

LOCALIZATION ACCURACY IN 3-D SOUND DISPLAYS: THE ROLE OF VISUAL-FEEDBACK TRAINING

P. Zahorik

Department of Psychology
University of California - Santa Barbara
Santa Barbara, CA 93106-9660

C. Tam, K. Wang, P. Bangayan, & V. Sundareswaran

Rockwell Science Center
1049 Camino Dos Rios
Thousand Oaks, CA 91360

ABSTRACT

Using an inexpensive headphone-based 3-D sound display, sound localization accuracy was assessed for six listeners, before, during, and after a perceptual feedback training procedure which provided listeners with paired auditory/visual feedback as to the correct sound source position. We show that feedback training markedly improved localization accuracy, with the largest improvements resulting from listener's enhanced abilities to distinguish sources in front from sources behind. Further, these improvements were not transient short-term effects, but appear to last a number of days between training and testing sessions. These results suggest that simple and relatively short periods of perceptual training can effectively mitigate technical deficiencies in low-cost 3-D sound systems due to the use of non-individualized head-related transfer functions.

INTRODUCTION

Recent technological advances have made it possible to accurately represent 3-dimensional acoustic spaces virtually over standard stereo headphones (Wightman & Kistler, 1989a), a technology that holds great promise for display applications where relevant spatial information needs to be conveyed or augmented by non-visual means. Unfortunately, it is often difficult to achieve acceptable spatial localization accuracy with many "off-the-shelf" virtual 3-D sound systems. This is because the spatial processing typically employed by these systems, although based on known acoustical cues to sound source direction and distance, does not tailor the cues to the individual user. It is known that localization error is greatly increased when certain types of acoustic cues -- namely, the spatially-dependant acoustical transfer functions of the external ear and head, which are commonly referred to as head-related transfer functions, or HRTFs -- are not individualized (Wenzel, Arruda, Kistler, & Wightman,

1993). Because providing for true individualized spatialization is a practical impossibility for commercial 3-D sound systems, it is of great interest to explore other ways of improving localization accuracy in such systems. One simple idea that has not been previously examined is the role of perceptual training which provides listeners with paired auditory/visual feedback as to the correct sound source position.

Past results with real sound sources suggest that some form of perceptual training can greatly improve the accuracy of spatial maps in the brain that result from impoverished stimulus conditions in which the acoustical transfer characteristics of the external ears (i.e. HRTFs) are degraded. In an important study by Hofman, Van Riswick, & Van Opstal (1998), the ears of 4 listeners were filled with putty in order to degrade each listener's HRTFs. These listeners were then allowed to continue with normal and uncontrolled interaction with the environment over a period of weeks. By the end of this period, each listener's localization accuracy had nearly returned to the level observed prior to inserting the putty. This suggests that the spatial maps in the brain that relate HRTF patterns to physical locations in the environment had been successfully remapped. From this study, it is not clear what listeners did to facilitate the spatial remapping, although it seems very likely that comparisons with visually perceived space were made.

Because the degraded stimulus conditions used by Hofman et al. (1998) are similar to those experienced in 3-D sound systems using non-individualized HRTFs, the possibility of substantial localization accuracy improvements with perceptual training in such systems seems promising. Here we examine the role of perceptual feedback training on localization accuracy under more efficient and controlled conditions, using a large number of sound source positions presented with unambiguously paired visual targets.

METHODS

Subjects: Six listeners (all male, median age of 29.5 years) voluntarily participated as listeners in the experiment. Three of the listeners were authors of this article and the remainder were recruited from the Rockwell Science Center. All listeners had normal hearing, as verified by audiometric screening.

Stimuli: The auditory test stimulus was a 100 ms duration Gaussian noise burst, presented from one of a variety of spatial positions using a virtual sound source technique. The spatial positions were uniformly distributed around an imaginary partial sphere, with origin at the center of the listener's head. This partial sphere included a full 360 degrees of azimuth, and ± 40 degrees of elevation relative to ear level (see Figure 1 for a display of this coordinate system). The radius of the sphere was 1.5 m.

