

Effects of categorization and discrimination training on auditory perceptual space

Frank H. Guenther^{a)}

Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, Massachusetts 02215 and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

pears to be warped, although apparently not as dramatically as for consonants.

It is very likely that some of the warping of auditory space is “built in” to the auditory nervous system. Evidence for this comes from studies of auditory perception in animals and newborn infants. For example, the discriminability by chinchillas of changes in VOT for stimuli varying between [S~] and [#~] is nonuniform and peaks at a VOT of about 30 ms, which is near the voiced/voiceless boundary in English (Kuhl and Miller, 1975, 1978; Kuhl, 1981). A similar result was also reported for macaque monkeys (Kuhl and Padden, 1982). Increased discriminability was also found at the /" /-/\$/ and /\$/-/,/ phonetic boundaries of a continuum of *F*2 transition onset frequencies in the macaque monkey (Kuhl and Padden, 1983). Eimas *et al.* (1971) showed that human infants 1–4 months old produced evidence of categorical perception for the voiced/voiceless distinction, further suggesting that this effect is a consequence of auditory mechanisms that are present at birth.

A. Experience-based warping of auditory space

Other aspects of the warping of auditory space appear to arise from learning, rather than from built-in properties of the auditory system. Evidence for this view comes from cross-language studies, since differences in the locations of warping in auditory space across languages are presumably the result of learning driven by linguistic experience. One example of such a difference is the small but systematic difference in the VOT boundary for the voiced/voiceless distinction across languages (e.g., Lisker and Abramson, 1970). Another example is the language specificity of the warping of auditory space for vowels as measured in studies of the perceptual magnet effect. In a study of 6-month-old English and Swedish infants presented with English and Swedish vowel stimuli, Kuhl *et al.* (1992) found that infants had more difficulty discriminating between stimuli falling near a prototypical vowel from their native language than stimuli falling near a prototypical vowel in the non-native language.

The experiments described in the current article were designed to investigate learned warpings of auditory perceptual space. Because the experiments were designed in part to test neural network models of the perceptual magnet effect (as described in the next section), and because the magnet effect is one of the most heavily studied examples of a learned warping of auditory space, we will frequently refer to it when discussing our experimental results. We do not mean to imply by this that the perceptual magnet effect should be considered as a separate phenomenon from learned instances of categorical perception.

Liberman (1957) identified two possible learning processes that might underlie categorical perception. The first, *acquired distinctiveness*, is defined as an increase in percep-

It is also commonly believed that, all else being equal, stimuli that have a larger cortical representation are more easily discriminated from one another than stimuli that have a smaller cortical representation. For example, the cortical representation of the fingers in human somatosensory cortex is disproportionately large when compared to the representation of the back, and, correspondingly, humans are typically much better at discriminating tactile stimuli with their fingers than with their backs (e.g., Kandel, 1985). Similarly, the primary visual cortex representation of the high-resolution foveal area of our retinas is much larger than the representation of the low-resolution visual periphery.

If one assumes that frequent exposure to a stimulus leads to a larger cortical representation, and that larger cortical representations lead to better discriminability, then one sees a paradoxical aspect of the perceptual magnet effect: in the magnet effect, discriminability of more frequently encountered stimuli (prototypical vowels) is *worse* than discriminability of less frequently encountered stimuli (nonprototypical vowels). Two recent neural network models posit explanations for the perceptual magnet effect in terms of experience-based formation of neural maps in the auditory system (

