

Representation of Sound Categories in Auditory Cortical Maps

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

Journal of Speech, Language, and Hearing Research, 2004, 47(1), pp. 46-57

* *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 St., Charlestown, MA 02129, USA*

Corresponding author: Prof. Frank Guenther, Boston University, 677 Beacon Street, Boston, MA 02115. Phone: 617-353-5765; Fax: 617-353-7755; Email: guenther@cns.bu.edu.

Keywords: speech, cortex, auditory comprehension

ABSTRACT

Functional magnetic resonance imaging (fMRI) was used to investigate the representation of sound categories in human auditory cortex. Experiment 1 investigated the representation of prototypical (good) and non-prototypical (bad) 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 perceptual distortion near sound categories: 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.

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 sounds is aided by the fact that our perceptual spaces are distorted near speech sound categories. For example, if listeners are presented with synthetic speech stimuli created by varying the second formant transition in small steps through a range corresponding to the phonemes /b/, /d/, and /g/, they show very poor discriminability when two stimuli both fall within one of the categories and very good discriminability for stimuli that straddle category boundaries, a phenomenon referred to as *categorical perception* (Liberman, Harris, Hoffman, and Griffith, 1957; Liberman, Harris, Kinney, and Lane, 1961; Eimas, 1963).

Speech perception has often been characterized as involving an *auditory mode* of processing, which involves a relatively rich representation of the acoustic signal, and a categorical *phonetic mode* of processing (e.g., Pisoni, 1973). The classical explanation of categorical perception involves the phonetic mode: two sounds that fall into the same phonetic category are difficult to discriminate, whereas two sounds that activate different phonetic categories are easy to discriminate (Liberman et al., 1957). Alternatively, the auditory mode of processing allows listeners to discriminate fine acoustic differences within phonetic category if listeners are able to compare acoustic signals over a short period (less than one second) (Pisoni, 1973). In other experiments, some investigators have found that perception may be warped such that prototypical members of a category are closer to each other in auditory space than non-prototypical category members (e.g., Kuhl, 1991; discussed further below).

Some aspects of the distortion of perceptual space near sound categories appear to be 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 boundary¹ (Kuhl, 1991; Kuhl et al., 1992). This effect has been the focus of numerous studies (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; Guenther, 2000), though not all vowel categories seem to show the effect (e.g., Sussman and Gekas, 1997) and some researchers have argued that there are no substantive differences between the perceptual magnet effect and classical categorical perception (e.g., Lotto et al., 1998). Category learning-based effects on perception have also been reported in other sensory modalities, such as visual perception of shapes (Lane, 1965) and faces (Beale and Keil, 1995).

¹ It has been noted that Kuhl's non-prototypical /i/ stimulus falls close to the border with /e/ or /I/ in acoustic space (e.g., Sussman and Lauckner-Morano, 1995; Lotto et al., 1998). Kuhl's results might therefore be interpreted as a difference in discriminability between central and borderline examples of a category rather than a difference in discriminability between prototypical and non-prototypical examples as interpreted by Kuhl.

Guenther et al. (1999) hypothesized that, during category learning, auditory space becomes distorted as a result of reorganization of the auditory cortical map. 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 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 the inferotemporal cortex (Kobatake, Wang, and Tanaka, 1998) and tactile stimuli in the 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.

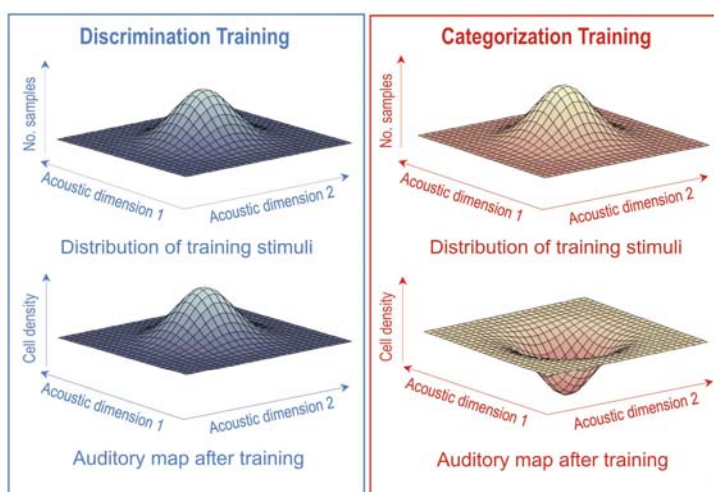


Figure 1. Hypothesized effects of discrimination training and categorization training on neural maps in the auditory cortical areas. The x and y axes of each plot correspond to two auditory dimensions, such as the first two formant frequencies for vowel sounds. For the same Gaussian distribution of training 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.

