# Adapting to supernormal auditory localization cues. I. Bias and resolution

Barbara G. Shinn-Cunningham,[a)] Nathaniel I. Durlach, and Richard M. Held
*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

~Received 23 September 1996; revised 27 January 1998; accepted 2 March 1998!

Head-related transfer functions ~HRTFs! were used to create spatialized stimuli for presentation through earphones. Subjects performed forced-choice, identification tests during which allowed response directions were indicated visually. In each experimental session, subjects were first presented with auditory stimuli in which the stimulus HRTFs corresponded to the allowed response directions. The correspondence between the HRTFs used to generate the stimuli and the directions was then changed so that response directions no longer corresponded to the HRTFs in the natural way. Feedback was used to train subjects as to which spatial cues corresponded to which of the

While previous studies provide few quantitative measures of adaptation, many such studies suggest that one of the most important factors affecting adaptation is the type of exposure to the rearranged spatial cues that the subjects receive. In particular, active motor tasks generally yield more complete adaptation than comparable experiments with passive exposure to the rearrangement ~Freedman and Zacks, 1964; Pick and Hay, 1965!. Two different types of feedback are investigated in the current study in order to determine how active motor involvement affects adaptation in our experimental paradigm.

Other studies of spatial perception have shown that the complexity of the visual or acoustic field can affect the perception of source motion. For example, if a single point light source is seen to move around a subject in an otherwise dark room, the subject perceives himself to be stationary and the source to be moving. However, if multiple lights move with a fixed angular velocity around a stationary subject, the subject perceives himself to be rotating within a fixed room ~Lackner, personal communication!. In our studies, results when subjects are presented with an acoustic field ~two ongoing, stationary sources in addition to the target! are compared to results from an experiment in which only the target source is presented.

The degree to which different stimuli can be resolved is determined in part by the range of target stimuli presented in an experiment ~Durlach and Braida, 1969!. The effect of stimulus range on resolution is examined by comparing results in adaptation experiments using a stimulus range of 120 degrees compared to a range of 60 degrees.

Finally, the strength of the cue rearrangement is systematically varied in order to examine how the rate and degree of adaptation depend upon the quantitative strength of the change in acoustic localization cues and the overall range of cues presented.

In the current study, subjects were asked to adapt to supernormal remappings of auditory localization cues. In the experiments described, subjects are first tested with ''normal'' localization cues to yield baseline measures, then with the ''supernormal'' cues to examine how performance changes as subjects adapt. Finally, at the end of each experimental session, subjects are retested with the ''normal'' cues to look for any aftereffects in performance that may result from the training with the supernormal cues. Two quantitative measures were used to track how subject performance changed over the course of the experimental session. Bias ~a measure of response error in units of standard deviation in subject response! was used to measure the degree to which subjects adapted to the supernormal cues. While bias is related to the error in mean response, the measures are not equivalent. In particular, since bias is measured in units of standard deviation, the absolute magnitude of response errors cannot be determined from bias results. As a metric, bias describes the magnitude of response errors relative to the variability in subject responses; thus, a decrease in bias could result either from a decrease in absolute error or an increase in variability. In the current study, bias is examined instead of mean response error in order to quantify the importance of errors relative to response variability. In particular, if response variability is large, the relative importance of a given magnitude error is much less than if response variability is small. The ability to resolve adjacent stimulus positions was measured in order to gain insight into whether better-than-normal resolution is achievable in a localization task using supernormal cues. Estimates of the standard psychophysical metric *d*8 ~again, a measure with units in standard deviation! were found for adjacent stimulus positions to summarize resolution.

## I. SUPERNORMAL CUES

Supernormal localization cues were created in this study by remapping the relationship between source position and normal head-related transfer functions, or HRTFs. Normally, to simulate a source at azimuth *u* and elevation *f*, one simply uses the empirically measured HRTF for that direction, denoted in the frequency domain by $H(v,u,f)$, where *v* denotes angular frequency. In the current study, the correspondence between HRTFs and azimuth values was remapped such that the HRTF used to simulate a source at position @*u*, *f*# was given by

