THE DESIGN OF MIXED –USE, VIRTUAL AUDITORY DISPLAYS: RECENT FINDINGS WITH A DUAL-TASK PARADIGM

Derek Brock, James A. Ballas, Janet L. Stroup, and Brian McClimens

Naval Research Laboratory
Washington, DC 20375
brock|ballas|stroup@itd.nrl.navy.mil, mcclimen@aic.nrl.navy.mil

ABSTRACT

In the third of an ongoing series of exploratory sound information display studies, we augmented a dual task with a mixed-use auditory display designed to provide relevant alert information for each task. The tasks entail a continuous tracking activity and a series of intermittent classification decisions that, in the present study, were presented on separate monitors that were roughly 90° apart. Using a 2-by-3 design that manipulated both the use of sound in each task and where sounds for the decision task were positioned, the following principal questions were addressed: Can tracking performance be improved with a varying auditory alert tied to error? To what degree do listeners use virtual auditory deixis as a cue for improving decision reaction times? Can a previous finding involving participants' use of sound offsets (cessations) be repeated? And, last, are there performance consequences when auditory displays for separate tasks are combined? Respectively, we found that: Tracking performance as measured by RMS error was not improved and was apparently negatively affected by the use of our auditory design. Listener's use of even limited virtual auditory deixis is robust, but it is probably also sensitive to the degree it is coincident with the location of corresponding visual stimuli in the task environment. On the basis of manually collected head movement data, listeners do make opportunistic use of sound offsets. And, finally, a significant interaction, as measured by average participant reaction time, was observed between the auditory display used for one task and the manipulation of the degree of auditory deixis encoded in the auditory display used for the other task in our paradigm.

1. INTRODUCTION

Auditory display research at the Naval Research Lab (NRL) is predominantly motivated by Navy ambitions to institute optimized manning concepts in its new platforms, founded in large part on advanced automated decision support and human-centric information and interaction technologies. Although new operational uses of sound as information are likely to have a secure place in this foundation, a great deal of basic, applied, and design research remains to be carried out.

2. BACKGROUND

Since 1998, an ongoing series of auditory studies in our lab has focused on the use of sound to manage attention and improve performance in a perceptually demanding dual task that combines a sequence of discrete decision events with a continuous tracking activity. In both of our previous studies in this series [2][3], the decision task was augmented with a set of three easily distinguished sounds that signaled and respectively identified each of the three types of decision events participants were expected to address. Listeners were then expected to examine the events visually and record their judgments. Since the sounds alone did not specifically indicate where to look, we used a generic head related transfer function (HRTF) to spatially correlate the presentation of each sound with the visual location of its corresponding event. The result of this use of auditory deixis [1] in both studies was a significant improvement in our measure of decision reaction times over the same measure when our spatialized sound cues were absent.

3. A MIXED–USE AUDITORY DISPLAY

Following these studies, the auditory display for the dual task was revised and expanded. A spatialized auditory alert whose pitch increased and decreased with operator error was added to the tracking task and a simpler, alternative spatialization scheme for the sounds augmenting the decision task was implemented. These additions allowed us to explore a further set of questions about auditory display-based attention management and to revisit the findings of our most recent study.

The immediate result of combining the new tracking alert with the existing set of decision task sounds was an instance of a mixed-use auditory display—an auditory display in which sounds are used for more than one activity. One of our first considerations was how performance would be affected by this mixed-use design.

3.1. Tracking Alert Issues

To reinforce its signal purpose, the new tracking alert was positioned in the virtual listening space to sound as if it were located in the center of the tracking task's visual display. We anticipated that the alert would both reduce the size of tracking errors and interact with the decision task to improve reaction times. However, we also suspected that the extent of these improvements might not be significant, as there were potential difficulties for this hypothesis in both the way the tracking alert was triggered and in the nature of the tracking task itself. Specifically, the alert was only sounded (i.e., triggered) when tracking exceeded a threshold that was determined in advance for each of the study's participants. It was then looped, with its pitch tied to excess in the error rate, until performance resumed below the threshold. Since the task itself requires constant attention to be performed well, the mismatch between its continuous nature and what was effectively an intermittent, or semi-continuous, alerting paradigm could be an inherently poor design. It was also conceivable that the use of each participant's predetermined threshold for the triggering point would provide little incentive for improving performance. Better tracking performance might correlate with some lower value, a standard deviation, for instance, less than the thresholds that were used.