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.

Experiment 1

Participants

Nine right-handed native adult speakers of American English (4 male, 5 female) between the ages of 18-55 participated in this experiment. All subjects had normal hearing and 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 participants.

Stimuli

Two synthetic vowel stimuli, a prototypical /i/ stimulus (i.e., one judged to be a good example of /i/ by most listeners in earlier psychophysical studies) and a non-prototypical /i/ stimulus (i.e., a sound judged by most to be a bad example of /i/, chosen from near the boundary with the vowel /I/), were presented in separate blocks. Stimuli were generated using the Sensyn speech synthesis software (Sensimetrics Corporation) with the parameter settings provided in Table 1. These parameters were chosen to match the prototypical and non-prototypical vowels used by Kuhl to demonstrate the perceptual magnet effect psychophysically (Kuhl, 1991). Stimulus duration was 0.5 seconds and the inter-stimulus interval was 1.5 seconds. The sounds were presented binaurally to the subject over headphones that reduced the scanner noise by approximately 20 dB. The sounds were played at a listening level deemed comfortable by the subject, and subjects reported that the stimuli could be clearly heard over the scanner noise.

Table 1: Stimulus parameters for good and poor exemplars of /i/.

	/i/ prototype	/i/ non- prototype	Formant bandwidth
F1	266 Hz	347 Hz	100 Hz
F2	2294 Hz	2095 Hz	120 Hz
F3	3010 Hz	3010 Hz	150 Hz
F4	3300 Hz	3300 Hz	300 Hz

Procedures

Data collection. Data for Experiment 1 were obtained using a 1.5 Tesla General Electric Signa imager. 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 perisylvian cortex were acquired using an asymmetric spin-echo echo-planar imaging sequence ($\tau=25\text{ms}$, $\text{TE}=70\text{ms}$, $\text{TR}=2\text{s}$, matrix size 64×64 , 5mm thick contiguous slices with in-plane resolution= $3.1\times 3.1\text{mm}$).

Functional runs. Between four and eight functional runs were conducted for each subject. During the functional runs, stimuli were presented binaurally in a block paradigm consisting of alternating 30-second blocks of the prototypical vowel stimulus or the non-prototypical vowel stimulus separated by 30-second silent blocks 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, although all of the stimuli from within a block were the same.

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, while 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). Details of this parcellation process follow.

The dorsal surface of the temporal lobe (ventral intrasylvian area) was separated into 3 parcellation units. Starting rostrally and continuing caudally these are planum polare (PP), Heschl's gyrus (HG), and planum temporale (PT). This region is bordered medially by the circular fissure of the insula and laterally by the lateral margin of the sylvian fissure. HG lies diagonally between PP and PT and, thus, serves as the caudo-lateral and rostro medial borders of these areas, respectively. The concurrent appearance of the first transverse fissure and Heschl's fissure mark the anterior borders of both HG and PT. HG extends posteriorly to the end of Heschl's fissure. The junction of the temporal and frontal lobes (temporofrontal junction) serves as the anterior border of PP. PT extends posteriorly to the posterior end of the Sylvian fissure. The lateral surface of the temporal lobe is subdivided into the superior temporal gyrus (T1) and the middle temporal gyrus (T2). These are further subdivided into anterior (T1a, T2a) and posterior (T1p, T2p) segments. T1a and T2a begin rostrally at the temporofrontal junction and end caudally at the coronal plane immediately anterior to the rostral end of Heschl's fissure. Posterior to this plane, T1p and T2p begin and continue caudally to the posterior end of the sylvian fissure. T1a and T1p are bordered ventrally by the superior temporal fissure and dorsally by the lateral margin of the sylvian fissure. T2a and T2p are bordered ventrally by the inferior temporal fissure and dorsally by the superior temporal fissure.

