To appear in: *Proceedings of the Speech Recognition as Pattern Classification (SPRAAC) Workshop,* Nijmegen, The Netherlands, July 11-13, 2001.

EFFECTS OF CATEGORIZATION TRAINING ON AUDITORY PERCEPTION AND CORTICAL REPRESENTATIONS

Frank H. Guenther
Boston University, Boston, MA and Massachusetts Institute of Technology, Cambridge, MA
Alfonso Nieto-Castanon, Jason A. Tourville, and Satrajit S. Ghosh
Boston University, Boston, MA

ABSTRACT

Our ability to discriminate sounds is not uniform throughout acoustic space. One example of auditory space warping, the *perceptual magnet effect*, appears to arise from exposure to the phonemes of an infant's native language. We have developed a neural model that accounts for the magnet effect in terms of neural map dynamics in auditory cortex. This model predicts that it should be possible to induce a magnet effect for non-speech stimuli. This prediction was verified by a psychophysical experiment in which subjects underwent categorization training involving non-speech auditory stimuli that were not "categorical" prior to training. The model further predicts that the magnet effect arises because prototypical vowels have a smaller cortical representation than non-prototypical vowels. This prediction was supported by an fMRI experiment involving prototypical and non-prototypical examples of the vowel /i/. Finally, the model predicts that categorization training with non-speech stimuli should lead to a decreased cortical representation for stimuli near the center of the training category. This prediction was supported by an fMRI experiment involving categorization training with non-speech auditory stimuli.

1. INTRODUCTION

It is well known from phenomena such as categorical perception that our ability to discriminate speech-like sounds is not uniform throughout acoustic space. In one heavily discussed example of auditory space warping, referred to as the *perceptual magnet effect* (Kuhl, 1991), prototypical examples of a vowel or semi-vowel are more difficult to discriminate from each other than non-prototypical examples. This effect appears to arise due to linguistic experience, since 6-month-old American babies show the effect for an American vowel but not a Swedish vowel, and Swedish babies show the opposite effect (Kuhl et al., 1992).

We have developed, experimentally tested, and refined a neural model that explains the perceptual magnet effect as the result of changes to neural maps in auditory cortical areas (see also Bauer, Der, and Herrmann, 1996). These changes are hypothesized to occur during vowel category learning in infancy (Guenther and Gjaja, 1996; Guenther, Husain, Cohen, and Shinn-Cunningham, 1999). In this paper, we describe the model and present the results of psychophysical and brain imaging experiments that support its account of the perceptual magnet effect and, more generally, the effects of categorization training on sensory cortical maps.

2. A MODEL OF THE EFFECTS OF CATEGORIZATION TRAINING ON AUDITORY CORTICAL MAPS

Many neurophysiological studies of sensory maps have shown that disproportionately large exposure to a particular type of stimulus typically leads to a larger cortical representation for that stimulus. For example, kittens reared in a visual environment consisting only of vertical stripes have more visual cortex cells tuned to vertical contours than kittens reared in a normal environment (e.g., Rauschecker and Singer, 1981). Analogous results have been found in other sensory modalities. Preferential stimulation of a digit in monkeys leads to a larger cortical representation for that digit in somatosensory cortex (Jenkins, Merzenich, Ochs, Allard, and Guíc-Robles, 1990). In the auditory realm, Recanzone, Schreiner, and Merzenich (1993) found that repeatedly exposing monkeys to tones in a particular frequency range during learning of a tone discrimination task resulted in an increase in the area of auditory cortex preferentially activated by sounds in the trained frequency range and a concomitant increase in the discriminability of the training tones.

