EFFECTS OF INTERFACE TYPE ON NAVIGATION IN A VIRTUAL SPATIAL AUDITORY ENVIRONMENT

Agnieszka Roginska[1], Gregory H. Wakefield[2], Thomas S. Santoro[3], Kyla McMullen[2]

[1]Music and Audio Research Lab, New York University, 35 West 4th St, New York, NY 10012

[2]EECS Department, The University of Michigan, Ann Arbor, MI 48109

[3]Naval Submarine Medical Research Lab, SUBASE NLON, Groton, CT 06349

roginska@nyu.edu, ghw@umich.edu, thomas.santoro@med.navy.mil, kyla@umich.edu

ABSTRACT

In the design of spatial auditory displays, listener interactivity can promote greater immersion, better situational awareness, reduced front/back confusion, improved localization, and greater externalization. Interactivity between the listener and their environment has traditionally been achieved using a head tracker interface. However, trackers are expensive, sensitive to calibration, and may not be appropriate for use in all physical environments. Interactivity can be achieved using a number of alternative interfaces. This study compares learning rates and performance in a single-source auditory search task for a head-tracker and a mouse/keyboard interface within a single source and multi-source context.

1. INTRODUCTION

The use of auditory cues to help navigators explore unfamiliar environments is of ancient origin. Horns have led ships through the foggy seas just as more contemporary portable sound devices have led the blind through urban environments (e.g. [4][5][9]). Similarly, spatial auditory displays can be used to communicate spatial information about a virtual environment to the user. In this type of interface, locations are represented as sound sources, and a user may navigate and explore this virtual environment as they would the more familiar natural environment.

An important factor in fully-immersive systems is the degree to which the participant and the virtual environment interact at the participant's sensorimotor level. Interaction supports the participant's active exploration of their environment through which they become better oriented spatially and can, therefore, navigate more accurately.

One of the challenges with virtual spatial audio is the type of interface used to inject the user into the virtual world. Hardware for sensing head orientation and position has been used extensively as a means to track users. However, there are several issues associated with the use of head-tracking systems – they are expensive, susceptible to calibration issues, require more specialized application development, and can't always be used in all physical environments. In addition, due to the fact that many users are unfamiliar with the interface, the head tracker involves training. An alternate interface is desirable, which doesn't compromise the user experience or performance. Although the sensorimotor integration between changes induced in a spatial audio display by other interfaces' motion may be less natural, we hypothesize that similar performance can be achieved with alternate interfaces and correspond to an equally compelling experience as with a head tracker.

In this study, we propose the use of an *avatar* interface as an alternative. The interface involves a mouse and keyboard as a means to navigate through and interact with an environment. The mouse controls the x/y position of the listener, and the keyboard controls their orientation.

We compare the use of the mouse/keyboard interface to the head tracker interface in a search and navigation task. We focus on comparing human performance in a search task of a single source within a single- or multi-source environment with both interfaces. We explore the differences in the use of the two interfaces. Through a subjective experiment, we look at how participants learn to use these interfaces, the effect each interface has on their performance, and search strategies developed and used by the participant during a search task.

2. EXPERIMENT

An experiment was designed to assess the extent to which auditory search in a virtual acoustic environment (VAE) can be mediated through an avatar interface. The VAE was comprised of acoustic sources arranged along a circle in an otherwise anechoic environment. Participants could move and orient through this environment either directly, by walking and turning their head, or indirectly, by moving the location and angular orientation of an avatar on a computer display presented in a top-down perspective. In what follows, the former will be called *natural mediation* and the latter will be called *avatar mediation* of user position in the VAE.

The task required that participants locate a source in the VAE by moving to the location of that source. To acclimate participants to the apparatus, the experiment was conducted in two phases. During the training phase, a single source was presented during a trial and the participant moved from the center of the circle to the location of the source as quickly as possible. During the test phase, four sources were presented

during a trial and the participant was to move to the location of each source until all four sources were found.