The peri-sylvian portion of the parietal lobe was divided into the parietal operculum (PO) and anterior and posterior supramarginal gyri (SGa, SGp). The coronal level where the postcentral fissure meets the sylvian fissure marks the anterior border of PO and SGa. Both then continue caudally to the posterior end of the sylvian fissure. PO is bordered medially by the circular fissure of the insula and laterally by the lateral margin of the sylvian fissure. The medial border of SGa is the postcentral fissure anteriorly and the intraparietal fissure posteriorly. SGa is bordered laterally by the lateral margin of the sylvian fissure. SGp lies immediately behind SGa and continues caudally to the intermediate fissure of Jensen. It is bordered ventrally by the superior temporal fissure and dorsally by the intraparietal fissure.

Functional data analysis. 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; Nieto-Castanon et al., 2002). 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.

Results

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 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 (Liegiois-Chauvel et al., 1999; Schlaug et al., 1995), and SGp has been implicated in phonetic discrimination and identification (Caplan, Gow, and Makris, 1995). No statistically significant differences between the prototype and non-prototype activations were found in the left hemisphere ROIs. The prototypical vowel also induced less activation across the auditory cortical areas as a whole, although this difference fell short of statistical significance ($p=0.081$).

These results support a simple and straightforward explanation for reduced discriminability for sounds from near the center of a category as compared to sounds from near a category boundary (e.g., Kuhl, 1991): category-central examples of a phoneme are more difficult to discriminate from each other than borderline 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 General Discussion section), thus leading to reduced psychophysical discriminability of the category-central stimuli.

It is also possible that the increased activation for the non-prototypical /i/ stimulus occurs because this sound activates the neural representation of two different sound categories, the /i/ category and the /e/ category that it borders, while the prototypical /i/ activates only one category. This interpretation, which seems more in line with the classical view of categorical perception (e.g., Liberman et al., 1957), is addressed further in the General Discussion.