The finding that sensory neural maps grow with heavy stimulus exposure has been explained by neural network models commonly referred to as self-organizing feature maps. Figure 1 schematizes a typical self-organizing feature map. Roughly speaking, a self-organizing feature map contains a subcortical layer (or layers) and a cortical layer of cells. The subcortical layer represents incoming sensory stimuli. Cells in the cortical layer compete with each other through inhibitory connections, with only the cells receiving the largest total input from the subcortical layer becoming active when a stimulus is presented. The amount of input to each cortical cell depends on the synaptic weights between the subcortical layer and the cortical layer. When a stimulus is presented, the weights projecting to the cortical cells that "win" the competition are changed in such a way that those cells become even more likely to win the competition when the same stimulus pattern is later presented to the network.

In the "classical" formulation of a self-organizing feature map, increased exposure to a set of stimuli leads to a larger cortical representation for those stimuli (e.g., von der Malsburg, 1973; Grossberg, 1976; Kohonen, 1982). Furthermore, it is widely believed that, all else equal, larger sensory cortical representations lead to better discriminability of the represented stimuli. For example, we have a larger somatosensory cortical representation (i.e., more cortical area per unit of skin surface) for our fingertips than our forearms, and we are better at two-point discrimination with a fingertip than with the forearm. This relationship makes sense when one considers that neural representations involving larger numbers of cells can better "average out" the noisy signals of individual neurons.

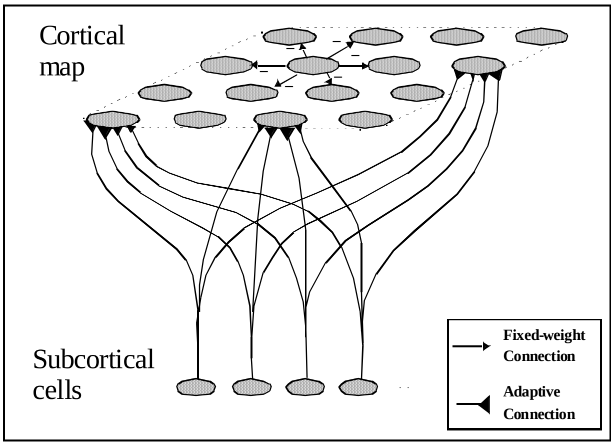

Figure 1. The basic architecture of a self-organizing feature map neural network.

Although prototypical vowels are presumably much more commonly experienced by a listener than non-prototypical vowel-like sounds, listeners are *worse* at discriminating the prototypical vowels. This clearly conflicts with the classical formulation of a self-organizing feature map, since in a classical self-organizing feature map there would be a larger representation for prototypical vowels, leading to better discriminability as compared to non-prototypical vowel-like sounds.

Bauer et al. (1996) have proposed a neural architecture that can produce a smaller cortical representation for the most frequently encountered training stimuli. Based on the results of our experiments designed to test the idea that the magnet effect is the result of changes in the neural maps in auditory cortex (Guenther et al., 1999), we favor this formulation over a closely related but somewhat different neural model of the magnet effect proposed by Guenther and Gjaja (1996). We have further determined that it is the type of training an infant undergoes with speech sounds, not the distribution of training stimuli, that leads to a shrinking of the neural map for speech sound stimuli, as discussed in Section 3. This leads to the model of the effects of categorization training schematized in Figure 2.

The model's explanation of the magnet effect is simple and straightforward: prototypical examples of a category are more difficult to discriminate from each other than non-prototypical examples because they have a smaller representation in auditory cortical maps (see also Bauer et al., 1996). The model further posits that this reduced cortical representation results from phoneme category learning during infancy. In particular, learning to treat sounds from a particular region of acoustic space as members of the same category leads to a decrease in the size of the auditory cortical representation of sounds near the center of that region.