3.2. Auditory Spatialization Issues

In addition to the tracking alert, another aspect of the mixed-use auditory design we wished to address was the role of auditory spatialization.

Previously, we had always spatialized the sounds used to augment the decision task in an egocentric manner. When decision events appeared in this task's visual display, we mapped their positions relative to the window's centerline to an arc of left to right positions in the virtual sound space in front of the listener. Sounds associated with events wide to the right in the display, for instance, were correspondingly heard to be coming toward the listener wide to the right of straight ahead in the listening space. Events in the display's center and on its left side were similarly heard to respectively approach from straight ahead and from corresponding positions to the left.

Ballas [1] characterizes the directional information conveyed by sounds presented in this or any spatially positioned manner as auditory deixis. Our working premise up to this point has been that the extent of performance gains we have observed in the sound-augmented decision task are due, at least in part, to the deictic component of our egocentric sound spatialization scheme. Under this scheme, each sound conveys at least three things to the listener: the presence of a pending decision event (onset and duration), the kind of event it is (identity), and where the event is located in the task's visual display (deixis). In the present study, though, we decided to examine this information structure more carefully by contrasting our egocentric scheme with the simpler spatialization scheme we alluded to earlier, at the beginning of Section 3—an alternative, screen-centric design that renders the same sounds as if they were located in the center of the decision task's visual display. Since this simpler scheme represents a reduction in deictic information, we reasoned that if the egocentric sound presentations were helping listeners with their visual search, then their performance with screen-centric sounds should be slower. An aspect of this contrast we failed to consider, though, was how the tracking alert's position in the listening space might affect decision performance under each of these spatialization schemes.

3.3. Effort

The most difficult predictions to make with regard to the mixed use of sounds in the dual task were seemingly those involving the performance consequences of combining the two auditory displays. Although we thought a beneficial interaction might result from the use of the tracking alert as measured by reaction times in the decision task, we were divided as to whether this might not come at the expense of greater effort. The significant reduction in head movements we had observed in our most recent study was easy to explain in terms of straightforward auditory monitoring. However, instead of providing an incentive to hasten decisions and encourage better tracking, might not the addition of a signal to return to the tracking task offset the benefits of auditory monitoring by pressuring participants to turn back and forth between the tasks more often, perhaps even more than they must when no sounds are present at all? Here too, our design for the onset of the tracking alert obfuscated the issue.

Also of interest under this heading was the unexpected finding in our prior study of listeners' apparent substitution of sound offsets for visual evidence of their decision entries. Not only did we hope to replicate this effect, we anticipated it would appear whenever the decision task was augmented with sounds

4. METHOD

Although we were able to reuse our previous materials, procedures, and physical setup in the present study, to address the mixed-use questions raised by the new tracking alert and the alternative screen-centric spatialization scheme for decision task sounds, it was necessary to enlarge our experimental design and add new decision task scenarios.

4.1. The Dual Task

The visual displays for the dual task's two activities (shown side-by-side in Fig. 1) were presented to participants in the present study on two computer monitors that were separated by roughly 90° in azimuth. The tracking task was presented on the right and the decision task was presented on the left. The motivation for this particular arrangement comes from the horizontal, three-screen configuration of an advanced, single person, operations desk for command and control the Navy expects to deploy in its next-generation platforms [7]. The performance consequences of this workstation's wide angle of separation between its outer two screens is a particular concern of our attention management studies, since the size of this angle this makes it difficult at best to peripherally monitor one screen while looking at the other.

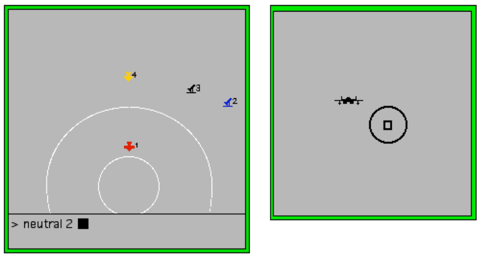

Figure 1. *The dual task's two visual displays, which were separated by almost 90°in the present study. The decision task is shown on the left and the tracking task is shown on the right.*

In the tracking task (shown on the right in Fig. 1), participants pursue a silhouette of a military aircraft with a right-handed joystick that controls a reticle and pipper. The target requires constant visual attention to be tracked well and its evasive behavior is continuous throughout each exercise. Because of the task's general difficulty, participants first train on it as a single task. When they proceed to the dual-task setting, the mean RMS error (average distance to the target) of their final training exercise plus two standard deviations is used as an individual performance benchmark. Instantaneous tracking error above this point is considered poor performance and is signaled visually by coloring the normally black reticle and pipper yellow.