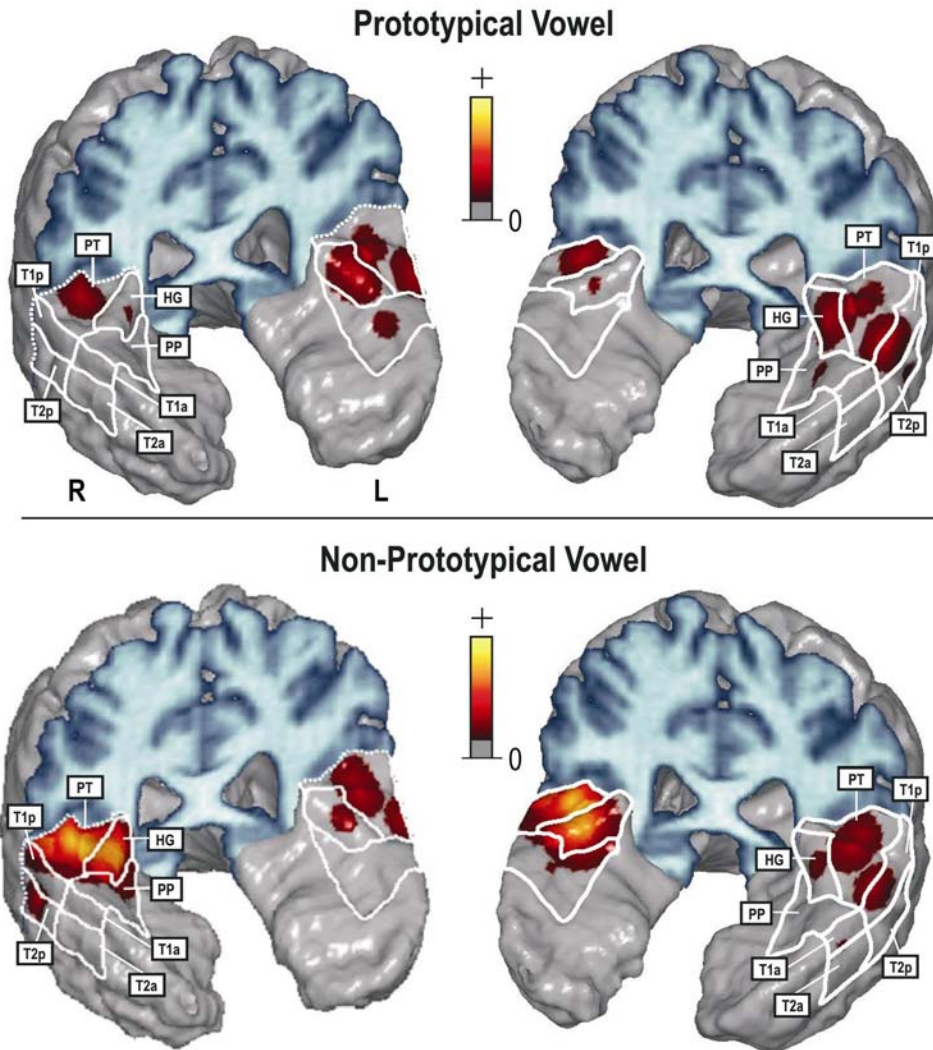


Figure 2. Average temporal lobe activations for the prototypical vowel (upper panels) and non-prototypical vowel (lower panels) conditions in Experiment 1. The frontal and parietal lobes have been removed to the posterior end of the sylvian fissure to expose the intrasylvian regions of the temporal lobe. Listening to the prototypical example of a vowel /i/ results in less activation than listening to the non-prototypical example in auditory cortical areas in the temporal lobe and supratemporal plane. HG = Heschl's gyrus; PT = planum temporale; PP = planum polare; T1a,T1p = anterior/posterior superior temporal gyrus; T2a,T2p = anterior/posterior middle temporal gyrus.

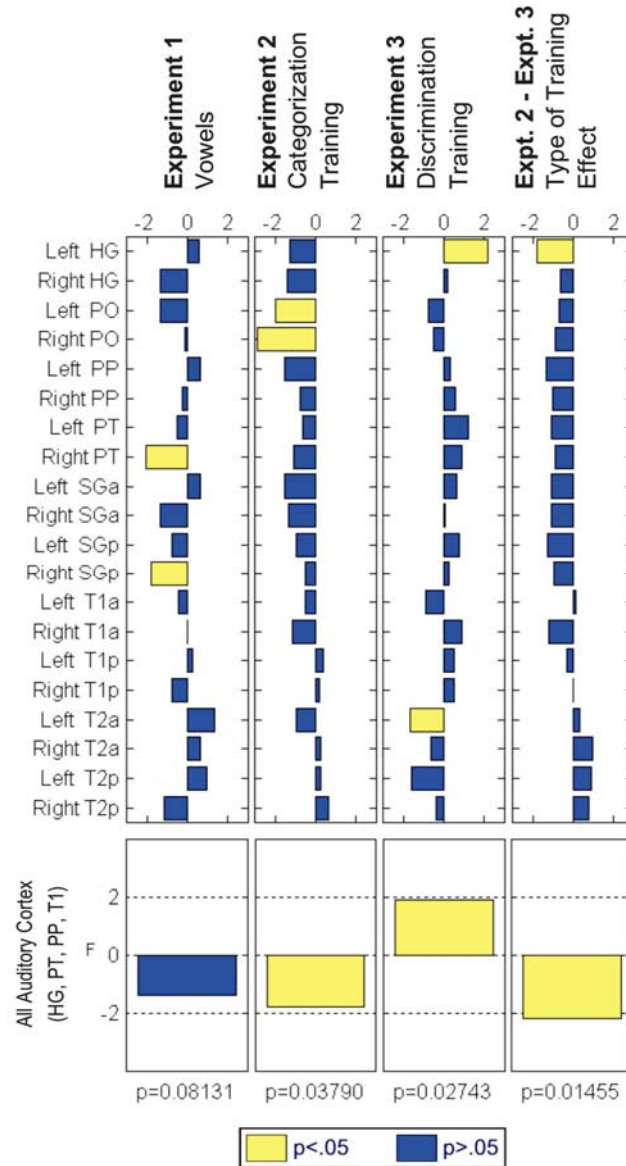


Figure 3. Comparison of results across experiments for each ROI individually (top panels) and for the auditory cortical areas considered as a whole (bottom panels). The first column shows the difference in activation between prototypical and non-prototypical examples of the vowel /i/. In general, less 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.