C. Goals of the current experiments

The first purpose of the current studies was to observe whether it is possible to induce acquired similarity for a category-relevant dimension of nonspeech stimuli (auditory noise stimuli) using a categorization training task. This type of induced “perceptual magnet effect” is predicted by the Guenther and Gjaja (1996) model since this model attributes the reduced discriminability near a category prototype to neural map formation principles that are not specific to speech. Although this sort of acquired similarity had been identified as a possible learning mechanism underlying categorical perception several decades ago (e.g., Liberman, 1957; Lane, 1965), it apparently has not been demonstrated experimentally (Goldstone, 1994; Liberman, 1996). A second purpose of the current study was to investigate some of the learning conditions that are necessary to reduce sensitivity for frequently encountered stimuli, if it is indeed possible to induce such an effect. A final purpose of this study was to test between the Guenther and Gjaja (1996) and Bauer *et al.* (1996) neural models of the perceptual magnet effect in order to form a clear and testable hypothesis concerning the properties of the nervous system that lead to this effect. Most of the experimental results reported herein have been presented in preliminary form in conference publications (e.g., Husain and Guenther, 1998a,b).

I. EXPERIMENTS

Four experiments were performed. All experiments consisted of four phases: a calibration phase in which a subject’s detection threshold for auditory stimuli like those used in later phases of the experiment was determined, a pretest phase to determine baseline sensitivity, a training phase, and a post-test phase

limina.hses phaslimitsbeen-2-298(trainin8(phasl3(totest)-413main)-913.4(b(pre-9135ted)-43relin9135tfromn)-913.4-298

stimuli spaced ± 1 , 1.5, and 2 jnd from it constitute the

E. Training phase

The type of training varied for each experiment, and the different training paradigms are explained along with the relevant experiments below. All experiments shared the following criterion for inclusion of a subject's results in the analysis: the subject must have responded correctly on half the trials of each of the ten training subsessions which comprised the training phase. If the subject did not meet this criterion, it was assumed that he/she did not succeed in learning the training task, and his/her results were thus excluded from the statistical analyses.

1. Experiment I

The main goal of the first experiment was to investigate whether it is possible to induce a decrease in discriminability along a category-relevant dimension of a set of nonspeech stimuli that was repeatedly encountered during a training session. This would constitute a demonstration of acquired similarity along a category-relevant dimension, and it would also be in keeping with models of the perceptual magnet effect that attribute the effect to neural map formation properties that are not specific to speech (Guenther and Gjaja, 1996).

a. Training. In the training phase of experiment I, subjects were trained to choose sounds that belonged to the training region (i.e., milestone B and its neighbors) from a list of sounds. Specifically, subjects were told that they were to learn to identify sounds from a category, referred to as the "prototype category" and corresponding to the training region of frequency space in Fig. 2, and that during training they would have to choose the prototype category sound from a list of sounds that included only one member of the prototype category. Since the subjects were taught to treat the training region sounds as members of the same category, we will refer to this type of training as *categorization training*. The subjects underwent two types of training trials: (1) listening trials in which they heard example sounds from the training region and did not have to make any response, and (2) identification trials in which they identified one sound from a list of sounds as belonging to the training region. During a listening trial, subjects heard four sounds randomly chosen from a set of nine sounds which were evenly spaced in 0.5-jnd increments within the training region. These included the milestone B and its six neighbors used in the testing procedure, plus the two stimuli falling ± 0.5 jnd from the milestone. During an identification trial, subjects heard a short list of sounds, only one of which came from the training region. The other sounds that comprised the identification trial were generated from the "band edges" regions flanking the training region (see Fig. 2

sounds in the control region before and after training, and Fig. 4(b) shows the same results for the training region. Subjects were significantly worse ($p < 0.05$) at discriminating stimuli in the training region after training compared to before training [$t(5) = -12.4$; $p < 0.05$], but not in the control region [$t(5) = -1.48$; $p > 0.05$]. Figure 4(c) compares the change in d' for the control and training regions. The change in d' was calculated as the percentage increase or decrease in d' from pretest to post-test. This figure indicates that the change in sensitivity for the training set of stimuli was significantly more negative [$t(5) = -5.14$; $p < 0.05$] than the change in sensitivity for the control region. All ten subjects showed a decrease in sensitivity for the training region, and eight of the ten showed a larger sensitivity decrease in the training region than in the control region.