Because it draws upon the standard means by which we, as listeners, navigate through our environment, we hypothesize that natural-mediated search will require fewer trials than avatar-mediated search to reach asymptote for the training phase. Nevertheless, because both forms of mediation engage a common representation of auditory space, we expect that the asymptotic search strategies of each will be similar.

When multiple sources are present, it is not clear how search times should be affected. An increase in the time it takes to locate the first source would be expected if the presence of multiple sources interferes with the cues used to locate any one source. Alternatively, a participant may choose to minimize total search time by using a portion of their first search to establish a general mapping of all the sources before moving to the first source. In the absence of interference or a global strategy, the time it takes to locate the first source during the test phase should be the same as the asymptote reached during the training phase.

Finally, we are interested in whether some users are generally faster than others when performing an auditory search and in the strategies they use. Accordingly, each participant was tested under both forms of mediation. Half the subjects were trained and tested first under natural mediation, before going on to training and testing under avatar mediation, while the other half underwent initial training and testing under avatar mediation. We hypothesize that experience in either modality (natural or avatar mediation) will transfer to the other as evidenced in fewer trials to reach asymptote when shifting to the alternative modality and that there will be a high degree of correlation between fastest and slowest performers across modality.

2.1. Procedure

For both training and test phases of the experiment, a trial began with a source (or sources) positioned randomly along a fixed circle placed horizontally in the 0-degree elevation plane and the participant positioned in the center of that circle. Participants were notified by a diotic auditory cue when they arrived within a fixed radius of the source. During the training phase, a single source was presented and the participant re-centered him or herself after notification to begin another trial. Training continued until a participant's current and past four search times had a standard deviation of 2.5 seconds or less.

The test phase consisted of four sources. At the beginning of a trial, participants were informed which source they should search for first by a diotically-presented four-second sample of the selected source. Following the cue, the four sources were presented and the participant began their search. Upon successfully locating the first source, the sources were turned off, and the second cue was presented, after which the sources were turned back on again. This sequence continued until all four sources were located.

Once a participant finished both the test and training phases for one modality, they repeated the procedures for the alternate modality.

2.2. Apparatus

Avatar mediation was controlled by a mouse/keyboard interface. A mouse controlled the position of the participant in the acoustic space. The left and right arrow keys controlled the yaw of the participant's head, in steps of two degrees. Natural mediation was controlled by a Polhemus Liberty electromagnetic 6DOF system head tracker, with a 240Hz update rate. The sensor was mounted at the center of the headphone band worn by the participant. The tracker emitter was mounted at the end of an arm placed at least 0.5 m above the participant's head. The system was maximally sensitive to within a 1.5-m radius sphere of the emitter, which is similar in size to a CAVE.

Yaw and position were sampled at a rate of 10 Hz to drive a real-time spatial audio system programmed in Matlab using HRIRs obtained from KEMAR using the NSMRL measurement facility [2]. Audio streaming was implemented as follows:

- A Matlab timer was programmed to generate a new frame of audio based on the participant's current position in the virtual environment. The timer called on routines to convolve audio input read from disk using the appropriate yaw-adjusted interpolated HRIRs and adjustment in position-dependent gain.
- Audio was controlled through the PsychToolbox extension of OpenAL by double buffering. The same Matlab timer was responsible for querying the OpenAL source to determine when one of its (two) buffers had finished playing. The next frame of audio was then loaded into the spent buffer and re-queued.

HRIR interpolation was implemented by constructing the minimum phase impulse response of a system whose magnitude spectrum is determined from a log mixture of the adjacent measured HRTFs (sampled every 10 degrees) and convolving the result with an all-phase system using a fractional-delay method.

Stimuli were presented over Sennheiser open ear HD650 headphones. The listening room was a sound-treated standard 4.5m x 7m acoustic research space in the Music Technology program at New York University. Depending on the condition, participants were either seated before the computer console in which a window with a listener icon was displayed or standing in the middle of the room beneath the Polhemus emitter.