Experiment 2

Participants

Seven right-handed native adult speakers of American English (3 male, 4 female) between the ages of 18-55 participated in Experiment 2. All subjects had normal hearing and 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 participants.

Stimuli

Stimuli were narrow-band filtered samples of white noise with different center frequencies. Center frequencies ranged between 1000 Hz and 3500 Hz. The bandwidths of the stimuli were chosen to be equal in mel space, with the stimulus at 2500 Hz having a bandwidth of 100 Hz and the bandwidths of all stimuli falling within the range of 90-130 Hz. The stimuli were generated at a sampling rate of 16 kHz using Entropic's ESPS/Waves software on a Sun SPARCstation 10 by filtering white noise through a bandpass filter (a linear phase finite impulse response filter created using a weighted mean square error criterion) with a roll-off of approximately 20 dB per 100 Hz. Stimuli were presented binaurally to the subject over headphones that reduced the scanner noise by approximately 20 dB. The sounds were played at a listening level deemed comfortable by the subject, and subjects reported that the stimuli could be clearly heard over the scanner noise.

Each subject's threshold for discriminating the stimuli was established at the beginning of the experiment. This was done to account for rather large inter-subject differences in the ability to discriminate between the stimuli. An adaptive up-down staircase method (AX same-different paradigm) was used to determine the discrimination threshold. Stimuli for this procedure consisted of narrow-band white noise centered at different frequencies around 2500 Hz with a bandwidth of 100 Hz. The step size that shifted the center frequency of the noise stimuli was fixed at 5 Hz. Thresholds were determined both for frequencies lower than and greater than 2500 Hz. The final threshold was the average of these two thresholds. This threshold, specified in mel units, was used as an estimate of the Just Noticeable Difference (JND) for that particular subject throughout the range of frequencies used in the experiment.

The stimuli for the remainder of the experiment were generated based on this JND measure, as shown in Figure 4. The hatch marks on the x axis of this figure are spaced one JND apart. First, a reference stimulus, labeled "Milestone B" in Figure 4, was located at 3200 Hz. This stimulus and six additional stimuli spaced 1, 1.5, and 2 JND from it on either side constitute the "training region" of frequency space. Next, a second reference stimulus, milestone A, was located at a frequency 11 JNDs less than milestone B. Milestone A and stimuli spaced 1, 1.5, and 2 JND from it constitute the "control region" of frequency space. Stimuli in the training region were involved in the training phase in a manner described below. The training regimes used in the experiments required the use of stimuli from outside of the training region but not in the control

region. These additional stimuli were chosen from a uniform distribution over two regions of frequency space labeled “band edges” in Figure 4: a 4-JND-wide region between the training region and the control region, and a 4-JND-wide region located above the training region in frequency space. There was a separation of 1.5 JNDs between the band edges and both the training and control regions.

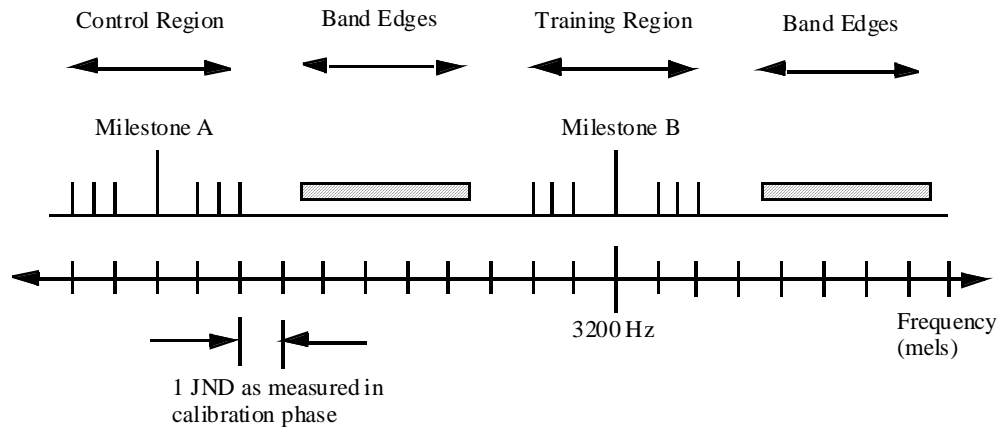


Figure 4. Locations of the center frequencies of narrow-band noise stimuli in Experiments 2 and 3. Milestone A and its neighbors form the control region and milestone B and its neighbors form the training region. Regions spanning 4 JNDs on either side of the training region, called “band edges”, are needed for the training task. See text for details.