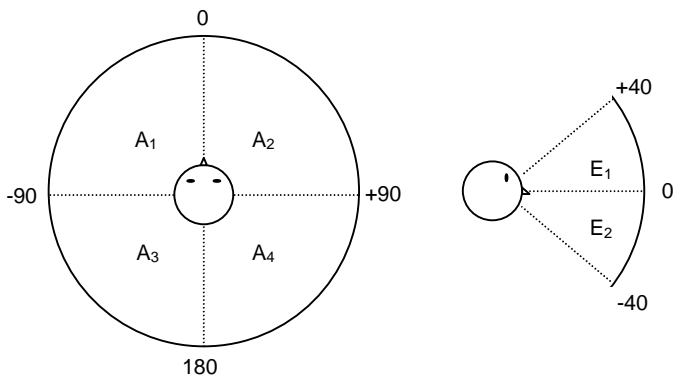


Figure 1. The source position coordinate system for azimuth angles (left) and elevation angles (right) in degrees. Within a block of trials, source position was chosen such that each of the spatial regions (Azimuth A_1 ... A_4 , and Elevation E_1 , E_2) were equally represented. To generate a location within a spatial region, a position was selected at random based on a uniform distribution with a range matching the given spatial region's range.

Two types of visual stimuli were used. The first was a head orientation "cross-hair" that was presented to the listener at all times through a head-mounted display (HMD). This cross-hair marked the vector that pointed straight ahead from the center of the listener's head. The second visual stimulus was one that, in certain conditions, provided the listener feedback as to the correct sound source location. This indicator stimulus was a small point of light with high contrast, also presented via the HMD. When presented, the indicator stimulus was always paired with an auditory stimulus at the same spatial location. This auditory stimulus was identical to the test stimulus, but was repeated at a rate of 1 Hz.

Apparatus: The spatialized auditory stimulus was presented using a standard PC-based 3-D sound card (Diamond Multimedia Monster Sound MX300, Vortex2 chipset) and headphones (Sennheiser HD265). This sound card uses HRTF-based processing for sound spatialization. Since this processing is based on a generalized set of HRTF measurements from a single individual -- an individual that was not one of the test subjects in this experiment -- this display likely suffers from certain performance degradations due to non-individualized HRTF usage (Wenzel et al., 1993). This expectation of degraded performance will actually facilitate investigation of training effects in this setting, however. Visual stimuli (both the cross-hair and visual target) were presented using a PC-based 3-D computer graphics system coupled to a head-mounted display (Sony Glasstron PLM-A35). The position of the listener's head, and the resulting position of the cross-hair, was tracked using a Logitech ultrasonic 6 DOF position/orientation sensor.

Procedure: There were three phases in the experiment: a "pre-test" baseline phase, a training phase, and a final "post-test" phase. In the pre-test phase, each listener's ability to judge the apparent angular position of sound sources was evaluated. No feedback as to the correct sound source position was given in this phase. As such, results from this phase represented a baseline level of localization accuracy for a given listener, using the previously described 3-D sound hardware. The procedure for a pre-test trial is shown graphically in panels 1-3 of Figure 2. At the start of a given trial (panel 1), the listener, while seated in a swiveling chair, oriented to a reference location straight ahead. The cross-hair was visible to the listener via the HMD, and was used to guide the listener to this reference location. Head position in this reference location was verified by the 6 DOF position sensor. Once the listener was correctly positioned at this reference orientation, the auditory stimulus was presented at a given spatial location (panel 2). After the presentation of the auditory stimulus, the listener turned to point the cross-hair in the direction of the perceived sound source location (panel 3). Once the listener placed the cross-hair in a position that he/she felt most properly matched the perceived sound source location, the listener pressed a button to signify the end of the response. During the response phase of the trial, the listener's head position was recorded by the 6 DOF sensor, which allowed for an accurate reading of the final cross-hair position (azimuth and elevation angles). A total of 144 spatial positions were tested, 18 from each of 8 spatial regions. Figure 1 displays these spatial regions, composed of 4 azimuth regions (A_1 ... A_4) and 2 elevation regions (E_1 , E_2). Spatial position within each of these regions was determined by selecting a random location (uniform distribution) within the given region. Within a block of trials, each of the 144