In the decision task (shown on the left in Fig. 1), participants use their left hand to classify instances of three types of potential threats as hostile or neutral based on a set of rules for their visual behaviors. The threats appear as enumerated icons that move down the screen through a set of radar range circles. Each threat's display time is about 25 seconds, and at any point, as many as six icons in either initial, pending, or final dispositions can be onscreen. Initially the icons are black, but after a few moments, an automation component assisting in the classification process paints them one of three colors. The participant is then allowed to enter his or her assessment with two left-handed keystrokes on the number pad of a standard keyboard. Red and blue icons are respectively hostile and neutral and only need to be confirmed; yellow icons are unknowns and must be assessed visually. All icons are painted white as soon as they are classified.

The spatialized auditory displays for both tasks were rendered and served by a Crystal River Engineering Acoustetron. No head tracking was used. For the tracking task, a synthesized bell was sounded as a loop in conjunction with the yellow reticle to signal poor tracking performance (also see discussion in Section 3.1). When listeners faced directly forward, the alert sounded as if it were coming from the center of the tracking window. Instantaneous tracking error above each participant's predetermined threshold was used to parameterize the loop's sounding pitch. Informally, this had the effect of creating an auditory sense of tension and release as tracking respectively worsened and improved. The alert was silenced when performance resumed below the threshold.

The sound loops used to signal each class of event in the decision task were a siren, airplane propeller noise, and a diesel truck horn. Loops started when each event's color assignment was made and ended when decisions were entered, but were only sounded one at a time and always corresponded to the oldest unacknowledged event whenever overlaps occurred. In separate manipulations, participants heard these signals positioned in the listening space either egocentrically or screen-centrically (these presentation schemes are described in Section 3.2). Like the tracking alert, decision event sounds in the screen-centric spatialization scheme were positioned to sound as if they were located in the center of the decision task window when listeners faced directly forward.

4.2. Recording Head Movements

As a measure of effort, participants' head movements were manually recorded by the experimenter on a handheld computer during each dual-task exercise[1]. In this data collection technique, which was also used in the latest of our previous studies [3][4], a template that represents events of interest is used as an overlay on the handheld's screen. As each event occurs, the experimenter taps on the appropriate event and the handheld computer time stamps and records each tap's location. Afterwards, desktop software is used to process the raw data and derive a variety of counts and statistics including event durations and transition frequencies.

In our previous study, we recorded how often participants turned their heads to look at five locations in the physical setup: to the computer monitors on the left and the right, to a monitor in the center, and to the decision response keyboard on the left and to the right-handed joystick. Although the center monitor was not used in the present study, we monitored the same locations again and used this data principally to derive counts of the number of times participants directed their attention from one location to another.

4.3. Experimental Design:

Eighteen NRL staff members (5 women and 13 men), ranging in age from 18 to 49, volunteered to participate in the study. Participants first trained to perform the two tasks separately and then carried out six dual-task exercises under different treatments within a two-factor, 2-by-3, repeated measures design. The first factor manipulated the use of the tracking alert (two levels: alert off vs. alert on) and the second factor manipulated the use and spatial presentations of sounds for the decision task (three levels: sounds off vs. egocentric sounds on vs. screen-centric sounds on). Treatments were presented to participants in diagram-balanced order and, independently of this, each dual-task exercise was successively scripted by a different scenario of 65 decision events

5. RESULTS

Our principal measures of performance were RMS error in the tracking task, counts of participants' head movements when carrying out the dual-task, and decision task reaction times. Each of these will be discussed separately.

5.1. Tracking Error

To evaluate the effect of the study's manipulations on tracking performance, we first normalized each participant's average RMS error in each dual-task exercise to compensate for individual differences. Our method for doing this was to divide the difference between each participant's mean tracking error in a given treatment and his or her predetermined mean (see Sections 3.1 and 4.1) by the standard deviation of the predetermined mean.