*H*8~ *v*, *u*, *f* ! 5 *H*~ *v*, *f*

this remapping, while two positions off to the side would give rise to more similar cues than are normally heard.[2] As a result, subjects were expected to show better-than-normal resolution in the front and reduced resolution on the side, creating an enhanced ‘‘acoustic fovea’’ toward the front in which supernormal auditory localization could occur. In addition to affecting resolution, however, this transformation was expected to cause a bias whereby sources were perceived farther off-center than were their ‘‘correct’’ locations. If the same family of transformations is used with $n$ , 1, the opposite would be true: sources heard in front of the listener would have smaller-than-normal differences in localization cues and sources off to the sides of the listener would have larger-than-normal differences. The experiments discussed in the current paper used values of $n$ 5 1 ~normal cues, or no transformation!, 2, 3, and 4.[3] The main questions of the study concern the extent to which ~1! bias could be eliminated by subjects over time such that subjects interpreted the new relationship between HRTF and spatial location accurately, and ~

all practical purposes, the two systems are interchangeable in their performance characteristics!. A 486-based PC controlled the Convolvotron and the head tracker. The Convolvotron took monaural input stimuli ~which were amplified, antialiased signals from the sources described above! and created appropriate binaural signals to simulate the stimuli from azimuthal locations specified by the controlling PC. Figure 2 shows a block diagram of the auditory virtual environment.

In most experiments, the binaural signals generated by the Convolvotron were played out through Etymotic Research 3A insert earphones with Bilsom earmuffs worn over the earphones. The combination of insert earphones and commercial hearing protectors helped to block background sounds from the subject during experiments. Taken together, the insert earphones and Bilsom earmuffs reduced sounds reaching the subjects via direct paths by roughly 40–50 dB across all audible frequencies. Experiment $T_1$ was performed before the insert earphones were incorporated into the experimental setup, using TDH-30 circumaural earphones. In this experiment, background sounds were not attenuated well. However, as this caused little difference in the observed pattern of results, this difference is thought to be of little consequence.

A visual display, consisting of 13 lights on a 5-ft-diam arc ~every 10 degrees from 2 60 to 1 60 degrees azimuth!, was located in front of the subjects at eye level throughout the experiments. The lights, labeled 1–13 from left to right, corresponded to the possible locations of the click-train target presented in the experiments ~see Fig. 3!. This visual display was used to present visual, spatial feedback about the simulated auditory sources used in the experiments.

## D. Test procedure

Each subject performed eight identical test sessions over a period of between two to six weeks. Each session lasted roughly 2 h, and consisted of multiple runs separated by two 5-min breaks. Two types of experimental runs were used: localization test runs and training runs, described in detail below. In the training experiments ~$T_1$ and $T_3$!, no feedback was given during localization test runs, but training runs were interspersed with the localization test runs. In the feedback experiments ~$F_3$, $F_{3mid}$, $F_2$, $F_{4a}$, and $F_{4b}$!, correct-answer feedback was given during localization test runs and no training runs were performed ~i.e., the training consisted of giving feedback during the test runs!.

### 1. Test runs

In each localization test run, subjects were presented with a target stimulus simulated as coming from 1 of the 13 possible locations marked by the visual display, chosen at random. Each of the possible locations was presented exactly twice in each run. For most experiments, all 13 positions were employed for a total of 26 trials per run; however, in experiment $F_{3mid}$

turned off, a 500-ms pause occurred, and a new random location turned on. Training runs lasted 10 min each, with a variable number of trials ~usually between 30 and 60 and determined by the speed with which subjects performed each trial! performed in each run. In training runs, exposure to the supernormal transformation entailed an active sensorimotor task ~turning to face the audiovisual target!. When this training method was employed, subjects never received feedback during the testing runs ~and thus received no explicit feedback regarding any errors made during the testing portion of the experimental session!.

#### *3. Run order*

In each session, auditory sources were first synthesized using ‘‘normal’’ HRTFs, then synthesized using the ‘‘supernormal’’ HRTF mapping, then synthesized with the normal HRTFs again. In the training experiments, a total of ten localization test runs were performed in each session: two normal-cue runs, five transformed-cue runs, and then three normal-cue runs. The training runs were performed in between test runs ~see Table I!. In the feedback experiments, 40 localization test runs were performed, one after another. In these experiments, 2 normal-cue tests were performed, followed by 30 supernormal tests, and then 8 normal-cue tests ~see Table II!. Two 5-min breaks were scheduled during each by comparing results from experiments $T_1$ and $T_3$. If a complex sound field allows subjects to extract more information about the cue transformation than does a sound field consisting of a single source, more complete adaptation might be found in experiment $T_3$ than in experiment $T_1$. Comparisons between results from experiments $T_3$ and $F_3$ contrast the effects of active sensorimotor training ~experiment $T_3$! versus correct-answer, cognitive feedback ~experiment $F_3$!. Experiments $F_3$ and $F_{3mid}$ address the question of how the number of stimuli presented affects adaptation and resolution. Different strength transformations were employed in order to gather data that could lead to the development of a quantitative model of the adaptation process ~comparison of results from experiments $F_3$, $F_2$, $F_{4a}$, and $F_{4b}$!. Finally, experiment $F_{4b}$ was identical to experiment $F_{4a}$ except that subjects were exposed to a transformation of strength *n* 5 0.5 after exposure to the supernormal transformation. This final experiment investigated whether exposure to an inverse transformation might allow subjects to readapt to normal localization cues more quickly than without explicit inverse training.