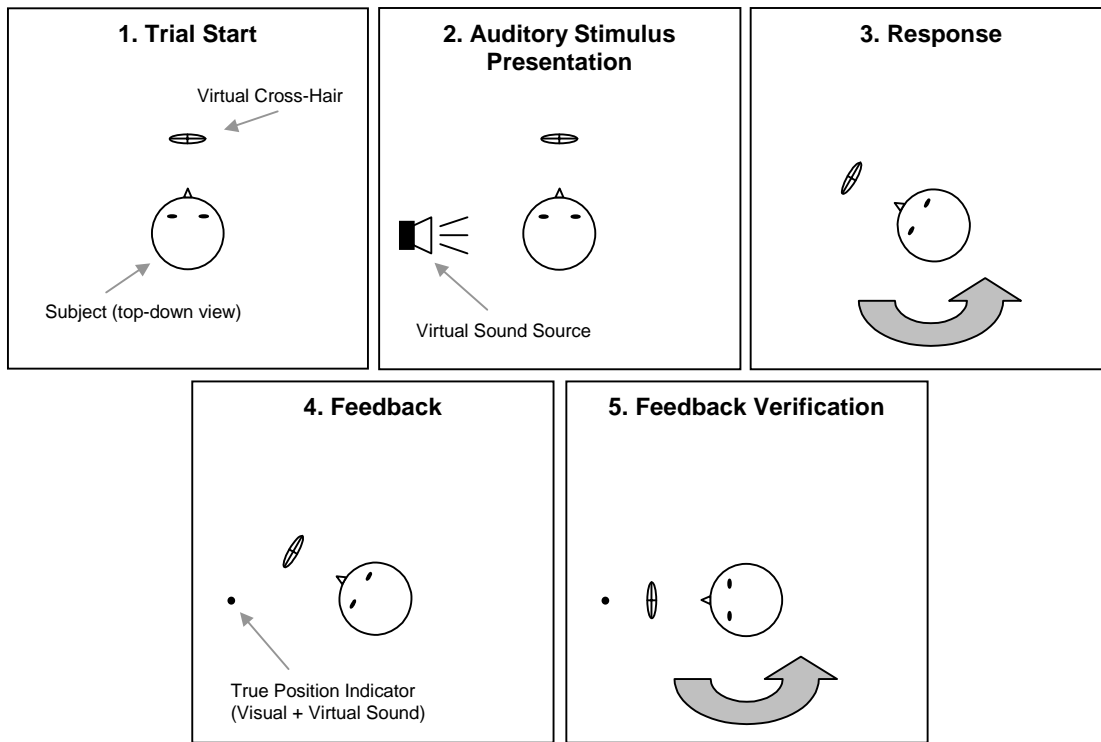


Figure 2. Experimental procedure for a single trial. During the pre-test and post-test phases of the experiment, only steps 1-3 were implemented. During the training phases, all steps were implemented.

spatial positions was presented once in a randomized order.

The training phase of the experiment took place after the pre-test phase. This phase was similar to the pre-test phase, except that visual feedback as to the correct source location was now provided to the listeners. The procedure for a training phase trial is shown graphically in panels 1-5 of Figure 2. The initial portions of a trial were identical to those of the pre-test phase (panels 1-3). After the listener had input his/her apparent position response (panel 3), the feedback portion of the trial began. A visual indicator of the correct source position paired with a repeating spatialized auditory stimulus (the same stimulus as in panel 2, only repeating) was then displayed to the subject via the HMD (panel 4). To verify that the listener was able to find this indicator, the listener was asked to aim the virtual cross-hair (via head rotation) at the position of the correct position indicator (panel 5). When the listener was confident that he/she had pointed to the indicator as accurately as possible, a button was pressed, at which point the cross-hair position was inferred from the measured head position, just as after the source position judgment (panel 3). Hence, this feedback procedure forced the listener to actively orient to the correct sound source position. The selection procedure for sound source positions was similar to that used in the pre-test phase. A total of 24 spatial positions were tested, 3 from each of the 8 spatial regions (see Figure 1). Within a block of trials each of the 24 spatial positions was presented 3 times, in

order to assess judgment variability for the same source position. This yielded a total of 72 trials per block, presented in a randomized order.

The training phase of the experiment lasted two experimental sessions. In each session, the listeners completed a block of 72 trials. After the training phase, listeners completed a final post-test phase of the experiment. This phase was identical to the pre-test phase, and was used to assess lasting effects of the training. The post-test phase was conducted at least 4 days after completing the requisite training phases of the experiment.

RESULTS

The data from all phases of the experiment were transformed to a three-pole coordinate system (Kistler & Wightman, 1992). In this coordinate system, azimuth angle is represented in terms of two angles: a "right | left" angle, which is the angle subtended by the judgment vector and the median plane, and a "front | back" angle, which is the angle subtended by the judgment vector and the vertical plane that passes through the ears. Elevation angle is the same in the three-pole transformation, and is referred to as an "up | down" angle.