Discriminability before and after training was also compared across groups using Gourevitch and Galanter's (1967) G statistic. Overall, as seen in Table II, there was a general pattern for sensitivity to worsen (indicated by the negative values) for the training region. On the other hand, sensitivity for the control region, across all the comparison steps, did not change significantly.

##!

duced by training would make this small fatigue effect difficult to detect. Because we are primarily interested in the relative effects of training on one region of frequency space (the training region) as compared to another (the control region), the source of the small negative changes in d' for the control region was not investigated further in this paper.

Figure 5(b) shows the results of the pre- and post-tests for the training region. Subjects showed a significant increase in d' [$t(5)=2.29$, $p<0.05$] after training. The increase in d' was significantly greater for the training region as compared to the control region [$t(5)=3.23$, $p<0.05$; see Fig. 5(c)].

The general pattern for sensitivity to improve for the training region, but not the control region, is also indicated by the G scores listed in Table III. Note that for the training region, the most positive change in sensitivity occurred to the right of the prototype of the training region. In fact, the sensitivity for the -2 and -1.5 jnd stimuli did not change significantly. Perhaps relatedly, subjects as a group showed far fewer errors for the -2 and -1.5 jnd stimuli during the pretest than they showed for the other four stimuli, with only

nine total errors for the -2 jnd stimulus and 20 total errors for -1.5 jnd stimulus as compared to 38, 53, 32, and 27 errors, respectively, for the -1 , 1, 1.5, and 2 jnd stimuli. We thus suspect that the lack of an increase in d' for the -2 and -1.5 jnd stimuli was a ceiling effect due to the very high level of sensitivity for these stimuli even before training, which was in turn apparently due to inaccuracies in calibrating the jnds for a subject across the entire range of frequencies used in the study.

c. Discussion. The results of this experiment indicate that the same distribution of training stimuli that led to a *decrease* in sensitivity for the training region in experiment I can lead to an *increase* in sensitivity if the training regime is changed to a discrimination training task. This is a case of acquired distinctiveness along a category-relevant dimension (see also Goldstone, 1994). Possible implications of this result for neural models of the perceptual magnet effect are treated in the General Discussion (Sec. II).

3. Experiment III

The third experiment was designed to elaborate on the training conditions required to induce the acquired similarity along a category-relevant dimension that was demonstrated in experiment I. The specific question this experiment sought to answer was whether training with only a single exemplar from a category is sufficient to induce decreased sensitivity in its immediate region of acoustic space. It is possible that a listener must experience many exemplars from the same category in order to induce acquired similarity. This scenario makes sense if one takes the view that acquired similarity is a case of learning to ignore differences between exemplars of the same category; if subjects hear only one exemplar of a category, there are no differences between category exemplars to learn to ignore.

Eleven adults participated in the third experiment. One subject's performance did not meet the established criterion, and this subject's results were thus not included in the analysis.

a. Training. This experiment involved a categorization training regime that differed from that of experiment I in only one respect: instead of hearing different exemplars from the training region when performing either a listening or identification trial, subjects always heard the same exemplar, milestone B (see Fig. 2).

b. Results. Figure 6 shows the main results for experiment III. Figure 6

cantly worse at discriminating stimuli within the control region [$t(5) = -2.98$, $p < 0.05$]. Again, general fatigue may have been a factor in this decrease in sensitivity. Subjects also became significantly worse at discriminating stimuli in the training region [Fig. 6(b); $t(5) = -2.04$, $p < 0.05$]. More importantly, the change in sensitivity for the training region was not significantly different from the change in sensitivity for the control region [$t(5) = 0.30$, $p > 0.05$; see Fig. 6(c)]. In