## III. RESULTS AND DISCUSSION

Although many experiments on adaptation to transformed sensorimotor cues have shown that exposure on one day can affect performance on a subsequent day ~e.g., see Welch and Warren, 1980!, no such effects are seen in the current experiments. However, this may be due to the fact that there are too little data to show any significant effects. In any case, any differences from session to session were small relative to the differences within a session. Thus, all the data reported here were combined across the eight identical sessions performed by each subject. This resulted in 16 trials ~2 trials from each of 8 sessions! for each position and run, for each subject.

### A. Analysis

We were interested in estimating how large subject response errors were relative to the variability in subject responses and how well subjects could distinguish stimuli from each other. Two metrics were used to summarize these quantities: bias and resolution.

A maximum likelihood method was used to estimate bias and resolution from the confusion matrices ~pattern of responses observed for every possible physical stimulus! for each subject and run ~combining data across the eight experimental sessions!. This approach assumed that each presentation of a physical stimulus gives rise to a random variable

The ability to resolve adjacent positions depends only on the distance between the means of the corresponding distributions, measured in units of standard deviation in the distributions ~

with circles and solid lines; altered-cue runs with diamonds and dashed lines. The open symbols represent runs prior to altered-cue training while filled symbols correspond to results from the tests following exposure to the altered cues. In all experiments, four runs are plotted: the initial run ~which uses normal cues!; run 3, the first run with altered cues; the final run with altered cues ~run 7 in the training experiments, run 32 in the feedback experiments!; and the first run with normal cues following altered-cue exposure ~run 8 in the training experiments, run 33 in experiments $F_3$, $F_{3mid}$, $F_2$, and $F_{4a}$, and run 37 in experiment $F_{4b}$!.

Independent of the exact training method employed, the range of source positions presented, the number of sources simulated in the acoustic field, and the strength of the cue transformation employed, bias results were similar. First of all, in all experiments, bias results are roughly left–right symmetrical, as expected. Since there is no reason to expect asymmetrical results, the degree of left–right symmetry in the results is one measure of variability in estimates of bias. Results from the first normal-cue run ~open circles! showed some bias, although these errors were much smaller than those found in other runs. A strong bias occurred in the first test with transformed cues ~

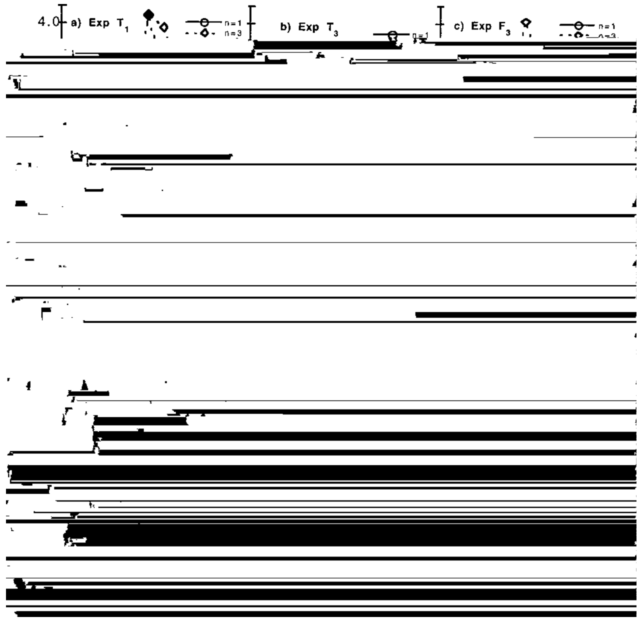


FIG. 6. Resolution results for the seven experiments. In each panel, the average estimated *d*8 is plotted as a function of correct source position ~in degrees!. For each panel, resolution was estimated for each subject and run in the experiment and then averaged across all subjects in the experiment. Four runs are plotted in each panel: the first run in the experiment ~open circles!, the first run with the transformed cues ~open diamonds!, the final run with altered cues ~filled diamonds!, and the first normal-cue run following altered-cue exposure ~filled circles!. In each panel, the legend details the strength of the cue transformation employed in the normal- ~*n* always equals 1! and altered-cue runs ~*n* 5 2, 3, or 4!.

the cues at adjacent positions! or ~2! change with time ~if performance depended upon high-level cognitive factors which were affected by training!.

