Representation of sound categories in auditory cortical maps

Frank Guenther, Alfonso Nieto-Castanon, Satrajit Ghosh and Jason Tourville

September, 2002

Technical Report CAS/CNS-2002-009

Permission to copy without fee all or part of this material is granted provided that: 1. The copies are not made or distributed for direct commercial advantage; 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and/or special permission.

Copyright © 2002

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA 02215
Representation of Sound Categories in Auditory Cortical Maps

Frank H. Guenther*,†, Alfonso Nieto-Castanon*, Satrajit S. Ghosh* & Jason A. Tourville*‡

* Department of Cognitive and Neural Systems, Boston University, 677 Beacon St, Boston, MA 02215, USA
† Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-413, 50 Vassar St, Cambridge, MA 02139, USA
‡ Center for Morphometric Analysis, Massachusetts General Hospital, Building 149, 13th Street, Charlestown, MA 02129, USA

Keywords: speech, cortex, auditory comprehension

ABSTRACT

We used functional magnetic resonance imaging (fMRI) to investigate the representation of sound categories in human auditory cortex. Experiment 1 investigated the representation of prototypical and non-prototypical examples of a vowel sound. Listening to prototypical examples of a vowel resulted in less auditory cortical activation than listening to non-prototypical examples. Experiments 2 and 3 investigated the effects of categorization training and discrimination training with novel non-speech sounds on auditory cortical representations. The two training tasks were shown to have opposite effects on the auditory cortical representation of sounds experienced during training: discrimination training led to an increase in the amount of activation caused by the training stimuli, whereas categorization training led to decreased activation. These results indicate that the brain efficiently shifts neural resources away from regions of acoustic space where discrimination between sounds is not behaviorally important (e.g., near the center of a sound category) and toward regions where accurate discrimination is needed. The results also provide a straightforward neural account of learned aspects of categorical perception: sounds from the center of a category are more difficult to discriminate from each other than sounds near category boundaries because they are represented by fewer cells in the auditory cortical areas.
1 INTRODUCTION

Our ability to sort stimuli into behaviorally relevant categories is a central aspect of human experience. The perception of a continuously varying speech signal as a discrete set of phonemes, syllables, and words is a vivid example of this phenomenon. The process of categorizing speech stimuli is aided by the fact that our perceptual spaces are warped near speech sound categories, a phenomenon often referred to as categorical perception (e.g., Liberman et al., 1957, 1961; Eimas, 1963). Some aspects of this warping are learned. For example, Kuhl and colleagues describe a language-specific “perceptual magnet effect” in which two prototypical examples of a vowel from an individual’s native language (i.e., two sounds judged to be good examples of the vowel) are harder to discriminate from each other than two non-prototypical examples that are near a category boundary (Kuhl, 1991; Kuhl et al, 1992). This effect has been the focus of numerous studies over the past decade (Kuhl, 1991; Kuhl et al., 1992; Iverson and Kuhl, 1995, 1996; Sussman and Lauckner-Morano, 1995; Guenther and Gjaja, 1996; Aaltonen et al., 1997; Lively and Pisoni, 1997; Lotto, Kluender, and Holt, 1998; Sharma and Dorman, 1998; Diesch et al., 1999; Frieda et al., 1999; Guenther et al., 1999), and similar effects have been reported for a variety of non-speech stimuli, including melodic musical intervals (Burns and Ward, 1978), simple visual shapes (Lane, 1965; Goldstone, 1994), and faces (Beale and Keil, 1995).