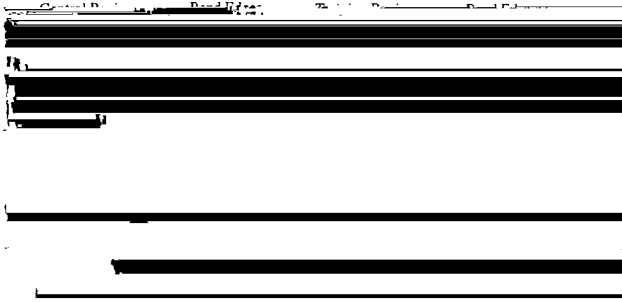


FIG. 7. The training and testing stimuli for experiment IV. Training stimuli were generated in exactly the same manner used in experiment I. Testing stimuli were more closely spaced than in experiments I–III to compensate for increased discriminability of the test sounds due to the shorter ISI and removal of the interstimulus noise burst. See the text for details.

ter characterized as a result of changes in the sensory-trace mode or the context-coding mode of auditory memory. It is usually assumed that increasing the ISI and/or adding a brief noise burst between two stimuli interferes with the sensory-trace mode of memory more than context-coding mode (e.g., Repp, 1984; Werker and Pegg, 1982). Given the relatively long ISI of experiment I and the use of a noise burst between the two stimuli in a discrimination trial, one might reasonably conclude that the effect measured in that experiment primarily involved the context-coding mode of auditory memory. In experiment IV, the ISI during discrimination training was reduced and the interstimulus noise was removed in order to better gauge whether the acquired similarity demonstrated in experiment I is also manifested in the sensory-trace mode of auditory memory.

a. Training and testing. The training and testing stimuli used in experiment IV are shown in Fig. 7. The training regime for experiment IV was identical to that of experiment I, and the training stimuli were generated in the exact same fashion as in that experiment. The testing procedure for experiment IV involved an ISI of 250 ms and there was no distractor noise between the two stimuli (see Fig. 3). In a pilot experiment, it was determined that these manipulations allowed subjects to discriminate the test stimuli almost perfectly. This invalidated the d' measures, since they are only accurate if a significant number of errors are made during testing. In order to obtain an accurate d' measure with the shorter ISI, the stimuli used in the testing sessions of experiment IV had to be more closely spaced than they were in the earlier experiments. Test stimuli for experiment IV were located at 0.75, 1.125, and 1.5 jnd units⁶ above and below the milestones in the control and training regions, as compared to a spacing of 1, 1.5, and 2 jnd units in experiment I. The placement of the milestones and the positioning of the band edges regions were not affected by this change.

b. Results. Figure 8 shows the collapsed d' scores for the control region [Fig. 8(a)] and training region [Fig. 8(b)] before and after training. A significant decrease in sensitivity occurred for both the control region [$t(5) = -5$, $p < 0.05$] and the training region [$t(5) = -3.8$, $p < 0.05$]. The change in the training region was not significantly different from the change in the control region [$t(5) = -0.63$, $p > 0.05$; see Fig. 8(c)]. The G scores for experiment IV are presented in Table V, with the group change in d' reaching significance