Resolution results for all seven experiments are shown in Fig. 6. The same four runs are plotted in this figure as were plotted in the bias results in Fig. 5. The initial test with normal cues is plotted with open circles and solid lines, the initial test with altered cues is plotted with open diamonds and dashed lines, the final test with altered cues is plotted with filled diamonds and dashed lines, and the first normal-cue test following altered-cue exposure is plotted with filled circles and solid lines.

As with bias, the basic patterns of results are the same across all experiments, and results are roughly left–right symmetric. Overall, resolution for normal cue runs showed a consistent pattern in which resolution appeared to be slightly *worse* for the center positions compared to positions just off center, rather than slightly better as was expected. This result may be due to positional dependencies on the accuracy of the simulation as well as true differences in resolution arising from perceptual issues. In any case, the results from the initial run using normal cues provides a baseline against which results from the other runs can be measured. As expected for the transformation employed, resolution on the first run using the transformed cues was enhanced for positions in the central region and degraded at the edges of the range compared to results from the initial run. Of particular interest are the results for the final, altered-cue test ~filled diamonds!: although resolution remains enhanced over that achieved with normal cues, there tends to be a decrease in resolution compared to results for the first test using the transformed cues, demonstrating that resolution does not depend solely on physical cues. A similar decrease can be seen when comparing the final normal-cue test ~following training! with the initial, preexposure normal-cue test. There is a tendency for resolution to decrease with training, both for the normal-cue and for the altered-cue results.

Comparing results for experiments $T_1$, $T_3$, and $F_3$ @Fig. 6~a!–~c!#, there is substantially more variability in the resolution results across these three experiments than was seen in the bias results. In particular, the increase in resolution with the first altered-cue run seen in experiment $T_1$ is larger than that seen in experiments $T_3$ and $F_3$. However, in all cases, the estimated value of *d*8 is quite large. Generally speaking, when *d*8 for two stimuli is larger than about 3.0, there are large changes in the amount of overlap of the distributions for relatively small changes in distance between the means of the distributions. As a result, estimates of *d*8 are very sensitive to small changes in the pattern of responses when *d*8 is relatively large. Thus, although the apparent differences between results for experiments $T_1$, $T_3$, and $F_3$ are pronounced, they arise in part from the numerical instability of estimating large values of *d*8. This numerical instability can also be seen in left–right asymmetries in many experiments for some of the large values of *d*8 @e.g., examine the estimates of *d*8 for the middle two positions in experiment $T_1$, particularly for the final test with altered cues ~filled diamonds!#. Results from experiment $F_{3mid}$ @Fig. 6~d!# are roughly consistent with results for experiments $T_1$, $T_3$, and $F_3$. The increase in resolution seen in experiment $F_2$ for the first test with altered cues is slightly smaller than was seen in the first four experiments, consistent with the fact that the transformation in experiment $F_2$ is less extreme than in the first four experiments. Finally, the increase in resolution in experiments $F_{4a}$ and $F_{4b}$ tends to be greater than in the other five experiments, all of which used a less extreme transformation. As with bias, there are no obvious differences in resolution results for experiments $F_{4a}$ and $F_{4b}$, despite the fact that a transformation of 0.5 was used in runs 33–36 in experiment $F_{4b}$.

## IV. CONCLUSIONS AND FUTURE WORK

The results demonstrate that subjects are able to learn remappings between acoustic cues and physical locations in the sense that they are able to reduce bias with training. However, subjects never completely overcome their systematic errors when responding to altered localization cues. Instead, over time, their errors grow smaller in magnitude, but retain the same pattern of results as is seen in their initial errors with altered cues ~i.e., larger errors at the center of the range, smaller errors at the edges of the range!. Since performance was stable by the final test with altered cues, it appears that subjects cannot adapt completely to the transformation employed in these experiments ~shown in Fig. 1!. This result is consistent with previous results investigating sensorimotor adaptation ~Welch, 1986! which show that adaptation usually occurs, but is seldom complete; instead, systematic biases remain even after performance is stable ~and additional exposure causes no further change in localization performance!. A negative after-effect was found for all experiments, implying that changes in performance were not based solely on conscious correction; instead, changes occurred gradually, with training. These gradual changes occurred both when subjects were first exposed to the altered cues ~adaptation!, and when subjects were returned to normal cues at the end of the experiments ~recovery!.