Guenther et al. (1999) hypothesized that this warping of perceptual space arises from reorganization of the auditory cortical map during category learning. Figure 1 contrasts the hypothesized effects of discrimination training and categorization training on the auditory cortical representation of sounds. Recanzone, Schreiner, and Merzenich (1993) demonstrated that training monkeys to discriminate tones from a particular frequency range leads to an increase

![Discrimination Training](image1)

![Categorization Training](image2)

**FIGURE 1.** Hypothesized effects of discrimination training and categorization training on neural maps in the auditory cortical areas. The x and y axes of each plot correspond to two auditory dimensions, such as the first two formant frequencies for vowel sounds. For the same Gaussian distribution of training stimuli, category learning is hypothesized to lead to a decrease in the size of the neural representation for stimuli from the center of the distribution, while discrimination training is hypothesized to lead to an increase in the size of the neural representation for the central stimuli.
in the size of the auditory cortical representation of this range (schematized in the left half of Figure 1) and a concomitant improvement in the discriminability of these tones. Analogous results have been reported for visual stimuli in inferotemporal cortex (Kobatake, Wang, and Tanaka, 1998) and tactile stimuli in somatosensory cortex (Recanzone et al., 1992). The right half of Figure 1 illustrates the hypothesized effects of categorization training on auditory cortical maps. In psychophysical experiments with human subjects (Guenther et al., 1999), we showed that, whereas discrimination training with sounds from a particular frequency range leads to an increase in discriminability for these stimuli, categorization training with the same sounds leads to a decrease in their discriminability. We hypothesized that this decrease was due to a decrease in the size of the cortical representation of the training stimuli, in direct contrast to the increase seen with discrimination training.

To test this account of the effects of category learning on auditory perception and cortical maps, functional magnetic resonance imaging (fMRI) was used to measure the amount of cortical activation, as evidenced by increases in regional cerebral blood flow, in three experiments involving speech and non-speech auditory stimuli.

2 METHODS

Subjects

Nine right-handed native adult speakers of American English participated in Experiment 1. Seven subjects participated in Experiment 2, and seven additional subjects participated in Experiment 3. All subjects had no history of language or other neurological disorders. The experimental protocol was approved by the Boston University and Massachusetts General Hospital committees on human subjects. Informed consent was obtained from all subjects.

Acoustic stimuli

In Experiment 1, subjects were stimulated binaurally with two synthetic vowel stimuli, a prototypical /i/ stimulus and a non-prototypical /i/ stimulus, presented in separate blocks. Stimuli were generated using the Sensyn speech synthesis software (Sensimetrics Corporation) with parameters chosen to match synthetic vowels used to demonstrate the perceptual magnet effect psychophysically (Kuhl, 1991). Stimuli were presented in a block paradigm consisting of alternating 30-second blocks of prototypical vowels and non-prototypical vowels separated by 30-second silent intervals for a total run length of 5-1/2 minutes. Subjects were told to attend to the stimuli by listening for differences from sound to sound. Between four and eight runs were conducted for each subject.

In Experiments 2 and 3, stimuli were narrow-band filtered samples of white noise with different center frequencies. Center frequencies ranged between 1000 Hz and 3500 Hz. Two fMRI sessions were completed for each subject, one before and one after a week of training sessions involving the acoustic stimuli. Stimuli were presented in a block paradigm consisting of alternating 30-second blocks of stimuli (800ms stimulus length, 2000ms ISI) belonging to either a training region or control region of frequency space. Sounds from the control region were not encountered during the training sessions. Total run length was 5-1/2 minutes. Subjects were told to attend to the stimuli by listening for differences from sound to sound. Four to eight runs were conducted for each subject.
Categorization training

In Experiment 2, imaging sessions were performed before and after subjects performed five sessions of a categorization training task (see Guenther et al., 1999 for details) over the course of one week. Training sessions lasted approximately 45 minutes. Subjects were trained to identify sounds from a “category” consisting of seven sounds from a small region of frequency space. During a training trial, subjects had to identify which sound from a short list of sounds was from the category. The length of the sound list was 2, 3, or 4 sounds, with only one sound from the category in each list. Each session involved 300 training trials. The subjects’ ability to identify sounds from the category generally increased during the week of training, while their ability to discriminate between sounds within the category, as measured by $d'$, decreased.

