Perceptual Plasticity in Spatial Auditory Displays

BARBARA G. SHINN-CUNNINGHAM, TIMOTHY STREETER, and JEAN-FRANÇOIS GYSS
Hearing Research Center, Boston University

Often, virtual acoustic environments present cues that are inconsistent with an individual's normal experiences. Through training, however, an individual can at least partially adapt to such inconsistent cues through either short- [Kassem 1998; Shinn-Cunningham 2000; Shinn-Cunningham et al. 1998a, 1998b; Zahorik 2001] or long- [Hofman et al. 1998] term exposure. The type and degree of inconsistency as well as the length of training determine the final accuracy and consistency with which the subject can localize sounds [Shinn-Cunningham 2000]. The current experiments of short-term adaptation measure how localization bias (mean error) and resolution (precision) change when subjects are exposed to auditory cue rearrangements simpler than those previously investigated. These results, combined with those of earlier experiments, suggest that there is plasticity at many different levels of the spatial auditory processing pathway with different time scales governing the plasticity at different levels of the system. This view of spatial auditory plasticity has important implications for the design of spatial auditory displays.

Categories and Subject Descriptors: [**General Literature - General**]—*Conference proceedings*

General Terms: Experimentation, Human Factors, Performance

Additional Key Words and Phrases: Auditory display, spatial hearing, virtual auditory space

1. INTRODUCTION

A basic goal of spatial auditory display research is to provide listeners with cues that allow accurate localization of sound sources. One approach to achieving this goal is to provide the most realistic cues possible so that the stimuli are essentially identical to the cues heard in a real environment. However, the inherent difficulty and expense of such an approach limits the veridicality that can be achieved in a practical, reasonably priced spatial auditory display. An alternative approach is to try to train listeners to accurately localize even when localization cues are different than normal experience [e.g., see Shinn-Cunningham et al. 1998; Zahorik 2001]. Of course, because the type of discrepancies between normal and altered cues will affect how rapidly and completely subjects can adapt (and ultimately, how well they localize sounds), it is important to explore how these factors are influenced by different cue rearrangements.

2. BACKGROUND

With long-term training, subjects can localize accurately even when the acoustic localization cues are inconsistent with previous experience [e.g., Hofman et al. 1998]. However, short-term training experiments suggest that subjects may be able to rapidly adapt only to linear transformations of auditory space [Shinn-Cunningham 2000; Shinn-Cunningham et al. 1998a, 1998b]. In particular, when subjects
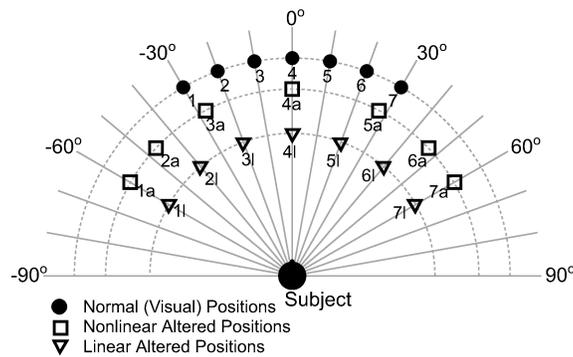
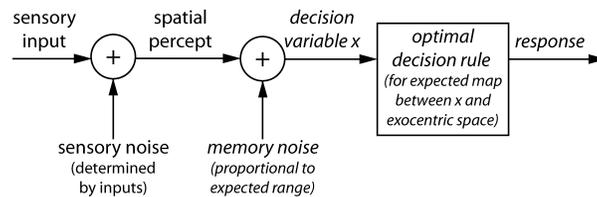Fig. 1.   Nonlinear and linear spatial remappings.

Fig. 2.   Block diagram of model of adaptation.

are asked to learn new, nonlinear associations between spatial cues and exocentric locations, residual errors remain even after performance has asymptoted.