Transformed data from pre and post-test conditions are shown for two listeners in Figures 3 and 4. In these figures, apparent source position is plotted as a function of true source position for each of the three angles: right |

left, front | back, and up | down. Perfectly accurate responses would lie on a positive diagonal shown by the dashed line. Error magnitude is therefore proportional to the displacement of data points from this line. For listener SCD (Figure 3), error decreased markedly (particularly in the front | back dimension) between pre and post-test phases of the experiment. For listener SCA (Figure 4),

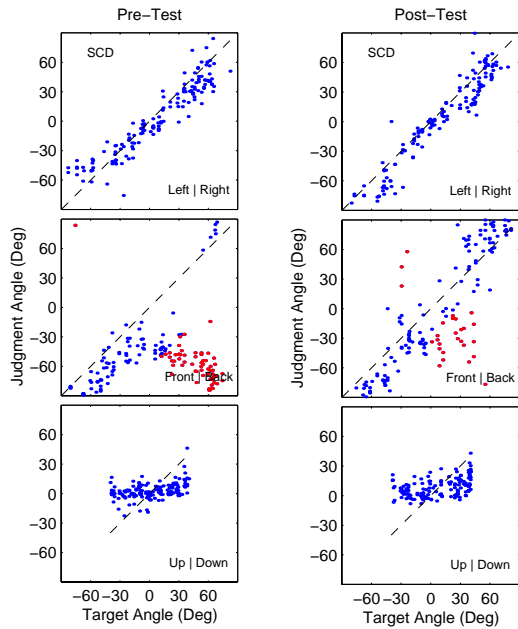


Figure 3. Apparent source position as a function of true source position, for a single listener (SCD), in Pre-Test and Post-Test conditions. The data are plotted in a three-pole coordinate system.

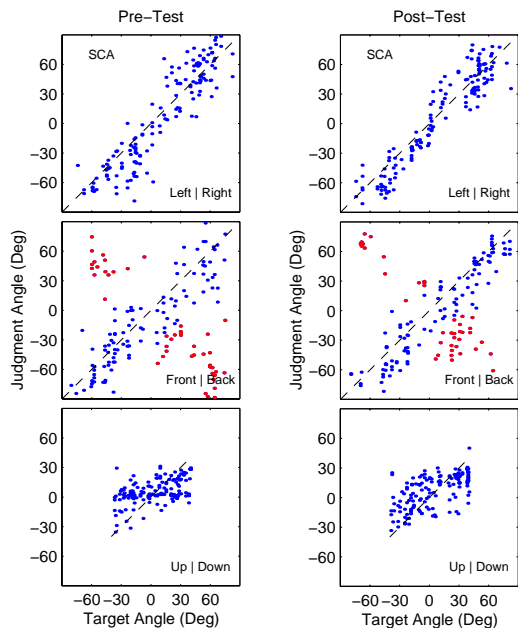


Figure 4. Apparent source position as a function of true source position, for a single listener (SCA), in Pre-Test and Post-Test conditions. The data are plotted in a three-pole coordinate system.

decreases in error magnitude were much more subtle. These listeners represented approximately the best (SCD) and worst (SCA) cases of accuracy improvement within the sample of listeners tested.

To more closely examine the effect of feedback training, we analyzed the time history of response errors throughout the experiment. These results are shown in Figure 5, which plots error magnitude averaged across all listeners as a function of trial number in the experiment. Separate curves are shown for each of the dimensions in the three-pole coordinate system (an explanation of the "resolved" curve is provided below). The most dramatic reduction of error occurred for the front | back dimension, falling from approximately 40 degrees to less than 25 degrees of error. It also important to note *when* the error reduction occurred: namely during the training phases of the experiment, and not during pre and post-test phases, where error remained approximately constant. This result provides strong evidence that the perceptual training alone was effective in reducing localization error in the front | back dimension. Other dimensions showed only minimal error reduction, on the order of a few degrees.

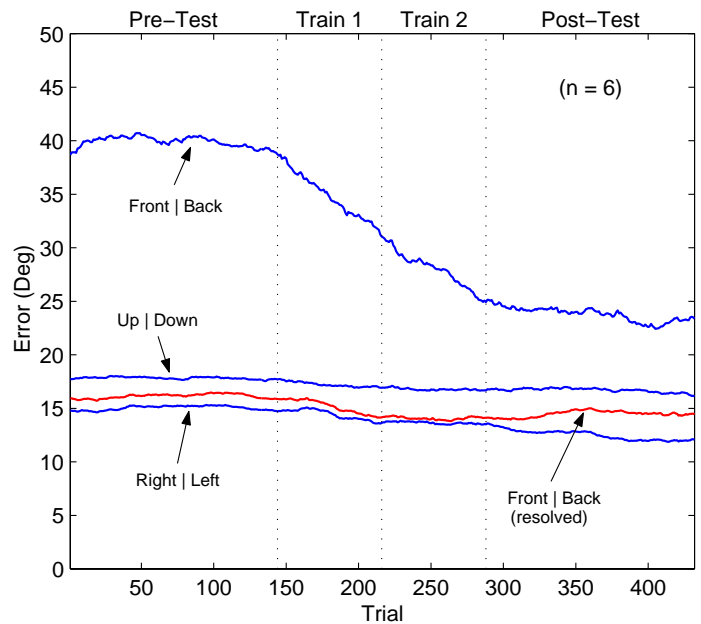


Figure 5. Average error magnitude ($n = 6$) as a function of trial number in all phases of the experiment (Pre-Test, Training 1, Training 2, and Post-Test). The data have been smoothed by a 72-point moving average function.