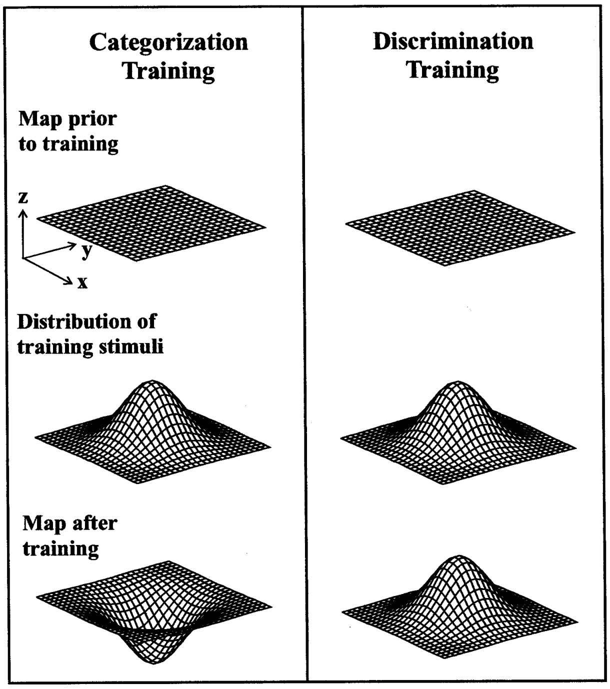

Figure 2. Hypothesized changes in the neural map in auditory cortex as a result of categorization training (left) and discrimination training (right). The x and y axes of all plots correspond to two acoustic dimensions, such as the first two formant frequencies of a vowel sound. The z axis corresponds to the number of cells in the map devoted to each region of frequency space (top and bottom plots) or the number of training stimuli from that region of frequency space (middle plots). According to this model, categorization training leads to a decrease in the number of cells coding the most frequently encountered stimuli, whereas discrimination training leads to an increase in the number of cells coding the most frequently encountered stimuli.

The model attributes the perceptual magnet effect to neural map formation properties that are not unique to speech stimuli. This leads to the prediction that it should be possible to induce a perceptual magnet-like effect if categorization training is carried out using non-speech stimuli. As described in the next section, this prediction was verified by psychophysical experiments involving bandpass filtered acoustic noise stimuli. Functional brain imaging experiments were then used to verify the model's prediction that stimuli from near the center of the newly learned category will have a reduced cortical representation, as described in Section 4.

3. INDUCING A MAGNET EFFECT WITH CATEGORIZATION TRAINING

We have performed psychophysical experiments to test the model's prediction that a perceptual magnet effect can be induced using stimuli that are not, prior to training, treated in a categorical manner (Guenther et al., 1999). In this experiment, subjects performed a category learning task in a 45-minute training stimuli involving bandpass-filtered acoustic noise stimuli that varied in center frequency of the pass band. Each subject's ability to discriminate these sounds was estimated before and after training using a d' measure. The stimuli were not perceived as speech-like by experimental subjects, and they were not perceived "categorically" prior to training.

Seven stimuli were generated for each of two regions of frequency space: a control region and a training region (see Figure 3). The ability to discriminate the "Milestone" stimulus from each range (i.e., the stimulus at the center of the range) from the other stimuli in the range was measured before and after training.

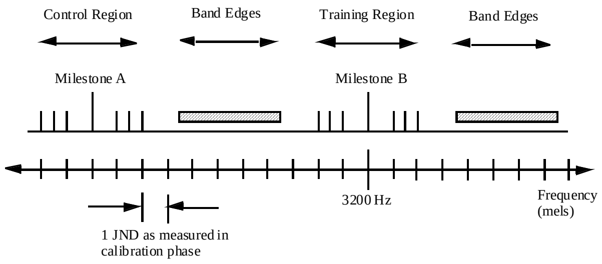

Figure 3. Stimuli for the psychophysical experiments investigating the effects of categorization and discrimination training on auditory perceptual space.

During each trial of the training task, the subject was presented with a sequence of two, three, or four stimuli. On each trial, one of these stimuli was from the training region, and the rest were from other parts of frequency space labeled "Band Edges" in Figure 3. The subject had to choose which of the stimuli in the sequence belonged to the training "category". Subjects generally got significantly better at the task over the roughly 45-minute training session. Sounds from the control region were not heard during training.