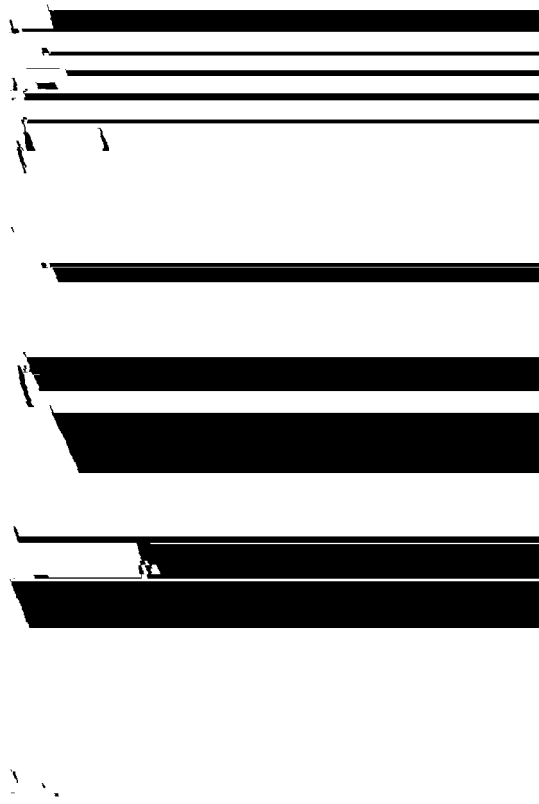
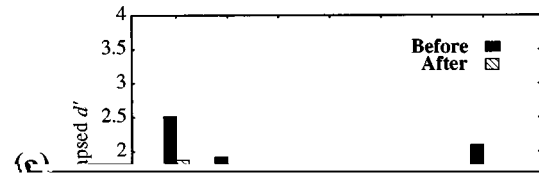


FIG. 8. (a) The collapsed d' scores for the control region of experiment IV, before and after training. (b) The collapsed d' scores for the training region of experiment III, before and after training. (c) Change in sensitivity after training for the control and training regions in experiment IV.

for only one test stimulus (-1 jnd in the training region).

c. Discussion. The results of this experiment indicate that the use of a shorter ISI and no noise burst between the two stimuli in the sensitivity testing trials essentially eradicates the acquired similarity found in experiment I despite the use of the same training regime as in that experiment. Since decreasing the ISI and removing the noise burst presumably favors a sensory-trace memory mode over a context-coding memory mode, this result suggests that the

TABLE V. G statistic comparison for experiment IV. Asterisk denotes statistically significant ($p < 0.05$) changes in sensitivity.

Stimulus (jnd)	Control G score	Training G score
-2	1.56	1.39
-1.5	0.54	1.42
-1	0.23	1.90*
1	0.32	-0.53
1.5	0.31	1.35
2	0.45	0.10

acquired similarity seen in experiment I was primarily associated with the context-coding mode of auditory short-term memory. This result is consistent with the hypotheses of Macmillan *et al.* (1988), Pisoni (1973), Repp (1984), and Werker and Pegg (1992) that a shorter ISI can diminish the categorical nature of the responses made by an observer.

II. GENERAL DISCUSSION

Figure 9 is a composite plot of the total d' measures collapsed across subjects before and after training in all four experiments. The left side of this figure illustrates that the change in sensitivity in the control region due to training in all four experiments was negative, though this change was relatively small and did not reach statistical significance in experiment I. Because the control region stimuli were not presented during training, we suspect that these small negative changes in d' were the result of generally poorer performance in the post-test as compared to the pretest, perhaps due to subject fatigue near the end of the roughly 1.5-h-long experimental session (see Sec. IE 2 *b*).

The right half of Fig. 9 illustrates the d' measures for

in combination with the neural map model of Bauer *et al.* (1996), is schematized in Fig. 10. The left side of the figure corresponds to a categorization training situation, as in experiment I. The top and bottom panels schematize the auditory map as a function of acoustic space before and after training, and the middle panel schematizes the distribution of training stimuli in acoustic space. In categorization training, heavy exposure to a set of training sounds leads to fewer cells coding these sounds in the auditory map, and the resulting smaller cortical representation diminishes a listener's ability to differentiate sounds in this region of acoustic space. This is how the Bauer *et al.* (1996) model, with an appropriate parameter choice that leads to a negative magnification factor for the cortical representation, accounts for the perceptual magnet effect. The right side of Fig. 10 corresponds to a discrimination training situation, as in experiment II. Here, more cells in the map become tuned to the most frequently encountered training stimuli, and the resulting larger cortical representation increases the listener's ability to differentiate sounds in this region of acoustic space. This learning situation corresponds to the "classical" formulation of a self-organizing feature map in the computational neuroscience literature, in which increased exposure to a set