Procedures

Data collection. Data for Experiment 2 were obtained using a 1.5T Siemens scanner using the same scanning sequence and scan parameters used for Experiment 1. Subjects underwent two scanning sessions, one before and one after a one-week categorization training procedure described below.

Categorization training. 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. Each training session lasted approximately 45 minutes. Subjects were trained to identify sounds from a “category” consisting of the seven sounds in the training region of frequency space (see Figure 4). 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 (training region) in the list. The remaining sounds in the list came from the “band edges” (see Figure 4). 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. For a detailed description of the psychophysical effects of this training task, see Guenther et al. (1999).

Functional runs. Four to eight functional runs were conducted for each subject. Subjects were told to attend to the stimuli by listening for differences from sound to sound. Stimuli were presented in a block paradigm consisting of alternating 30-second blocks of stimuli (800ms

stimulus length, 2000ms ISI) belonging to either the training region or the control region of frequency space, separated by 30-second silent blocks. Total run length was 5-½ minutes. Cortical parcellation and functional data analysis were carried out as described for Experiment 1 above.

Results

Figure 5 shows the difference between post- and pre-training activations for the training stimuli as compared to control stimuli (i.e., the activation comparison was [training post – control post] – [training pre – control pre]) 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 sounds from within the category in most of the auditory cortical areas, as evidenced by the blue areas of decreased activation in the top half of Figure 5.

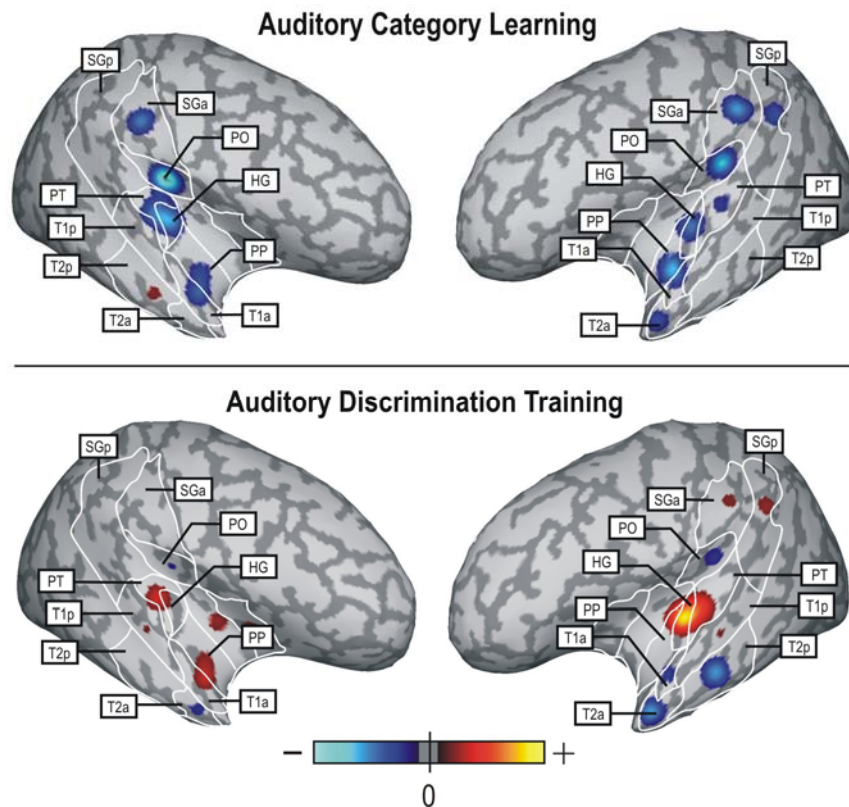


Figure 5. Results of Experiments 2 and 3 displayed on an inflated cortical surface. Plots show difference between post- and pre-training activations for the training stimuli minus the control stimuli in each region of interest. **Upper panels:** Effects of learning a novel auditory category (Experiment 2). Blue activations indicate regions 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). In contrast to category learning, discrimination training leads to a relative increase of activation (red) for the training stimuli in the peri-sylvian cortical regions. 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, SGp = anterior/posterior supramarginal gyrus.