As shown in Figure 4, categorization training was shown to produce decreased discriminability of the stimuli from the center of the training category, as in the perceptual magnet effect. Although the training region stimuli were encountered more frequently than the control region stimuli during the experiment, subjects showed a reduction in their ability to discriminate stimuli in the training region as compared to the control region. This verifies the model's prediction that one can induce a perceptual magnet-like effect using non-speech stimuli in a categorization training task.

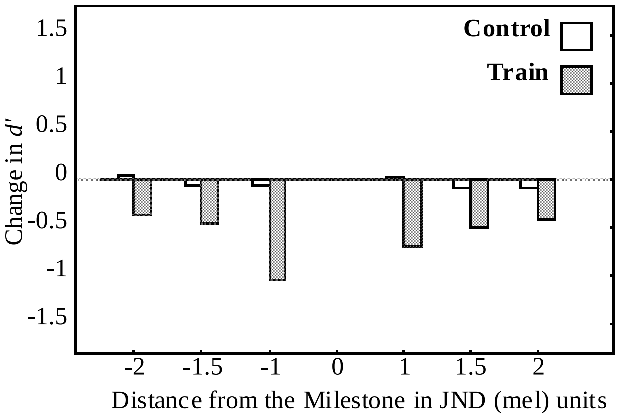

Figure 4. Effects of categorization training on discriminability of stimuli within a newly formed category.

In a second experiment, the same stimuli from the categorization training experiment were used in a discrimination training task rather than a categorization training task. On each trial, the subject was presented with two stimuli from the training region and was asked to report if the stimuli were "same" or "different". The results of this experiment are shown in Figure 5. Whereas categorization training led to a decrease in the discriminability of the training stimuli, discrimination training with the same set of stimuli led to increased discriminability. This indicates that it is the nature of the training task, and not just the distribution of the training stimuli, that leads to the perceptual magnet effect seen in the first experiment.

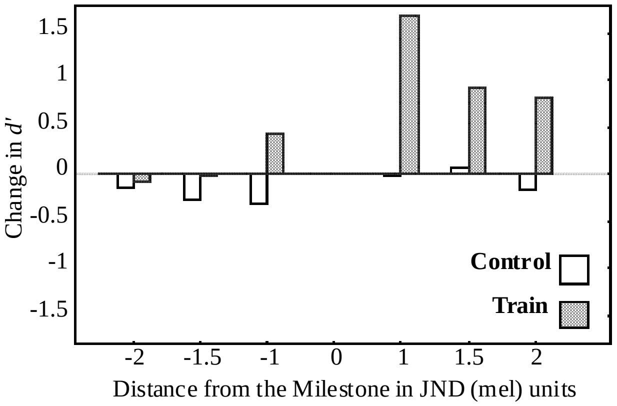

Figure 5. Effects of discrimination training on the discriminability of training stimuli.

4. fMRI EXPERIMENTS

The model's explanation for the perceptual magnet effect and the effects of categorization training was tested with two functional magnetic resonance imaging (fMRI) experiments. The first tested the prediction that prototypical examples of a vowel have a smaller auditory cortical representation than non-prototypical examples. The second tested the model's prediction that non-speech auditory stimuli from within a newly learned category will have a smaller cortical representation than stimuli that were not treated as members of a category during training.

Subjects. Nine right-handed native speakers of American English (4 male, 5 female) ages 18-40 participated in Experiment 1. Experiment 2 involved three subjects ages 18-40. Subjects for both experiments had no history of language or other neurological disorders. The experimental protocol was approved by the Boston University committee on human subjects. Informed consent was obtained from all subjects.

Image collection. Data for Experiment 1 were obtained using a 1.5T General Electric Signa imager. Data for Experiment 2 were obtained using a 1.5T Siemens 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 peri-sylvian cortex were acquired using an asymmetric spin-echo echo-planar imaging sequence (τ=-25ms, TE=70ms, TR=2s, matrix size 64x64, 5mm thick contiguous slices with in-plane resolution=3.1x3.1mm).