2.3. Sources and VAE

Four sources were selected from a publicly available database of audio recordings [7]. Because the present experiment is the first in a broader study on auditory-guided search through multiple-source virtual environments, the sources were chosen to be (1) sufficiently varying in spectro-temporal features, (2) mutually discriminable, and (3) mutually *inconsistent*, e.g., unlikely to be commonly occurring together in naturally occurring acoustic environment. Among the variety of options, we selected recordings of a **typewriter, street crowd**, **brook**, and **electronic sounds**, as might be heard in a piece of computer music. Each recording was between 23 and 80 seconds in duration and repeated continuously.

Stimulus levels were balanced by one of the authors using method of adjustment to achieve equal sensation level by determining the detection threshold of one source in the presence of the other three when presented diotically. These levels were confirmed by informal listening among all authors.

An inverse square law was used to determine the amplitude of the source as the participant moved through the environment. The dimension of the circle in the VAE was scaled so that there was a 13 dB drop in gain for a source that was one diameter away from the participant. Under natural mediation, this scaling created a non-veridical percept as the attenuation within the VAE is much greater than that associated with a 1.5 m displacement. The radius of the acceptance zone for locating a source was approximately 7.5% of the radius of the circle and was chosen to be roughly within the size of a participant's quarter step under natural mediation.

2.4. Subjects

Eighteen paid volunteers participated in the experiment. Half began with training and testing using the natural mediation while the other half trained and tested on avatar mediation first. Training and testing under both modalities took approximately 75 minutes. Each participant completed the experiment in one session.

3. RESULTS

3.1. Training

Training data was obtained from all subjects before each interface was used for testing. During the training period, subjects were presented with a single source and asked to either physically move to (in the natural mediation), or position their mouse (in the avatar mediation) at the location of the source. The search path and amount of time taken for a subject to "find" the source were measured A minimum of ten trials were presented. When ten trials were completed, results were analyzed. If the standard deviation of the last five contiguous trials was less than 2.5 seconds, it was said that the subject had reached optimal performance. If optimal performance was not reached, training was continued until optimal performance was reached.

Figure 1 and Figure 2 contains an example of search times for a subject who began training under the avatar mediation condition. The white bar represents the trial at which optimal performance was reached. The results for this subject show evidence of substantial learning before reaching optimal performance under avatar mediation, but little improvement in performance over time under natural mediation. For most subjects, the natural mediation (regardless of whether it was the first or second training condition) did not exhibit the type of substantial learning demonstrated when training under avatar mediation.

The scatter plot in Figure 3 represents the mean search times, once optimality is reached, for avatar-mediation first (x-marker) and natural-mediation first (circles). The x-axis shows the avatar search times, the y-axis represents the natural mediation search times. In general, there is considerable scatter in performance across subjects with some trend towards fast and slow subjects being such under both forms of mediation. Order of training does not appear to have an effect: prior exposure to natural mediation (or avatar mediation) does not appear to help nor hinder search times achieved under subsequent training. Finally, when averaged over all subjects, there is no significant difference between avatar-mediated and natural-mediated search times.

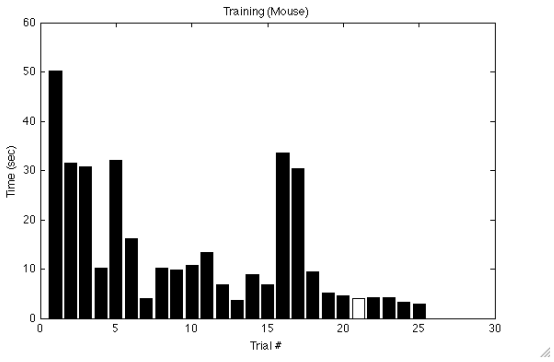

Figure 1 Example of training time results for a subject using the avatar mediation. The subject was presented with the Natural mediation first. The trial marked in white represents when subject has reached optimal performance.