Discrimination training

In Experiment 3, a discrimination training task (see Guenther et al., 1999 for details) was used in place of the categorization training task. This experiment involved the same distribution of training stimuli used in Experiment 2. Subjects were asked to respond “same” or “different” when presented with a pair of training stimuli chosen from the small region of frequency space corresponding to the category in Experiment 2. The subjects’ ability to discriminate between the sounds, as measured by $d'$, generally increased during the week of training.

Image collection

Data for Experiment 1 were obtained using a 1.5 Tesla General Electric Signa imager. Data for Experiments 2 and 3 were obtained using a 1.5T Siemens scanner. Imaging sessions began with the acquisition of anatomical images that were later used to parcellate the regions of interest. T2-weighted functional images encompassing the entire peri-sylvian cortex were acquired using an asymmetric spin-echo echo-planar imaging sequence (TR=25ms, TE=70ms, TR=2s, matrix size 64x64, 5mm thick contiguous slices with in-plane resolution=3.1x3.1mm).

Parcellation of cortical regions

Functional data analysis was carried out on ten peri-sylvian cortical regions of interest (ROIs): Heschl’s gyrus (HG), planum temporale (PT), planum polare (PP), anterior and posterior portions of the superior temporal gyrus (T1a, T1p), anterior and posterior middle temporal gyrus (T2a, T2p), parietal operculum (PO), and anterior and posterior supramarginal gyrus (SGa, SGp). HG includes primary auditory cortex, and PT, PP and T1 are commonly considered to be auditory association areas (Gloor, 1997). PO, SG, and T2 are multimodal areas that have been implicated in some speech and language tasks (Caplan, Gow, and Makris, 1995; Mazoyer et al., 1993). For each subject, sagittal anatomical images were positionally normalized relative to the anterior-posterior commissure line and the interhemispheric fissure and then resliced into coronal images for parcellation. The cerebral cortex of each brain was identified on relevant coronal slices and subdivided into regions of interest (ROIs) spanning the peri-sylvian cortex based on visible anatomic landmarks and fissures of the individual brains (Caviness et al., 1996).
Individual functional runs were temporally realigned and coregistered to the structural T1 series using the SPM99 software package. Preprocessing was applied to each subject’s functional series separately for each ROI and included estimation of noise autocorrelation and prewhitening, band-pass filtering, and data reduction (principal components analysis). The resulting functional components for each ROI were then fitted across all subjects using a multivariate general linear model (Ghosh et al., 2001). Regressors were temporal series defined by half-sine shaped blocks following the experimental protocol convolved with a standard hemodynamic response function. The temporal series defined by the estimated motion parameters were also included as regressors in the linear model. Hypothesis testing for each ROI was performed using a Likelihood Ratio Test.

3 RESULTS

Speech sounds

In Experiment 1, subjects listened to prototypical (good) and non-prototypical (poor) examples of the American English vowel /i/. This phoneme has been shown to exhibit worse discriminability near category prototypes than near category boundaries in numerous studies (Kuhl, 1991; Kuhl et al., 1992; Sussman and Lauckner-Morano, 1995; Lotto, Kluender, and Holt, 1998). Figure 2 shows the averaged activations in the prototype and non-prototype conditions (compared to a baseline silent condition) projected onto the temporal lobe ROIs. Higher total activation across the auditory cortical areas for the non-prototype condition is clearly evident, particularly in right hemisphere regions HG and PT.

Figure 3 details the results of all three experiments for each ROI individually (top portion of figure) and averaged across the primary and higher-order auditory cortical areas (bottom). In Experiment 1 (Figure 3, left column), significantly less activation was found for prototypical vowels than for non-prototypical vowels in right PT and right SGp. PT has been hypothesized to play a role in phoneme and pitch perception (Liegeois-Chauvel et al., 1999), and SGp has been implicated in phonetic discrimination and identification (Caplan, Gow, and Makris, 1995). The prototypical vowel also induced less activation across the auditory cortical areas as a whole, although this difference fell short of statistical significance (p=0.081).