Data analysis. Individual functional runs were realigned (motion-corrected) using rigid body transformations to the first image in each scan, then coregistered with a structural T1 scan for each subject. Two runs were rejected for scanner data collection problems not detected during scanning. The remaining runs were visually inspected to meet noise and residual motion criteria, then tested for paradigm-correlated observed motion. Three runs showed excessive correlated motion and were thus removed from the analysis. Structural T1 images were parcellated individually for each subject to define 10 brain regions of interest (ROIs) on the basis of anatomical markers according to the procedure described by Caviness et al. (1996). The use of this parcellation procedure for each individual avoids the need for spatial averaging of the statistical parameter maps (and the subsequent loss of spatial resolution). The ROIs were ten peri-sylvian cortical areas, including areas known to become active during perceptual processing of auditory speech stimuli: Heschl’s gyrus (H1), parietal operculum (PO), planum polare (PP), planum temporale (PT), anterior and posterior supramarginal gyrus (SGa, SGp), anterior and posterior superior temporal gyrus (T1a, T1p), and anterior and posterior middle temporal gyrus (T2a, T2p). HG, PT, and T1 are commonly considered to be auditory areas. PO, SG, and T2 are multimodal areas that become active during some speech or language tasks. Figure 6 illustrates the ROIs on the temporal lobe.

Data reduction was applied to each ROI to obtain one temporal activation profile characterizing the response of all voxels within a given region. This was defined for each subject as the first eigenvariate of the response of all voxels inside each ROI. Significance of specific contrasts for each ROI activation profile were obtained using the general linear modeling (GLM) framework within the SPM statistical analysis package.

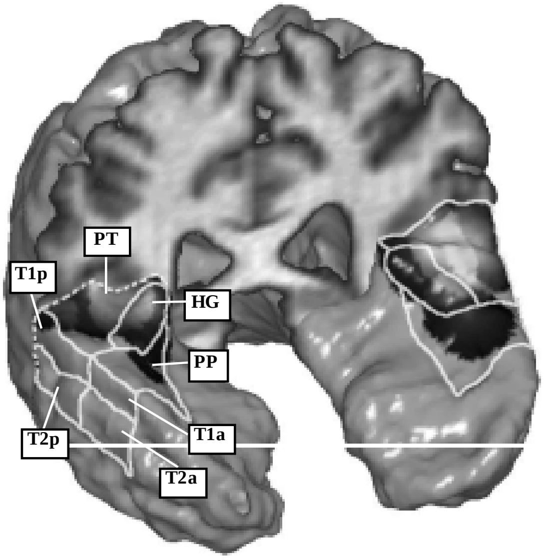

Figure 6. Temporal lobe regions of interest for the fMRI experiments. The frontal lobe has been removed to expose the temporal plane. Activation while listening to a non-prototypical example of the vowel /i/ can be seen in Heschl’s gyrus (HG), planum temporale (PT), and planum polare (PP), and the posterior superior temporal gyrus (T1p).

4.1. fMRI Experiment 1

Stimulus presentation. 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 the following parameters: sampling frequency 8KHz, amplitude of voicing 60, and 4 formant frequencies (266Hz, 2294Hz, 3010Hz, 3300Hz for the prototypical stimulus, and 347Hz, 2095Hz, 3010Hz, 3300Hz for the non-prototypical stimulus, with bandwidths of 100Hz, 120Hz, 150Hz, and 300Hz respectively). These parameters were 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. Four subjects heard the prototypical vowel block first, and five heard the non-prototypical vowel block first. Between four and eight runs were conducted for each subject.