Unlike previous experiments investigating sensorimotor adaptation, our experiments imply that there is no qualitative difference in the final level of adaptation achieved when using training paradigms that involve subjects in active sensorimotor tasks ~experiments $T_1$ and $T_3$! compared to the adaptation achieved in experiments in which simple correct-answer feedback is provided ~experiments $F_3$, $F_{3mid}$, $F_2$, $F_{4a}$, and $F_{4b}$!. In fact, the relative insensitivity of bias and resolution results to the various experimental conditions is somewhat surprising. Bias and resolution appeared to be insensitive to the complexity of the auditory scene, since results from experiments $T_1$ and $T_3$ are comparable. Even when subjects are explicitly trained to an inverse transformation in an attempt to allow their normal-cue test results to return to preexposure patterns more rapidly, no clear effect is seen ~compare results for experiments $F_{4a}$ and $F_{4b}$!.

While changing exposure conditions causes little difference in results, changing the strength of the transformation ~compare results for experiments $F_3$, $F_2$, $F_{4a}$, and $F_{4b}$! and/or the range of stimuli used ~compare results for experiments $F_3$ and F

$df_n(u)/du$ equals one is given by $\|u\| 5 \frac{1}{2} \cos^{2 1} (n^2 2 2n 1 1)/(n^2 2 1)$ . Thus, the range of positions that have larger-than-normal changes in physical cues for a given azimuthal increment are ( 2 35, 1 35) degrees when $n$ 5 2; ( 2 30, 1 30) degrees when $n$ 5 3; and ( 2 27, 1 27) degrees when $n$ 5 4.

[3]In Experiment $F_{4b}$, subjects were exposed to a transformation of strength $n$ 5 0.5 after the normal ''supernormal'' exposure period in order to test whether such retraining might increase the speed with which their performance returned to preexposure patterns.

[4]In the current experiments, only azimuthal position was transformed to achieve better-than-normal resolution. In the future, we hope to perform similar experiments involving elevation and distance. Since the main cue for azimuthal position is interaural time delay ~ITD; see, for example Wightman and Kistler, 1992!, it is likely that similar adaptation results would obtain for experiments involving only ITD transformations.

[5]Since the standard deviation of the internal decision variable is assumed independent of the physical stimulus, nonuniform sensitivity to the physical variable of interest ~i.e., azimuth! implies a nonlinear dependence of the decision-variable distribution mean on the physical stimulus. In particular, the means of the decision variable distributions for sources at 0 and 10 degrees ~sources that are relatively easy to distinguish from one another! will be farther apart than will the means of the decision variable distributions for sources at 50 and 60 degrees ~sources that are harder to distinguish from one another!. This is consistent with other decision-theory models, for instance, see Durlach and Braida ~1969!.

[6]It is possible that subjects change the placement of criteria as a run progresses to reflect the fact that there is less uncertainty in which stimuli will be presented at the end of the run than at the start of the run. For instance, if the position directly in front of the listener is not presented until the last two trials of the run, in principal it is possible that the listener is aware that the last two trials are more likely to come from the center than anywhere else, and may shift his criteria to reflect this fact. In our analysis, we ignore any such effects. In practice, there were enough trials in each run that subjects did not keep track of how often each of the stimuli had been presented during the run.

[7]While this is true in general, mean response for sources at the edges of the response range will not have this property. For instance, in Fig. 4, the mean response to stimulus 3 will always be less than 3, since the left tail of the distribution for stimulus 3 falls into the ''2'' response range and there is no ''4'' response allowed to counterbalance the effect of these ''2'' responses on the mean response to stimulus 4.

[8]Note that, using this simple estimation method, bias values are calculated for each of the $N$ stimulus values presented, not for the $N$ 2 1 criteria. Assuming the same underlying stimulus distributions and Thurstonian decision model, the $i$th bias estimated in this manner reflects ''error'' in the placement of all the criteria in the vicinity of the mean of the distribution of the $i$th stimulus value, while the bias estimated by the maximum likelihood estimate reflects only the ''error'' in the placement of the $i$th criteria.


Auel, J. ~**1980**!. *Clan of the Cave Bear* ~Crown, New York!.

Begault, D. ~**1993a**!te''Call sign intelligibility improvement using a u 66(eouremerdria.-7.465 -1.188 TD [(auditory)-286(display,'')-286(Technical)-286(Report)-286(10401

Begault, D. ~**1993b**!.''Head-up -1.188 itory y.for.trafficito(collision6cavoidCe-7.399 -1.18835,D [(ance)-393(system)-393(advisories:)-3ce)-AS(anc3(prelimi