Figure 1 illustrates how cue rearrangements were realized in an example experiment [cf. Shinn-Cunningham et al. 1998a]. Filled circles indicate the seven spatial angles that listeners were asked to choose among in a seven-alternative forced-choice experiment. In "normal" runs, these positions were simulated over headphones using head-related transfer functions [HRTFs; e.g., see Wightman and Kistler 1989] for these locations (from $-30°$ to $+30°$ azimuth, every $10°$). In "altered" runs, the azimuths of the HRTFs used to simulate the sources were displaced laterally (the HRTF angles used are shown by open squares in the figure). Thus, locations 3 and 5 were displaced by nearly $15°$, while positions 2 and 6 were laterally displaced by about $8°$. After training with this complex rearrangement, systematic errors remained that varied with source laterality. These remaining errors are well predicted by assuming that listeners approximate the true rearrangement by linear mapping. This result suggests that if a linear mapping were used, subjects could fully adapt, with no systematic residual error. The triangles in Figure 1 show such a linear remapping, where a source at angle θ is simulated using the HRTF from angle 2θ.

A simple model, shown schematically in Figure 2, was developed that can predict changes in mean localization judgments and spatial resolution in short-term training experiments [Shinn-Cunningham 2000]. The model assumes that internal noise arises both in the computation of spatial position (sensory noise) and at higher levels of the system (memory noise). Short-term adaptation only affects how internal positions are mapped to responses (and memory noise), not how locations are computed; long-term adaptation can affect how spatial percepts are computed. Such a model can account quantitatively for short-term training effects in many experiments [Shinn-Cunningham 2000]. Qualitatively, the model accounts for the fact that resolution is degraded in short-term experiments when novel combinations of interaural and spectral cues are presented simultaneously [e.g., see Kassem 1998]. In order to account

for the data, the model assumes that listeners approximate the new nonlinear relationship between internally computed location and exocentric space with a linear map. In the model, the slope of this mapping decreases exponentially with time and asymptotes at the optimal slope (i.e., the slope that minimizes the mean square difference between the expected and the actual internal percept for each response location). Finally, this change in internal slope directly determines the amount of memory noise affecting the response. These correlated changes in the mapping and internal decision noise completely explain changes in mean response and resolution during the course of the adaptation experiments [Shinn-Cunningham 2000].

The current experiments are designed to directly probe the assumptions of this model by testing subjects with a purely linear remapping (triangles in Figure 1). In contrast to previous experiments, which used nonindividualized HRTFs, the current experiment uses personalized HRTFs, which should, if anything, reduce the sensory noise in the spatial percepts. Such experiments improve our understanding of the limitations of auditory spatial plasticity and provide insight into how to provide robust spatial auditory information that can be accurately perceived by human listeners.

3. METHODS

Seven subjects with normal hearing (confirmed by an audiological screening) performed the experiment. For each subject, HRTFs were recorded in a moderate-sized room for sources at a distance of 1.2 m using a maximum-length sequence technique [Rife and Vanderkooy 1989]. "Pseudoanechoic" HRTFs were created from these measurements by windowing out reflections (a feat possible given the source distances and room size). Short, broadband noise tokens were convolved with the appropriate HRTFs using a Tucker–Davis Technologies (TDT) PowerDAC to generate spatialized stimuli, which were presented over insert earphones (no compensation was used to correct for the spectral characteristics of the playback system). The TDT equipment was controlled from a Windows PC. The PC also controlled a small handheld terminal containing a two-line text display and a number/letter keypad. This device (which connected to the PC via an RS-232 port) provided instructions to the subject and was used by the subject to enter responses.

Each subject performed ten identical adaptation sessions. Each session consisted of 42 runs. The four initial and 10 final runs in each session used "normal" HRTFs. The middle 28 runs used a 2x-linear transformation (triangles in Figure 1). Thus, with altered cues, the range of spatial cues presented is doubled; however, the response range is fixed. Each run consisted of 14 trials, randomly ordered, in which all positions were presented exactly twice. In each trial, a sound source was presented and the subject had to identify (in a seven-alternative, forced-choice paradigm) the position of the source by entering the corresponding number on the handheld terminal. The response positions were always visible on labels affixed to an array of lights positioned in front of the subject. Following each response, the light at the "correct location" was turned on to provide correct-answer feedback. Subjects were instructed that the purpose of the experiment was to see how quickly they could adapt when spatial auditory cues were intentionally distorted. Subjects were not told anything about the manner in which the cues would be altered, but were explicitly reminded when the cues changed from normal to altered and from altered back to normal. In all cases, they were instructed to "do their best" to answer correctly.