The results of Experiment 1 support a simple and straightforward explanation for the perceptual magnet effect of speech perception: prototypical examples of a phoneme are more difficult to discriminate from each other than non-prototypical examples because they are processed by a smaller neural representation in the auditory cortical areas involved in sound discrimination. These smaller neural representations are more susceptible to the effects of noisy processing in individual neurons (Bauer, Der, and Herrmann, 1996; Guenther et al., 1999; see Discussion section).

Non-speech sounds

Experiment 2 examined the effects of categorization training on non-speech auditory maps by imaging subjects before and after one week of training on a category learning task involving non-speech auditory stimuli. Experiment 3 was the same as Experiment 2 except discrimination training was used instead of categorization training in order to identify differences
in the effects of these tasks on auditory cortical maps. Figure 4 shows post-training activations while listening to the training sounds (as compared to listening to control sounds not encountered during training) in Experiments 2 and 3. The cortical surface has been inflated in this figure to expose cortical areas buried in the sylvian fissure. Learning of a sound category results in a reduction of activation for prototypical members of the category in most of the auditory cortical areas. Discrimination training, in contrast, leads to an increase in activation in the auditory cortical areas.

Comparing post-training activations to pre-training activations in Experiment 2 (Figure 3, second column) reveals a significant decrease in activation across the auditory cortical areas for category-prototypical sounds, as was the case for speech sounds in Experiment 1. The biggest
FIGURE 3. Comparison of results across experiments for each ROI individually (top panels) and for the auditory cortical areas considered as a whole (bottom panels). The first column shows the difference in activation between prototypical and non-prototypical examples of the vowel /i/. In general, less activation is found for the prototypical vowel. The second column compares post-training activation to pre-training activation for Experiment 2. Categorization training leads to a decrease in the size of the auditory cortical representation of stimuli from within a newly learned category as compared to control stimuli not encountered during training. The third column compares post- and pre-training activations in Experiment 3. In contrast to categorization training, discrimination training leads to an increase in the auditory cortical activations for the training stimuli. The fourth column shows the difference between the change in activation caused by categorization training (Experiment 2) and the change in activation caused by discrimination training (Experiment 3). The differential effect of the two training types is statistically significant in left HG and in the auditory cortical areas considered as a whole.

activation decreases in Experiment 2 were seen in PO bilaterally, as compared to right PT and SGp in Experiment 1. Together these results suggest that (i) different sound types may be
FIGURE 4. Results of Experiments 2 and 3 displayed on an inflated cortical surface. Upper panels: Effects of learning a novel auditory category (Experiment 2). Plots show difference between post- and pre-training activations for the training stimuli (as compared to control stimuli). Blue activations indicate areas in which training has led to a relative decrease in activation for stimuli from within the newly learned category. Lower panels: Effects of discrimination training (Experiment 3). Plots show difference between post- and pre-training activations for the training stimuli (as compared to control stimuli). In contrast to category learning, discrimination training leads to a relative increase of activation (red activations) for the training stimuli in the peri-sylvian cortical areas. HG = Heschl’s gyrus; PT = planum temporale; PP = planum polare; T1a, T1p = anterior/posterior superior temporal gyrus; T2a, T2p = anterior/posterior middle temporal gyrus; PO = parietal operculum; SGa,SGr = anterior/posterior supramarginal gyrus.

represented in different parts of peri-sylvian cortex, and (ii) though not usually considered to be auditory cortex, parietal cortical areas PO and SG play a role in sound representation (see also Caplan, Gow, and Makris, 1995).

After discrimination training (Experiment 3), a significant increase in activation was found in left HG and in the auditory cortical areas considered as a whole, contrasting with the decrease in activation seen for the same stimuli after categorization training in Experiment 2. As shown in the rightmost column of Figure 3, statistically significant differences in the effects of the different training tasks were found in left HG and in the auditory cortical areas considered as a whole.