Results. Significant activation ($p<0.05$) in response to vowel sounds was found in 17 of the 20 ROIs (10 regions x 2 hemispheres = 20 ROIs). Only left T1a, left

T2a, and left T2p did not show significant activation for either the P or NP stimulus. The averaged activations for auditory cortical regions on the temporal lobe and supratemporal plane for the P and NP conditions are shown in Figure 7. As predicted by the neural models of Guenther et al. (1999) and Bauer et al. (1996), less activation is seen for the prototypical vowel than the non-prototypical vowel in auditory cortical areas, thus supporting a simple explanation for the perceptual magnet effect: prototypical vowels are more difficult to discriminate from each other than non-prototypical vowels because they have a smaller cortical representation, and smaller cortical representations are more susceptible to noise in the neural signals.

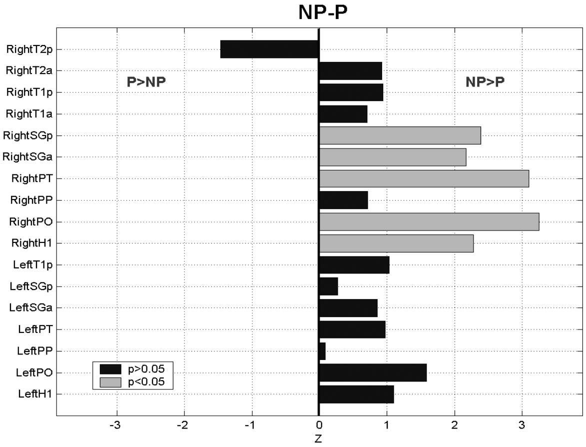

Figure 7. Results of fMRI Experiment 1. See text for details.

Figure 7 shows the difference in activation between the P and NP conditions for all ROIs that were significantly activated by the stimuli. Significant differences ($p<0.05$) were found in five right hemisphere regions: Heschl's gyrus (H1), anterior and posterior supramarginal gyrus (SGa, SGp), planum temporale (PT), and parietal operculum (PO). H1 includes primary auditory cortex and PT is a higher-order auditory cortical area. SG and PO are peri-sylvian parietal areas that have been implicated in phoneme discrimination (Caplan, Gow, and Makris, 1995).

These results support the model's simple explanation for the perceptual magnet effect: prototypical examples of a category are more difficult to discriminate from each other than non-prototypical examples because they have a smaller representation in auditory cortical maps.

4.1. fMRI Experiment 2

Stimulus presentation. In Experiment 2, the same bandpass auditory noise stimuli used in our psychophysical experiments (Section 3) were presented to subjects in the scanner. Scans were performed before and after the subject underwent a week of categorization training involving these stimuli.

Results. The results of these scans are presented in Figures 8 and 9. In these figures, P stands for category-prototypical stimuli, which corresponds to stimuli from within the training region, and NP stands for non-prototypical stimuli, i.e. those from the control region.

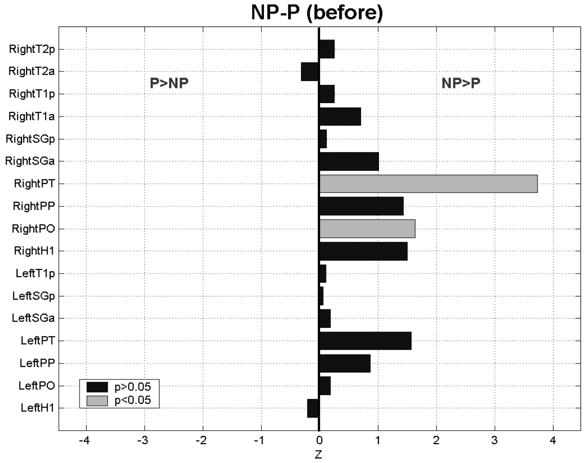

Figure 8. Results of pre-training scans in fMRI Experiment 2. Bars indicate difference in activation between the control region (NP) and training region (P) stimuli for the seventeen ROIs that showed significant activation in fMRI Experiment 1.