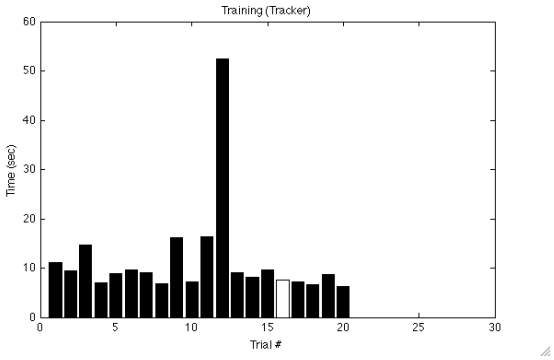

Figure 2 Example of training time results for the same subject as in Figure 1, using the Natural mediation. The trial marked in white represents when subject has reached optimal performance.

Results from the number of trials it took to reach optimality are shown in Figure 4. The scatter plot shows the number of trials for avatar mediation along the x-axis and that for natural mediation along the y-axis. As above, results are shown by the x-markers for those first trained under avatar-mediation, while circles indicated results for those first trained under natural-mediation. Out of the 18 total subjects, 5 showed virtually equal learning times with both interfaces; 7 (4 who used the natural mediation first, 3 avatar first) reached their optimal performance faster using the avatar; and 6 (3 avatar first, 3 natural mediation first) reached optimal performance faster using the natural mediation.

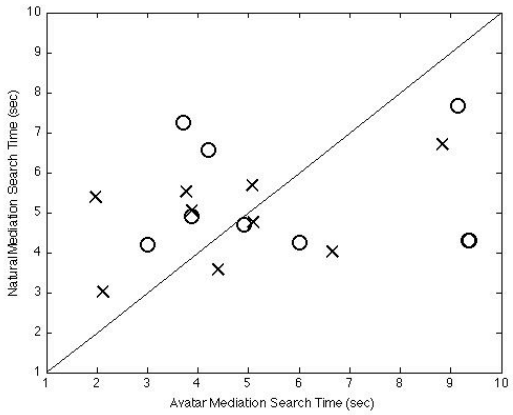

Figure 3 Mean search times are shown for the avatar-mediation first (x-marker) and natural mediation-first (circles) during the training phase.

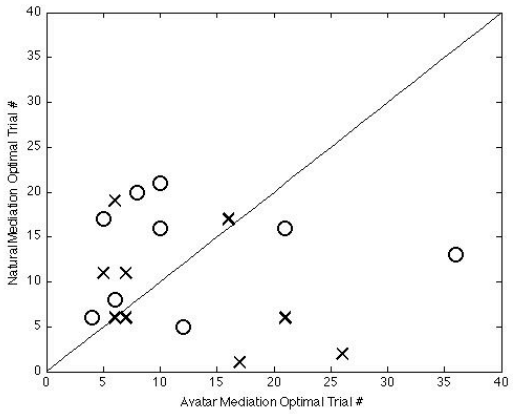

Figure 4 Optimal trial number for the avatar-first (x-marker) and natural mediation-first (circles).

Training results show that the mean search times for all subjects for the natural and avatar mediations are comparable: 5.29 sec for the avatar, and 5.11 sec for the natural mediation. These results suggest that, in the configuration used in this experiment, the type of interface does not play a role on the resulting search time.

The search time does not significantly decrease from the first to the second interface, nor does the number of trials to reach optimal performance decrease from the first to second interface. Based on these two facts, there does not appear to be any transference of performance from one interface to the other.

3.2. Test Results

Results were analyzed separately for (first) target search in the single- and four-source environments. It is beyond the scope of this paper to analyze search times and strategies for all sources in the four-source environment.