Of particular interest is the reduction of error in the front | back dimension. Certain types of front | back errors, known as front | back confusions, are often observed when spectral cues to sound source position are insufficient to resolve the inherent ambiguities in the interaural time difference (ITD) cue and interaural level difference (ILD) cue (Mills, 1972), such as when non-individualized

HRTFs are used (Wenzel et al., 1993). Numerous front | back confusions can be seen in Figure 3 (pre-test) as points lying along the negative diagonal. For these points, the listener mistakenly thought that the sound source was behind when it was actually in front. While the rate of front | back confusion can be seen to decrease substantially between pre and post test for listener SCD (Figure 3), this effect is very small for listener SCA (Figure 4). To examine these effects for each listener more closely, we first identified responses that were front | back confusions, using the following selection rule for confusions:

$$|R - T| > |-R - T| > \text{average error}$$

where R and T are front | back response and target angles for a given trial, and *average error* is defined as the unsigned error across all trials in the given condition. Rate of front | back confusion is therefore simply the number of confusions identified within a given condition divided by the total number of trials in that condition. Table 1 displays front | back confusion rates for each listener in the pre and post-test conditions. A statistically significant decrease in front | back confusion rate was observed between pre and post-tests, $t(5) = 2.962, p < 0.02$.

Table 1. Front | Back confusion rates.

Listener	Pre-Test	Post-Test
SLO	0.10	0.06
SCA	0.31	0.28
SCB	0.36	0.27
SCC	0.30	0.24
SCD	0.42	0.19
SCE	0.25	0.17

Because the existence of front | back confusions can distort measures of error magnitude, the error analyses shown in Figure 5 also included a measure of front | back error in which confusions were resolved. In this resolving process, responses that were identified as confusions were multiplied by -1. When confusions were resolved, front | back error was similar to that observed in the other dimensions (Figure 5). This suggests that the reduction in front | back error observed in Figure 5 is primarily a result of decreasing front | back confusion rates through training phases of the experiment.

At least two final aspects of the data should be noted. First, the overall magnitude of errors observed in Figure 5 after the training phases of the experiment ranged from approximately 12 to 24 degrees, which is within the range of error magnitude observed by (Wightman & Kistler,

1989b) for localization using a 3-D sound display with individualized HRTFs. Second, recall that at least 4 days elapsed between the second training session and the post-test. As shown in Figure 5, this period of time does not appear to affect error magnitude, since error observed at the end of Training 2 is nearly the same as that at the start of Post-Test.

CONCLUSIONS

This experiment has demonstrated that a perceptual training procedure that provides listeners with feedback as to true target locations via paired visual and auditory indicators can substantially improve localization accuracy for low-cost 3-D sound displays that do not use individualized HRTFs. Although the amount of accuracy improvement differs from listener to listener, on average the magnitude of response error after training was similar to that observed in other localization studies using 3-D sound displays with individualized HRTFs (Wightman & Kistler, 1989b). The greatest single area of improvement was in the front | back dimension, in which average error magnitude decreased substantially, as did the rate of front | back confusion.

One aspect of these results that may be particularly useful is the lasting effects of the perceptual training. Here we observed that at least 4 days between training and the post-test did not effect accuracy in the least. Past research (Hofman et al., 1998) suggests that perceptual remapping to degraded HRTFs may actually cause two spatial maps to form in the brain, one for normal conditions, and a second map especially for the degraded conditions. It is possible this took place in the current experiment. Although further study is needed to fully examine this hypothesis, such a result would have great practical significance. Listeners could train to use a given 3-D sound display, which in turn would develop a spatial map in the brain for the display that would not interfere or degrade the spatial map used for normal hearing conditions in the real world.

REFERENCES

- Hofman, P. M., Van Riswick, J. G., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, *1*(5), 417-21.
- Kistler, D. J., & Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, *91*(3), 1637-1647.
- Mills, W. (1972). Auditory localization. In J. V. Tobias (Ed.), *Modern Auditory Theory*. New York: Academic Press.

- Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94(1), 111-123.
- Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening: I. Stimulus synthesis. *Journal of the Acoustical Society of America*, 85(2), 858-867.
- Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening: II. Psychophysical validation. *Journal of the Acoustical Society of America*, 85(2), 868-878.