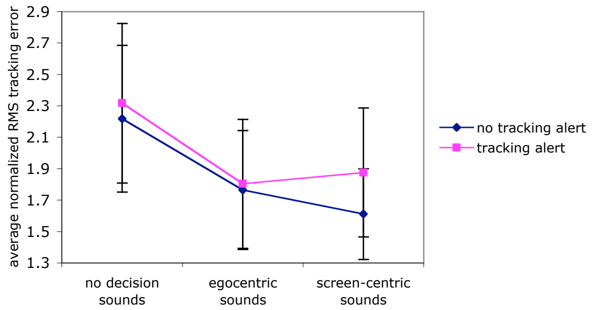

Figure 2. *Interaction plot of average normalized tracking error in the study's six conditions.*

A two-factor, within-subjects ANOVA showed a main effect only for the decision task sound manipulation ($F(2, 16) = 3.937$, $p < .041$). Pairwise comparisons showed that differences between the estimated marginal means of this factor's three levels (no decision sounds, egocentric sounds, and screen-centric sounds) were only significant between the no sound and the two sound treatments (no decision sounds – egocentric sounds = .484, $p < .012$; no decision sounds – screen-centric sounds = .524, $p < .046$). In other words, tracking performance was significantly better when the decision task was augmented with sounds, but relatively insensitive to the method of spatialization.

From the interaction plot shown in Fig. 2, it is apparent that our expectations for the tracking alert regarding tracking error—that average error would be reduced—did not materialize. In fact, the alert may have even worsened tracking performance: in addition to the lack of a main effect, average normalized tracking error was consistently greater whenever the tracking alert was present. The alert's impact on participants' reaction time performance in the decision task is examined in Section 5.3 below.

5.2. Head Movements

Much like our tracking error result, a two-factor ANOVA of total head movements only showed a main effect for the decision task sound manipulation ($F(2, 16) = 24.868$, $p < .0005$). A similarly robust effect on this measure was seen in our previous study and was expected here. Pairwise comparisons of this factor's estimated marginal means show that, as is the case for participants' tracking error, the main effect is due only to the use of sound (no decision sounds – egocentric sounds = 105.750, $p < .0005$; no decision sounds – screen-centric sounds = 115.222, $p < .0005$) and not due to the use of either of the spatialization schemes.

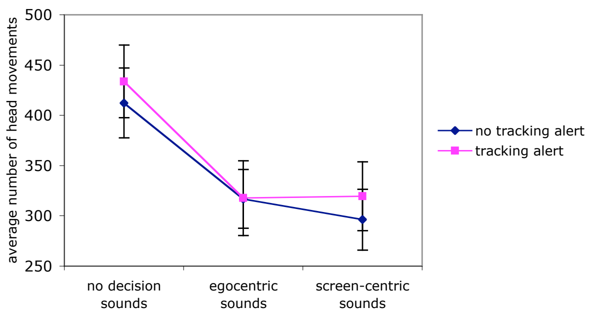

Figure 3. *Interaction plot of the average number of participant head movements in the study's six conditions.*

The marginal effect of presenting the tracking alert proved to be a pervasive but slight increase in overall numbers of head movements. The pattern, which can be seen in Fig. 3, is qualitatively similar to the pattern of our tracking error data. While neither trend is significant, their similarity is notable and suggests that additional effort correlates with greater tracking error.

Counts of participants' head movements (transitions) from one location to another in each of the dual-task exercises are shown numbered by condition in Table 1. As in our previous study, counts of movements to and from the joystick and the center position between the left and right task displays were negligible, so these data are not included (out of 37,730 events, a total 6 head movements were directed to these positions). We will refer to the counts in this table by condition number for the remainder of this section.

5.2.1. *Conditions 1 and 2*

The counts in conditions 1 and 2 are of immediate interest since these conditions are equivalent to the two conditions in our previous study in which we unexpectedly found evidence that listeners were occasionally substituting decision task sound offsets for visual confirmation of the effect of their decision responses.

In condition 1, participants heard no sounds whatsoever, and in condition, 2 they heard only the egocentrically spatialized decision task sounds. As we found in our previous study, a comparison of the respective marginal totals in these conditions in Table 1, particularly those for turns away from the tracking task on the right (3246 in condition 1 and 2417 in condition 2), confirms that participants readily used the egocentric sounds as an effective way to reduce the effort needed to perform the decision task. In addition to affording a systematic reduction in their head movements, this auditory strategy also allowed participants to devote more attention to, and thus measurably improve, their performance in the tracking task (see Fig. 2).

