	58
L Inf Parietal	976	3.57	−24	−56	46
R Pallidum/Putamen/Caudate	1224	3.23	14	10	−10
		3.06	20	16	−2
L Mid/Sup Temporal	**1080**	**3.15**	**−56**	**−36**	**2**
L Cuneus/Lingual	1784	3.15	−18	−80	14
		3.12	−14	−76	4
M0 > M6 (Post)					
None					

*
24 voxels.

Table 2.

FMRI-RA results of Experiment 3. Brain regions with categorical selectivity to the trained auditory stimuli, using the contrast of M6 > M0 (p<0.001, uncorrected, at least 50 contiguous voxels), or the contrast of M6& M3$_{between}$ > M0&M3$_{within}$ (p<0.005, uncorrected, at least 50 contiguous voxels), masked by Stimuli > Silence (p<0.005, uncorrected).

Region	mm³	Z$_{max}$	MNI Coordinates		
			X	Y	Z
M6>M0					
L Lingual/Mid Occipital	992	4.71	−16	−94	2
L/R SMA/Sup Med Frontal	7040	4.67	8	20	50
		3.91	−2	22	44
		3.86	−8	24	38
L Mid/Inf Frontal	**2168**	**4.17**	**−48**	**8**	**18**
		3.80	**−46**	**4**	**34**
L Inf/Sup Parietal	856	3.72	−32	−46	42
		3.45	−28	−56	42
R Inf Frontal	512	3.63	46	12	36
R Inf/Mid Frontal	640	3.45	56	32	26
		3.33	44	26	20
M6 & M3$_{between}$ > M3$_{within}$ & M0					
L Inf Frontal	**848**	**4.08**	**−30**	**28**	**2**
L Inf Frontal	**1328**	**3.58**	**−50**	**8**	**20**
		2.87	**−48**	**4**	**32**
L/R SMA/Sup Med Frontal	6768	3.49	8	16	52
		3.39	−2	16	52
		3.38	−6	2	60
R Cuneus	764	3.17	16	−98	2
		2.93	18	−86	12
		2.89	16	−88	4
R Inf/Mid Frontal	400	3.08	42	28	26
M0 > M6					
R Sup Temporal	472	4.21	48	−34	18
R Sup Temporal/ R Postcentral	1296	3.63	58	−14	4
		3.32	66	−18	16
		3.30	50	−14	12
L Heschl/Insula	432	3.61	−40	−16	6

M3$_{within}$ & M0 > M6 & M3$_{between}$

Region	mm³	Z_{max}	MNI Coordinates		
			X	Y	Z
R Inf/Mid/Sup Temporal	3600	3.95	54	−14	−2
		3.19	54	−2	−12
R Sup Temporal/R		2.89	52	−10	−14
Supramarginal	920	3.27	62	−36	26
		3.24	70	−34	18
L Sup Temporal	904	3.14	−52	−20	10

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
SPM8	The Wellcome Trust Centre for Neuroimaging	http://www.fil.ion.ucl.ac.uk/spm/software/spm8/
MATLAB R2015B	The MathWorks Inc.	http://www.mathworks.com
MarsBaR 0.44	Matthew Brett	https://sourceforge.net/projects/marsbar/files/
SPSS Version 24	IBM	http://www-01.ibm.com/support/docview.wss?uid=swg24041224