In all three experiments, the middle temporal gyrus (T2a, T2p) activations differed from those in the auditory cortical areas of the superior temporal gyrus (see Figure 3). In particular, category prototypical examples tended to cause more activation in T2 than non-prototypical sounds, and discrimination training led to a decrease in the amount of T2 activation. Although
they will not be discussed in further detail here since our focus is on the auditory cortical areas, these results are suggestive of a more categorical mode of processing in T2 as compared to the auditory cortical areas of the superior temporal gyrus and temporal plane.

4 DISCUSSION

Taken together, these experiments indicate that learning a sound category leads to a decrease in auditory cortical activity while processing prototypical members of the category. It has been demonstrated recently that the fMRI signal can be approximated by a delayed and temporally smoothed version of neural activity as measured by local field potentials (Logothetis et al., 2001). A natural interpretation of the current results, therefore, is that fewer cells in the auditory cortical maps have receptive fields near the center of learned categories as compared to parts of auditory space where sound discrimination is more important behaviorally, as in the model schematized in Figure 1. Since it has long been believed that smaller cortical representations lead to worse discriminability of the represented stimuli, this interpretation is also in accord with psychophysical experiments showing reduced discriminability for category prototypes (Liberman et al., 1957, 1961; Eimas, 1963; Lane, 1965; Burns and Ward, 1978; Kuhl, 1991; Kuhl et al., 1992; Goldstone, 1994; Beale and Keil, 1995; Iverson and Kuhl, 1995, 1996; Sussman and Lauckner-Morano, 1995; Guenther et al., 1999). In fact, the model schematized in Figure 1 constitutes an account of the neural bases of these psychophysical findings. Since longer periods of activation will also result in a higher fMRI signal, an alternative possibility is that the length of time that auditory cortical cells remain active after stimulus presentation is smaller for category prototypes than for non-prototypes. In other words, the brain may be conserving neural resources by reducing the processing time for category prototypes, rather than by reducing the number of cells representing the category prototypes. Differentiating between these possibilities will require measurement of the time course of activity in the auditory cortical areas with techniques that afford better temporal resolution than fMRI, such as electrophysiology in animals or fMRI combined with electroencephalography or magnetoencephalography in humans.

Because decreased auditory cortical activity for category prototypes is found for both speech and non-speech auditory stimuli, it is tempting to attribute examples of categorical perception in other sensory modalities (Lane, 1965; Burns and Ward, 1978; Goldstone, 1994; Beale and Keil, 1995) to the same property of sensory map formation identified in Experiments 1 and 2. In keeping with this view are results indicating less activation in visual cortical areas BA 17/18 for categorically processed visual stimuli (Reber, Stark, and Squire, 1998). This property may have evolved as a means for efficiently allocating neural resources for speech perception and other categorical processing tasks. When perceiving speech sounds, it is far more important to detect between-category differences than within-category differences. The learning processes identified here allocate more neural resources to “ambiguous” sensory stimuli lying near category boundaries than to stimuli from near the center of a category, thus effectively positioning these resources where they will be most useful for parsing the speech signal into distinct sound categories and words.

In addition to showing that discrimination training with auditory stimuli leads to an increase in activation for the training stimuli, Experiment 3 also served as a test of an alternative hypothesis regarding the reduced activations seen for category prototypes in Experiment 1, namely that the reduced activation is simply due to greater familiarity with these stimuli rather than from categorization training per se. Although the same distribution of training stimuli was
used in Experiments 2 and 3, opposite effects were seen for the post-training brain activations, indicating that it is the type of training, not just the amount of experience with the training stimuli, that determines what happens to the auditory cortical representation of these stimuli.

ACKNOWLEDGMENTS

We thank Julie Goodman, Jennifer Holmes, Fatima Husain, Bruce Rosen, David Kennedy, Andrew Worth, Mukund Balasubramanian, and Mark Shiffer for their assistance. We also thank the Massachusetts General Hospital NMR Center for the use of their facilities. Supported by the National Institute on Deafness and Other Communication Disorders (R01 DC02852, F. Guenther PI).

REFERENCES