Cond. 1: tracking alert OFF / decision sounds OFF

	>Right	>Left	>Keybd	
Right>	0	3127	119	3246
Left>	3062	0	488	3550
Keybd>	184	422	0	606
	3246	3549	607	7402

Cond. 2: tracking alert OFF / egocentric sounds ON

	>Right	>Left	>Keybd	
Right>	0	2296	121	2417
Left>	2185	0	478	2663
Keybd>	234	366	0	600
	2419	2662	599	5680

Cond. 3: tracking alert OFF / screen-centric sounds ON

	>Right	>Left	>Keybd	
Right>	0	2174	108	2282
Left>	2095	0	414	2509
Keybd>	189	333	0	522
	2284	2507	522	5313

Cond. 4: tracking alert ON / decision sounds OFF

	>Right	>Left	>Keybd	
Right>	0	3314	140	3454
Left>	3213	0	490	3703
Keybd>	242	388	0	630
	3455	3702	630	7787

Cond. 5: tracking alert ON / egocentric sounds ON

	>Right	>Left	>Keybd	
Right>	0	2278	139	2417
Left>	2222	0	460	2682
Keybd>	194	404	0	598
	2416	2682	599	5697

Cond. 6: tracking alert ON / screen-centric sounds ON

	>Right	>Left	>Keybd	
Right>	0	2361	90	2451
Left>	2302	0	442	2744
Keybd>	148	384	0	532
	2450	2745	532	5727

Table 1. *Head movements (transitions) from one location to another in the study's six conditions. Throughout, the tracking task was displayed on the right, the decision task was displayed on the left, and the keyboard was positioned for use with the left hand.*

The only transition in condition 2 that is not less than or essentially equal to its counterpart in condition 1 is the count of turns from the keyboard to the tracking display on the right. These counts (184 in condition 1 and 234 in condition 2) replicate the effect seen in the equivalent conditions in our previous study and thus support our contention that listeners can and do make use of sound offsets as a strategy for improving their performance in the dual task.

How this works is relatively straightforward. Presumably, participants look at the keyboard to be sure of their key presses. They must then look up at the decision display on their left if they wish to see that their response had the intended effect. But, in the context of condition 2, the corresponding sound loops also cease, and so, offer an alternative, eyes-free way to verify the effect of responses. If listeners were not making use of this alternative verification strategy, we would expect the qualitative relationship between the counts in question to be similar to that of the other counts or, put another way, for the proportion of keyboard-to-right transitions to be essentially the same in both conditions. Calculated for each condition as a percentage of the marginal total of turns away from the keyboard, these proportions are, however, quite far apart: 30.36% in condition 1 and 39.0% in condition 2.

Despite having replicated the previously observed qualitative relationship between corresponding transition counts in conditions 1 and 2, a number of other important pairwise relationships in Table 1 remain to be evaluated. In addition to addressing other relevant differences in these comparisons, we will continue to focus on evaluations of the keyboard-to-right transition, since we expected to find additional evidence of participants' use of decision event sound offsets in these counts.

5.2.2. Condition 3

In condition 3, listeners again heard only decision event sounds, but instead of hearing them presented in condition 2's egocentric scheme, they heard them presented screen-centrically. A comparison of this condition's marginal totals with those in condition 1 shows that in this condition too, participants readily made use of the sounds to reduce their effort. Again, the reduction in head movements is systematic, and, in fact, is greater than in any of the other conditions. This trend is also mirrored in Fig. 2's tracking data, where an even further, though non-significant, reduction in RMS error makes participants' tracking performance in this condition best overall.

All of condition 3's transition counts are smaller than their counterparts in condition 1, with the exception of the keyboard-to-right transition. However, unlike condition 2, here the keyboard-to-right transition is very nearly equal to condition 1's. The diminished size of this count, relative to the difference between conditions 1 and 2, could be taken as evidence against participants' use of the sound offset monitoring strategy in this condition. In fact, part of the explanation for this count may be the reduction in auditory deixis that was introduced by the use of the screen-centric spatialization scheme. In condition 2, listeners could be relatively confident in their substitution of sound offsets for visual checks of the effect of their responses because the egocentric spatialization scheme provided them with a useful form of information redundancy: sounds on the left corresponded to icons on the left in the decision display, sounds on the right corresponded to icons on the right, and so on. In the screen-centric scheme, though, this redundancy was removed. The sound onsets were still just as valuable (perhaps even more so in taking the form of an unambiguously located source: note that all of condition 3's transition counts are smaller than condition 2's), but the sound offsets lacked the disambiguating position information that made the offset monitoring strategy both effective and reliable in condition 2. Lack of redundant position information, then, implies that participants had to rely on visual checks in condition 3.