4. RESULTS

Following after the analysis performed in similar adaptation experiments [Shinn-Cunningham et al. 1998a, 1998b], responses from each of the 42 runs were combined across all sessions. This processing tracks how performance changes over the course of the experimental session; however, it ignores any effects across sessions. As in previous experiments [Shinn-Cunningham et al. 1998b], in all runs, the mean response location is a linear function of the physical cues presented. Also, like in previous
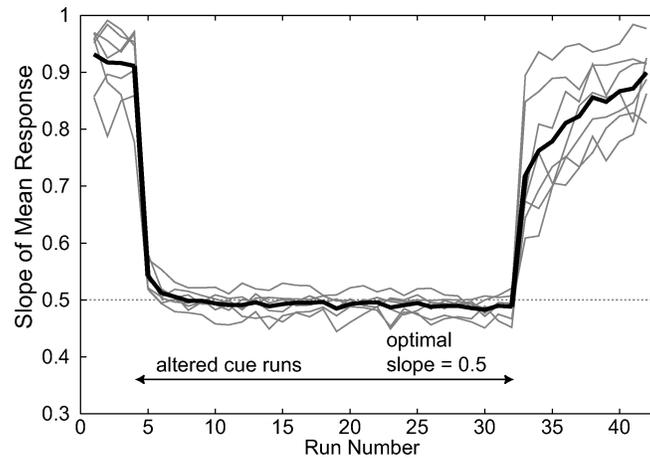
Fig. 3.   Slope of mean response versus HRTF position normal cue position. Individual subjects in gray; across subject average in black.
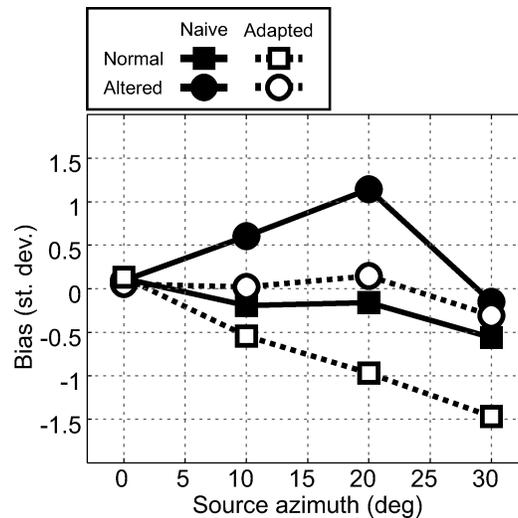

Fig. 4.   Bias (mean localization error in units of standard deviation in the response) as a function of source azimuth. Positive bias indicates subject responses err away from the median plane.

experiments, during the altered-cue runs, the slope describing this relationship decreases as subjects adapt to the larger-than-normal cue range. Figure 3 plots these slopes as a function of run number for both individual subjects (gray) and averaged across subjects (black). Similar to earlier nonlinear cue rearrangement experiments [Shinn-Cunningham et al. 1998b], during the altered cue runs, the slope decreases and asymptotes at the "optimal" slope. Unlike previous experiments, in the current experiment, this optimal slope produces a mean square error of zero, since the rearrangement can be perfectly fit by a line.

Figures 4 and 5 show how the bias and resolution (respectively) of subject responses evolves during the experiment. In order to find these values, the raw confusion matrices, which describe the number of times a subject responded that the source was at location i given that the source was actually from
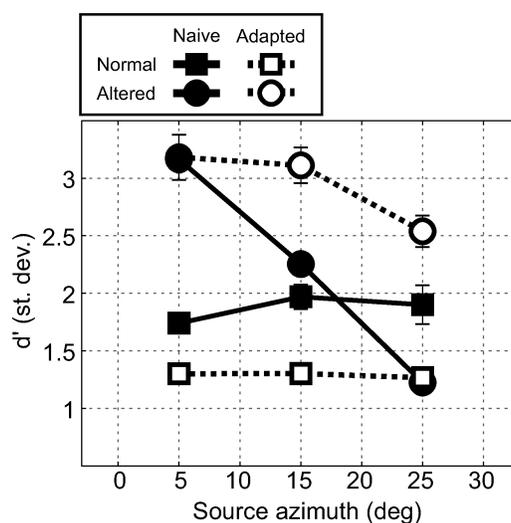
Fig. 5. Resolution (difference in mean responses for adjacent locations in units of standard deviation in responses) as a function of source azimuth.

position j, were analyzed for the final normal-cue run, the initial altered-cue run, the final altered-cue run, and the first normal-cue run following altered cues. From the confusion matrices, the mean response values for each source position were estimated along with the standard deviation in the responses. The difference between the mean and the correct response, normalized by the standard deviation in the responses, measures the bias (Figure 4).