When comparing the search time results between the training session and the test trials in the single-source environment, we see similar performance between the two phases of the experiment. Figure 5 compares training search times (x-markers) and the test search times (open circles) for the two interfaces. In most cases, very similar results can be seen during the test trials, as when optimal performance is reached during training. In other words, it appears that once a subject reached a certain level of performance during the training phase, they managed to maintain this level after a break.

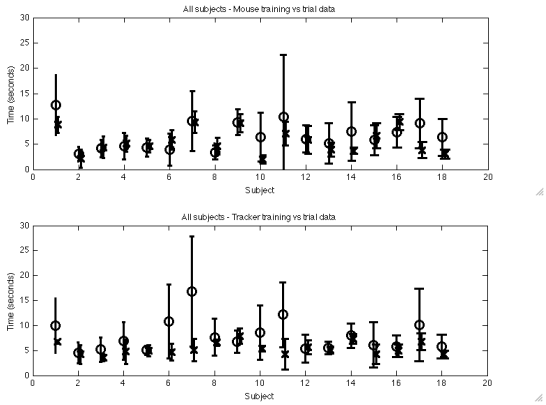

Figure 5 Error bar plot comparing results of training data (x-markers) to test data (open circles) for the avatar (upper) and natural (lower) mediation.

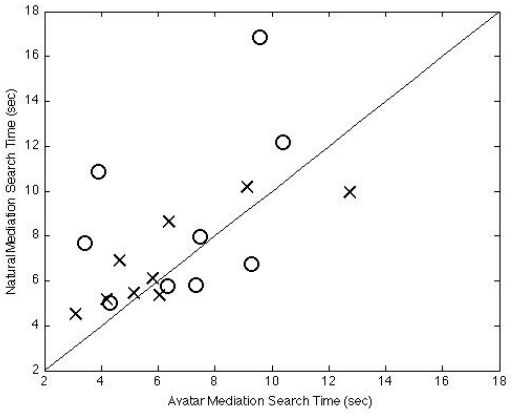

Figure 6 Single-source environment search times for avatar mediation-first (x-marker) and natural mediation-first (circles).

The mean search times during the testing phase for the single-source environment are presented in the scatter plot in Figure 6, for the avatar (x-markers) and natural (open circles) mediation. For many subjects (almost 50%) the search times for both types of mediation for each subject are very similar. A subject tended to spend an equal amount of time finding a single source regardless of whether they used avatar or natural mediation. This is consistent with our results from the training sessions, where we saw a similar search time for both types of mediation. However, when looking at the raw data for all

subjects for the 1-source context, we see an overall increase in search time going from the avatar to the natural mediation. When looking at all subjects, the mean search time for the avatar is 6.2 sec, and 7.8 sec for the natural mediation. Subjects who were presented with the avatar first, show a mean response time of 6.4 sec with the avatar mediation, and 6.9 sec with the natural mediation. Subjects who were presented with the natural mediation first have a response time of 6.9 sec using the avatar, and 8.8 sec using the tracker.

Search times for the first source in a 4-source environment are similar to those for the single-source environment. The mean search time for all subjects increases from 6 sec, using the avatar mediation, to 7.6 sec with the natural mediation. This is confirmed with subjects who were presented with the avatar mediation first, where the mean time is 5.9 sec with the avatar mediation, and 7.2 sec with the natural mediation. Subjects who were presented with the natural mediation first also exhibit a similar increase from 6.2 sec with the avatar mediation to 8 sec with the natural mediation. This overall increase in search times from the avatar to the natural mediation can be seen in the scatter plot in Figure 7.

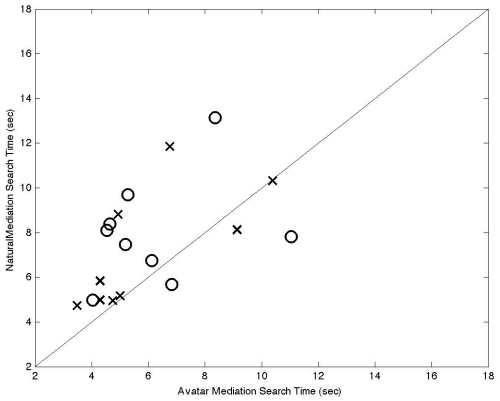

Figure 7 Search times for the first source within the 4-source context for avatar mediation-first (x-marker) and natural mediation-first (circles).