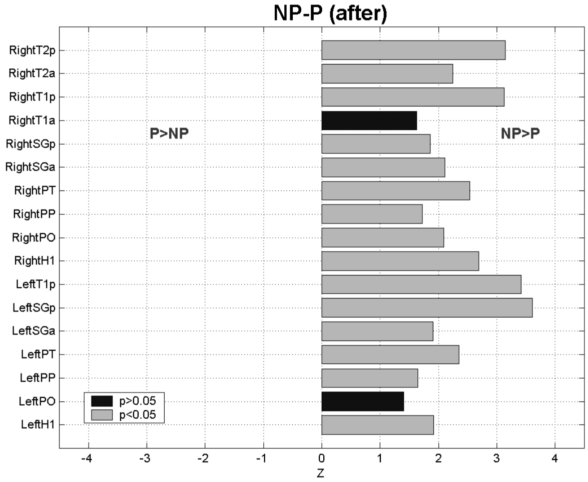

Figure 9. Results of post-training scans in fMRI Experiment 2. Bars indicate difference in activation in the non-prototype (NP) and prototype (P) conditions for the seventeen ROIs that showed significant activation in fMRI Experiment 1.

The pre-training scans (Figure 8) indicate that the stimuli were somewhat biased in their cortical response even before any training, with the non-prototypical stimuli causing a slightly larger activation. We suspect that this is in part due to the small number of subjects (3) used in this experiment or because the non-prototypical stimuli stood out better from the acoustic noise produced by the MRI scanner. The post-training scans, on the other hand, show a greatly increased difference in the size of the activations. After training, the sounds from the training region (P) induced a much smaller activation relative to the sounds from the control region (NP). This supports the model's prediction that categorization training leads to a decreased cortical representation for stimuli from near the center of the category.

5. DISCUSSION

The experiments described in the current article were designed to investigate learned warpings of auditory perceptual space by testing a neural network model of the perceptual magnet effect. This model posits that phoneme category learning in infancy leads to the perceptual magnet effect because it causes a reduction in the size of the auditory cortical representation of prototypical examples of a vowel category. The model's assertion that general neural map formation properties were responsible for the effect implies that it should be possible to induce the effect in non-speech stimuli. This prediction was verified by a psychophysical experiment showing that subjects learning a new category for non-speech auditory stimuli get worse at discriminating central examples of the category from each other. An fMRI analysis revealed that listening to prototypical examples of the vowel /i/ leads to less activation in peri-sylvian cortical areas than listening to non-prototypical examples. A second fMRI study suggests that categorization training leads to a decrease in the relative size of the cortical representation for central members of a category.

Taken together, these results strongly support the following assertions of the Guenther et al. (1999) neural model of auditory map formation:

- Categorization training leads to a relative decrease in the size of the cortical representation of prototypical examples of a category.
- Similarly, vowel category training in infancy leads to a decrease in the size of the cortical representation of prototypical examples of some speech sounds.
- This decreased representation is responsible for the perceptual magnet effect (see also Bauer et al., 1996).

6. ACKNOWLEDGEMENTS

We gratefully thank Dr. Julie Goodman and the Massachusetts General Hospital NMR Center for the use of their imaging facilities, and Dr. Fatima Husain for her help in generating the auditory stimuli. Address correspondence to: Frank Guenther, Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston MA 02215. Email: guenther@cns.bu.edu.

7. REFERENCES

Bauer, H.-U., Der, R., and Herrmann, M. (1996). Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, **8**, 757-771.

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. *J. Cog. Neurosci.*, **8**, 566-587.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics,* **23**, pp. 121-134.

Guenther, F.H., and 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., and 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.

Jenkins, W.M., Merzenich, M.M., Ochs, M.T., Allard, T., and Guíc-Robles, E. (1990). Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of Neurophysiology*, **63**, 82-104.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**, 59-69.

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

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

Rauschecker, J.P., and Singer, W. (1981). The effects of early visual experience on the cat's visual cortex and their possible explanation by Hebb synapses. *Journal of Physiology*, **310**, 215-239.

Recanzone, G.H., Schreiner, C.E., and 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.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striata cortex. *Kybernetik,* **14**, pp. 85-100.