The difference in the mean responses for adjacent locations, normalized by the average standard deviation in responses, estimates d', or the ability of the listener to perceive differences between adjacent source positions (Figure 5). In both figures, responses were assumed to be mirror symmetric and averaged across left/right, as well as across subject. Intersubject standard deviations are plotted as error bars in both figures (too small to be seen in Figure 4).

Bias results follow the pattern seen in previous results of short-term training experiments. Initial bias using normal cues is small. The bias is large when altered cues are first introduced in the expected direction: subjects hear the sources as more lateral than the "correct" response. Bias decreases with altered-cue training, and, unlike previous results, there is no systematic bias at the end of the altered-cue training. This result supports the hypothesis that subjects can completely overcome bias for linear rearrangements of auditory space. Finally, when normal cues are reintroduced at the end of the experiment, an aftereffect is seen: there is a large bias in the direction opposite that is introduced with the altered cues. This aftereffect arises even though subjects are explicitly told when cues change back to normal at the end of the session. Despite this knowledge, subjects make localization errors with normal cues. This result supports the idea that subjects are not using some conscious strategy during the altered-cue phase of the experiment; they cannot "turn off" their adapted behavior at will when returning to normal cues, even though doing so would yield better overall performance.

Figure 5 shows estimates of resolution as a function of source location. In this figure, initial, normal-cue results provide a baseline metric against which other runs can be compared. In fact, this baseline shows that subject resolution in the current experiment is better than in previous, comparable experiments in which a nonlinear transformation was used with nonindividualized HRTFs (see Figure 6). In general, resolution varies from subject to subject; thus, one possible explanation for this difference
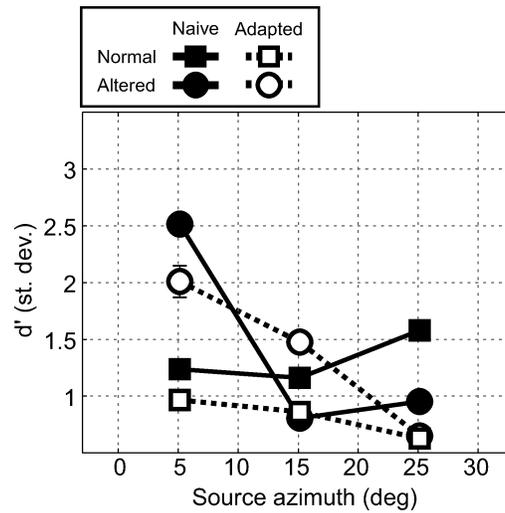
Fig. 6. Resolution (as in Figure 5) from a previous, nonlinear adaptation experiment using a comparable range of stimuli and experimental paradigm, but nonindividualized HRTFs (from [Shinn-Curringham et al. 1998a]).

may be that the subjects performing the current experiment were more skilled than those in the earlier experiment. However, it is interesting to note that individualized HRTFs were used in the current experiment, whereas a set of "generic" HRTFs was used for the subjects in the previous experiment, a fact that may have led to better spatial resolution overall in the current results.

When altered cues are first introduced, resolution increases for the center positions compared to using normal cues. This is not surprising, in that "adjacent" stimuli are actually generated using HRTFs with twice the normal angular separation. However, resolution is poor at the edges using altered cues. This result, which also arises in the results from the previous experiment, is most likely due to edge effects. In the identification task, if the source is perceived as coming from outside the range of allowed responses, the expected mean response for the location will be the edge. Such an edge effect makes it difficult to estimate d' for positions at the edges and this effect is exacerbated when overall resolution is good, as in the current experiment.

Following training with altered cues, resolution improves in the current experiments. This result is inconsistent with results from a number of previous experiments [Shinn-Cunningham et al. 1998a]. It may be that for linear transformations, resolution with altered cues actually improves with training. Further work is necessary to determine whether this effect is repeatable or if is only due to the systematic underestimation of resolution in the initial altered-cue run. If the effect is "real," it calls into question the way in which memory noise changes with adaptation in the model. Finally, when normal cues are reintroduced at the end of the experiment, resolution decreases. This finding is consistent with previous results and can be explained by the preliminary model of short-term adaptation described in Shinn-Cunningham [2000].