3.3. Search strategies

To further evaluate the subjects' performance using each form of mediation, we analyzed the paths taken by each subject when finding the source. These paths are indicators of the search strategies used by the subject, and give us insight into how well the user's spatial knowledge of the interface is being utilized. In his study of navigation behavior, Tellevik [5] found that a listener's search strategy changes over time as a result of learning. Many virtual environment and spatial cognition researchers (Buechner et. al [1], Hill et. al [3], and Thinus-Blanc & Gaunet [7],) have classified spatial search patterns into those that indicate novice search performance and those which indicate a more experienced search technique. It is by these classification schemes that we have categorized each subject's path data. Figure 8 shows an example from our data of a path that would be classified as a novice (left) strategy and a path that would be classified as an experienced (right) strategy.

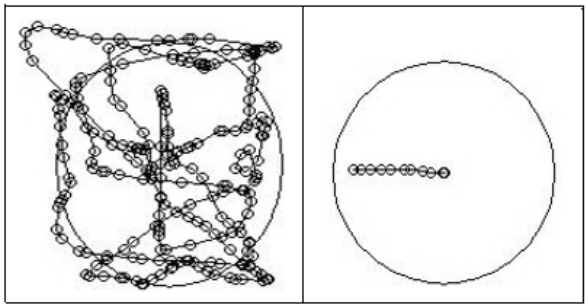

Figure 8 Classification of search strategies. The path on the left is classified as a *directed random* strategy and the path on the right is classified as an *enfilading* strategy.

Figure 9 shows the frequency of usage across subjects of an experienced search strategy during training. Subjects using natural mediation who trained first under avatar mediation exhibited the highest proportion of experienced search strategies. This trend can also be seen in the frequency of usage results during the test phase of the experiment as shown in Figure 10. Subjects using natural mediation, who trained on the natural mediation first also exhibited a high proportion of usage of experienced search strategies, although slightly lower than that of the subjects who trained first under avatar mediation. For the single-source environment, subjects who trained on the natural mediation first, when moving to the avatar mediation, showed a decline in performance. The subjects in this condition began the test, using a high proportion of sophisticated search strategies and later ended the test, using the least proportion of experienced search strategies.

Figure 11 examines the usage of an experienced search strategy to find a single source in the 4-source environment. Here, we can see that performance is very similar under natural and avatar mediation: there is no performance difference in the usage of sophisticated search strategies for either form of mediation.