The results of Experiment 1 showed that good examples of the sound /i/ cause smaller activation in the auditory cortical areas than non-prototypical examples. The results of Experiment 2 indicate that category learning leads to decreased auditory cortical activation for members of the category as compared to sounds from outside the category. Together these results suggest that the decreased activation for good examples of /i/ may arise from category learning during infancy, which may result in a reduced auditory cortical representation for these sounds as compared to sounds that are not clear examples of the category.

The second column of Figure 3 details the results of Experiment 2 by individual ROIs and in the auditory cortical areas considered as a whole. The greatest 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 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).

Experiment 3

Participants

Seven right-handed native adult speakers of American English (2 male, 5 female) between the ages of 18-55 participated in Experiment 3. All subjects had normal hearing and 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 participants.

Stimuli

The stimuli for Experiment 3 were narrow-band filtered samples of white noise with different center frequencies, generated in the manner detailed in the *Stimuli* section of Experiment 2 and presented binaurally to the subject over headphones that reduced the scanner noise by approximately 20 dB. The sounds were played at a listening level deemed comfortable by the subject, and subjects reported that the stimuli could be clearly heard over the scanner noise.

Procedures

Data collection. Data for Experiment 3 were obtained using a 1.5T Siemens scanner using the same scanning sequence and scan parameters used for Experiments 1 and 2. Subjects underwent two scanning sessions, one before and one after a one-week discrimination training procedure described below.

Discrimination training. In this experiment, 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 training region of frequency space (corresponding to the category in Experiment 2; see Figure 4). The subjects’ ability to discriminate between the sounds, as measured by d' , generally increased during the week of training. Note that this is the opposite of the effect of categorization training on discriminability in Experiment 2. See Guenther et al. (1999) for further details on the psychophysical effects of this training procedure.

Functional runs. Four to eight functional runs were conducted for each subject. Subjects were told to attend to the stimuli by listening for differences from sound to sound. Stimuli were presented in a block paradigm consisting of alternating 30-second blocks of stimuli (800ms stimulus length, 2000ms ISI) belonging to either the training region or the control region of frequency space. Note that the stimuli were not paired as in the discrimination training task; instead a stream of individual stimuli was played, and the subject listened for differences from sound to sound in the stream. Total run length was 5-½ minutes. Cortical parcellation and functional data analysis were carried out in the same manner as in Experiments 1 and 2.

Results

After discrimination training, there was significantly more activation for the training stimuli than before training in the auditory cortical areas. This can be seen on the inflated cortical surfaces in the bottom half of Figure 5, which shows the difference between post- and pre-training activations for the training stimuli as compared to control stimuli (i.e., [training post – control post] – [training pre – control pre]). In contrast to the decreased activation after categorization training in Experiment 2 (blue patches in top panels of Figure 5), discrimination training resulted in increased activation in the auditory cortical areas (red patches) when the training stimuli were presented after training. The third column of Figure 3 details these results by ROI. A significant increase in activation was found in left HG and in the auditory cortical areas considered as a whole.

The results of a test for differential effects of the different training regimes in Experiments 2 and 3 are presented in the rightmost column of Figure 3, which 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). 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.

General Discussion

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 produce poorer discriminability near the category center compared to stimuli further away from the category center (and nearer to the category boundary with /I/) in numerous studies (e.g., Kuhl, 1991; Kuhl et al., 1992). Less activation was found in auditory cortical areas when subjects listened to a prototypical /i/ than when they listened to a non-prototypical /i/. It is commonly believed that stimuli with larger representations in sensory cortical areas are easier to discriminate from each other than stimuli with smaller cortical representations. In light of this, the current results suggest a very simple and straightforward account of the perceptual magnet effect: prototypical members of the /i/ category are more difficult to discriminate from each other than non-prototypical members because they are processed by a smaller auditory cortical representation.