5. CONCLUSIONS

Current results confirm that with short-term training, listeners can completely adapt to linear transformations of auditory space. The imposed linear remapping caused large response bias initially, but, after training, no residual errors are found. Results also confirm previous findings that resolution on spatial identification tasks depends not only on the stimuli presented, but also on the state of the

subject. As in previous experiments, resolution for "normal" cues decreased when subjects were trained to expect a larger-than-normal range of stimuli. However, unlike previous experiments, resolution using "altered" cues actually increased as subjects learned to attend to the larger-than-normal range of stimuli. Additional experiments are necessary to examine how resolution is affected by training with linear-cue transformations. In particular, subjects adapted so well to the transformation that it is difficult to estimate resolution. Results were further confounded by the fact that in the first "altered-cue" run, edge effects (inherent in the identification paradigm) made it nearly impossible to discern differences in mean response; this later effect, which is accentuated by the overall better-than-expected performance, causes the resolution to be consistently underestimated for lateral source positions in the initial altered-cue run. In order to test whether resolution using altered cues improves with training obtains generally or is an artifact of the experimental paradigm, additional experiments are planned which (1) use more source locations to improve resolution estimates and (2) employ an analog localization response rather than an identification task to get a more direct measure of the distribution of spatial percepts elicited by different stimuli during the course of the experiment.

6. DISCUSSION

Subjects can adapt to linear transformations of auditory space more completely than to more complex transformations, given relatively short training times. However, they can adapt to very complex rearrangements with sufficient training [Hofman et al. 1998]. Thus, the fact that with short-term training, subjects can only adapt to linear transformations suggests that spatial auditory plasticity occurs at many different stages in the computational pathway. In particular, short-term training may not change how spatial cues, such as interaural differences and spectral cues are computed and combined to form spatial percepts, but only how these percepts are mapped to exocentric space. Such a model implies that internal noise in the spatial percept is fixed over short time scales, a view which may explain the very good performance of subjects using individualized HRTFs compared to those using nonindividualized HRTFs. While previous results have shown that the number of gross localization errors is reduced when individualized HRTFs are used in virtual auditory displays [Wenzel et al. 1993], current results suggest that use of individualized HRTFs also decreases variability in localization judgments. This result is consistent with the idea that the computation of source location is optimized for "normal" experience; presenting unnatural combinations of cues should yield a more diffuse sound image that is harder to localize precisely (an idea also suggested by the results of Kassem [1998]. However, the current experiments were not designed to explicitly test this hypothesis. Further work is necessary to confirm this result. Overall, results of perceptual plasticity experiments provide insights into what aspects of a spatial auditory display are critical to "get right," in order for listeners to localize accurately.

REFERENCES

HOFMAN, P. M., VAN RISWICK, J. G. A., AND VAN OPSTAL, A. J. 1998. Relearning sound localization with new ears. *Nature Neuroscience 1*, 5, 417–421.

KASSEM, S. 1998. Adapting to Auditory Localization Cues from an Enlarged Head. Unpublished M. S., Massachusetts Institute of Technology, Cambridge, MA.

RIFE, D. D. AND VANDERKOOY, J. 1989. Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society 6*, 419–444.

SHINN-CUNNINGHAM, B. 2000. Adapting to remapped auditory localization cues: A decision-theory model. *Perception and Psychophysics 62*, 1, 33–47.

SHINN-CUNNINGHAM, B. G., DURLACH, N. I., AND HELD, R. M. 1998a. Adapting to supernormal auditory localization cues. I. Bias and resolution. *Journal of the Acoustical Society of America 103*, 6, 3656–3666.

SHINN-CUNNINGHAM, B. G., DURLACH, N. I., AND HELD, R. M. 1998b. Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response. *Journal of the Acoustical Society of America 103*, 6, 3667–3676.

WENZEL, E. M., ARRUDA, M., KISTLER, D. J., AND WIGHTMAN, F. L. 1993. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America 94*, 111–123.

WIGHTMAN, F. L. AND KISTLER, D. J. 1989. Headphone simulation of free-field listening. I. Stimulus synthesis. *Journal of the Acoustical Society of America 85*, 858–867.

ZAHORIK, P. 2001. Localization accuracy in 3-d sound displays: The role of visual-feedback training. In *Proceedings of the Advanced Displays and Interactive Displays Federal Laboratory Consortium*.