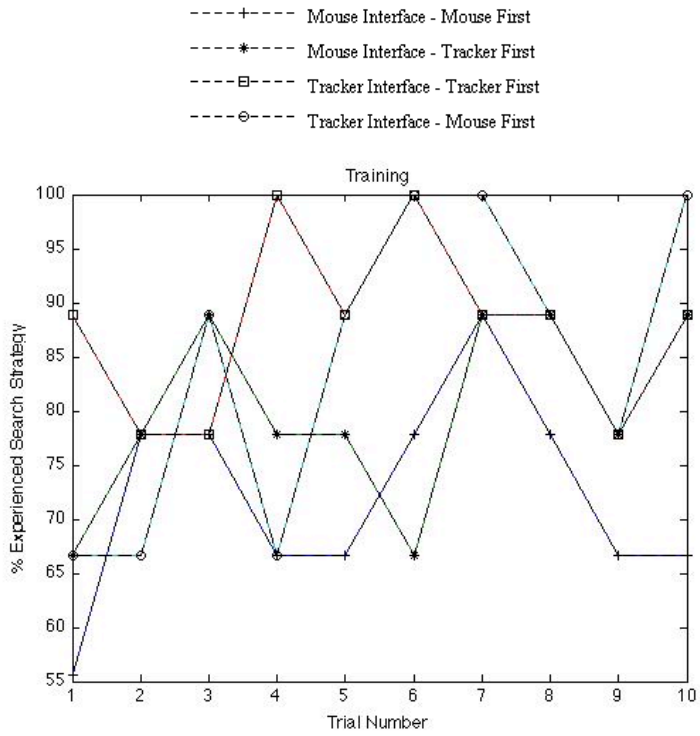

Figure 9 Usage of an experienced search strategy while training

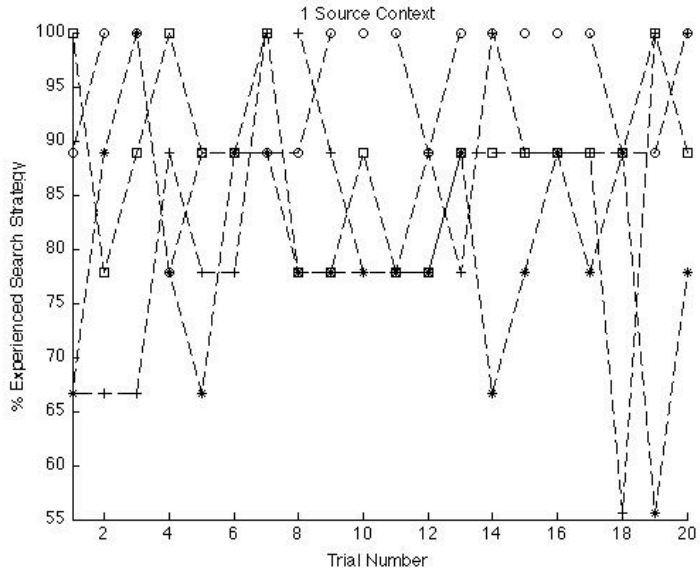

Figure 10 Usage of experienced search strategy in 1 source context environment

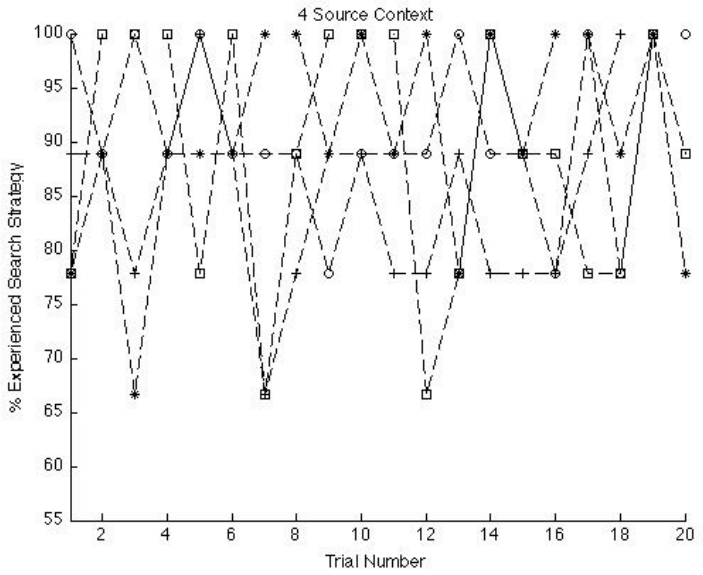

Figure 11 Usage of experienced search strategy in 4-source context environment

4. DISCUSSION AND CONCLUSIONS

This paper presents results of an experiment that compares search times and strategies of an avatar and natural mediation interface when finding one source within the context of a single and four-source auditory environment. Results from the avatar and natural mediation training and test phases of the experiment suggest that, although the training trends for the two interfaces are different, the resulting search times are similar.