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. The results showed a reduced auditory cortical representation for sounds within the category after the week of training; this reduced representation provides an account for the decreased discriminability seen for these test sounds after categorization training (Guenther et al., 1999). One possible explanation for the relatively large activation seen for non-prototypical members of the /i/ category in Experiment 1 is that prototypes only activate cells representing one category, while non-prototypical sounds from near a category border may activate cells representing more than one category. In Experiment 2, however, sounds from a learned category were compared to sounds from a control region of frequency space far away from the learned category. These control region sounds would not be expected to activate cells representing more than one category. The results of Experiment 2 thus indicate that the auditory cortical representation of sounds in a category actually shrinks as a result of category learning.

Experiment 3 was the same as Experiment 2 except that discrimination training was used instead of categorization training in order to identify differences in the effects of these tasks on auditory cortical maps. 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 central category members in Experiments 1 and 2, 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, which determines what happens to the auditory cortical representation of these stimuli.

In all three experiments, activations of the middle temporal gyrus (T2a, T2p) differed from those in the auditory cortical areas (see Figure 3). In particular, clear examples of a category (specifically, prototypical /i/ sounds or sounds from the center of a learned non-speech category) tended to cause more activation in T2 than sounds that were not clear members of a category, and discrimination training led to a decrease in the amount of T2 activation; this pattern is

opposite to the pattern in the auditory cortical areas. 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.

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 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 other sounds. In other words, the brain may be reducing the processing time for category prototypes, rather than 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 category-related perceptual distortion in other sensory modalities (e.g., 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 usually 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.

Acknowledgements

We thank Julie Goodman, Jennifer Holmes, Fatima Husain, Bruce Rosen, David Kennedy, Andrew Worth, 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

- Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America*, *101*, 1090-1105.
- Bauer, H.-U., Der, R., & Herrmann, M. (1996). Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, *8*, 757-771.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217-239.
- Burns, E. M., & Ward, W. D. (1978). Categorical perception -- phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456-468.
- Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, *45*, 293-298.
- Caviness, V. S. Jr., Meyer, J., Makris, N., & Kennedy, D. N. (1996). MRI-based topographic parcellation of human neocortex: An anatomically specified method with estimate of reliability. *Journal of Cognitive Neuroscience*, *8*, 566-587.
- Diesch, E., Iverson, P., Kettermann, A., & Siebert, C. (1999). Measuring the perceptual magnet effect in the perception of /i/ by German listeners. *Psychological Research*, *62*, 1-19.
- Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, *6*, 206-217.
- Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: implications for the perceptual magnet effect. *Perception and Psychophysics*, *61*, 561-577.
- Ghosh, S., Nieto-Castanon, A., Tourville, J., & Guenther, F. (2001). ROI-based analysis of fMRI data incorporating individual differences in brain anatomy. *NeuroImage*, *13* (6) Pt. 2, 125.
- Gloor, P. (1997). *The temporal lobe and limbic system*. Oxford: Cambridge University Press.
- Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology*, *23*, 178-200.
- Guenther, F. H. (2000). An analytical error invalidates the "depolarization" of the perceptual magnet effect. *Journal of the Acoustical Society of America*, *107*, 3576-3580.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*, 1111-1121.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, *106*, 2900-2912.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*, 553-562.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1130-1140.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, *80*, 324-330.

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*, 93-107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606-608.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review*, *72*, 275-309.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*, 379-388.

Liegeois-Chauvel, C., de Graaf, J. B., Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex*, *9*, 484-496.

Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 1665-1679.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150-157.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, *103*, 3648-3655.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467-479.

Nieto-Castanon, A., Ghosh, S. S., Tourville, J. A., and Guenther, F. H. (2002). ROI-based analysis of functional imaging data. *Boston University Technical Report CAS/CNS TR-02-010*.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*, 253-260.

Reber, P. J., Stark, C. E., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 747-750.

Recanzone, G. H., Merzenich, M. M., Jenkins, W. M., Grajski, K. A., & Dinse, H. R. (1992). Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained in a frequency-discrimination task. *Journal of Neurophysiology*, *67*, 1031-1056.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience*, *13*, 87-103.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*, 699-701.

Sharma, A., & Dorman, M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *Journal of the Acoustical Society of America*, *104*, 511-517.

Sussman, J. E., & Gekas, B. (1997). Phonetic category structure of [I]: extent, best exemplars, and organization. *Journal of Speech, Language, and Hearing Research*, *40*, 1406-1424.

Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: identification and change/no-change discrimination. *Journal of the Acoustical Society of America*, *97*, 539-552.