Results from the training phase, during which subjects familiarized themselves with each interface and the task, show that in many cases, the number of trials necessary to reach optimal performance with each interface was similar. However, looking closely at the trial data we notice that there is clear evidence of learning to use an avatar to interact with the acoustic environment. . Such steep learning curves were not seen in most subjects under natural mediation. These results are independent of whether the avatar mediation was presented to the subject as the first or the second interface and suggest no transfer of experience across the two interfaces.

The asymptotic search times achieved during training were very similar in both interfaces. During the testing phase we saw an increase in the search times in both the single and multi source context conditions (from 5.29 sec to 6.2 sec with the avatar, and from 5.11 sec to 7.8 sec). Although the training and test search times in most subjects were very similar, in a few subjects we observed an increase in search times in the testing phase of the experiment, which could be due to fatigue. Further testing is needed to validate the cause of the search time increase.

Regardless of the number of sources in the context (one or four sources), results show nearly identical search times. This suggests that the number of sources in the background does not create a distraction for the subject, at least for finding the first source. Further analysis is needed to describe search times for the remaining sources in a multi source context.

Congruent with the search time data, the search strategy data also indicate that there is no clear difference in the quality of a user's search strategy under natural mediation compared to avatar mediation for finding a single source in a four source-environment. Small differences exist in the proportion of experienced search strategies used, while training as well as in the one source environment. Although these differences can be teased out, they are not significant enough to suggest that one form of mediation is significantly superior to another.

The experiment was setup in a room where the physical configuration and the limitations of the sensitivity and range of the tracker used in the natural mediation limited the physical space during testing to a radius of 1.5 meters. Although we have not performed any testing in different sized configurations, we speculate that the size of the effective area during the testing was one of the contributing factors to the similar time scales of the results. Had the area been much larger, the physical constraints of human movement would have most likely produced different results, as it is doubtful, for example, that a virtual acoustic source placed somewhere in a football field would be found by most players in under 10 seconds! The key finding of our experiment is that the only penalty in using an avatar to explore one's acoustic environment is that of learning to use the interface in the first place. Once learned, participants appear to use it as effectively as they would their own bodies in exploring a new acoustic space.

5. ACKNOWLEDGMENT

This work was supported by the Office of Naval Research Award Number N0001409WR20103, 2006-10

6. REFERENCES

[1] Buechner S. J., Hölscher, C. & Wiener, J., (2009) “Search Strategies and their Success in a Virtual Maze”. Proceedings of the 31th Annual Conference of the Cognitive Science Society, 1066-1071

[2] Cheng, C. and Wakefield, G. H. (2001). "Moving Sound Source Synthesis for Binaural Electro-acoustic Music Using Interpolated Head-Related Transfer Functions (HRTF's)," Computer Music Journal, 25(4), 57-80.

[3] Hill E.W, Rieser J.J., Hill M.M., Halpin J., (1993) “How persons with visual impairments explore novel spaces: strategies of good and poor performers.” Journal of Visual Impairment and Blindness, 87(8)

[4] Sandberg S., Hakansson C., Elmqvist N., Tsigas P., and Chen F.. (2006) “Using 3D audio guidance to locate indoor static objects”. Human Factors and Ergonomics Society Annual Meeting Proceedings, 50(4), 1581-1584.

[5] Shoval, S., Borenstein, J. and Koren, Y. (1998) “Auditory guidance with the Navbelt - a computerized travel aid for the blind”, IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, August, 28(3), 459-466.

[6] Tellevik, J.M., (1992) “Influence of spatial exploration patterns on cognitive mapping by blindfolded sighted persons”. Journal of Visual Impairment & Blindness, 92, 221-224.

[7] The BBC Sounds Effects Library. Princeton, N.J.: Films for the Humanities & Sciences vol. 1-40 (1991)

[8] Thinus-Blanc, C. & Gaunet, F., (1997) “Representation of space in blind persons: Vision as a spatial sense?”. Psychological Bulletin, 121, 20-42.

[9] vOICe: http://www.seeingwithsound.com/