However, the full explanation may not be as straightforward. As it also happens, condition 3's proportion of keyboard-to-right transitions, calculated as a percentage of its marginal total of turns away from the keyboard, is 36.21%. Since this is almost 6% larger than the same proportion in condition 1 and we would expect it to be much closer in size if, in fact, participants were not making use of sound offsets, we now have evidence that suggests that they may have used this auditory response verification technique after all. Thus, it is likely that participants made some use of the offset strategy in condition 3, but not as much as in condition 2 due to the reduction in auditory deixis that was imposed by the screen-centric spatialization scheme.

Whether or not this is the full explanation, there are also the trends toward fewer head movements and smaller tracking error in condition 3 to be explained. Though not statistically significant, the pattern of these trends and the fact that they are the best instances of these two measures in the study suggest that an additional factor may have been at work in condition 2 that was not present in condition 3. Since the only difference between the two conditions was the spatialization scheme that was used, it is necessary to consider the possibility that unintended looks to the tracking task due to onsets of wide-to-the-right sound events in condition 2's egocentric scheme are also part of its keyboard-to-right counts. Since this could also bump up the screen-to-screen counts, as any unintended, sound-prompted looks to the right would naturally prompt looks back to the left, the result would be higher transition counts, which is what was observed in condition 2. Further, there appears to be a positive correlation between the number of head movements in each condition and the size of participants' tracking error, which can easily be seen by comparing the plots in Figs. 2 and 3. Since all sounds in condition 3 came from the left, there were no auditory prompts that could have resulted in unintended looks to the right. Hence, there were fewer head movements overall and thus better tracking performance. Although this explanation weakens the case for the extent of participant's use of the sound offset response verification strategy in condition 2, it also does nothing to rule it out.

5.2.3. Condition 4

The introduction of the auditory tracking alert in condition 4 had the effect of making all but one of the transition counts in this condition greater than their counterparts in condition 1. The simple and most plausible explanation for this effect is that the operation of the tracking alert gave participants license to attend to the decision task and the keyboard more frequently. The effect was compounded by the alert's repeated onsets, which prompted participants to return their attention to the tracking task. The fact that condition 4's keyboard-to-left transition count is smaller than its counterpart in condition 1 bears the latter part of this explanation out.

Since there were no decision task sounds in condition 4, it almost need not be said that the fact that condition 4's keyboard-to-right transition count is greater than its counterpart in condition 1 cannot have the same explanation as this transition's greater size in condition 2 (nor in condition 3). Without an auditory display for decision events, as was also the case in condition 1, operators must adopt a taxing strategy of repeatedly glancing away from the tracking task to check on the status of the decision task. Periodically, these glances must take the form of relatively sustained attention in order to carry out the demands of observing icon behaviors and making decision responses. In condition 1, the sum of this attention away from the tracking task and vice versa accounts for participants' significantly poor tracking performance and long reaction times. Whatever wobbly equilibrium of divided attention participants were able to achieve in condition 1's sound-free setting was evidently further destabilized by the seemingly haphazard onsets of condition 4's tracking alert and the urgency of its pitch shifts. In fact, all of our measures for condition 4 (tracking error, head movements, and reaction times) are the most extreme in the study.

5.2.4. Condition 5

Except for the additional use of the tracking alert, the relationship between conditions 5 and 4 is conceptually similar to the relationship between conditions 2 and 1. The utility of condition 5's decision task sounds for reducing effort is readily apparent. For the most part, its transition counts are smaller than their counterparts in both conditions 1 and 4. The exceptions are the right-to-keyboard transition (larger than in condition 1 but essentially equal with condition 4), the keyboard-to-right transition (slightly larger than in condition 1), and the keyboard-to-left transition (larger than in condition 4). Also worth noting is how close condition 5's marginal totals and transition counts are to condition 2's, with the exception of the keyboard-to-right and keyboard-to-left transitions. As in condition 2, the decision task sounds in this condition were egocentrically positioned, so it is reasonable to ask why the replicated keyboard-to-right transition effect in condition 2 was not repeated here.

The clearest explanation is that it is the result of a conflict, or overlap, of virtual positions in the auditory listening space, which is a consideration we somehow overlooked in the design phase of our study. This conflict was caused by the virtual location of the tracking alert, which you may recall from Section 3.1 was positioned to sound as if it were coming from the center of the tracking display. Assuming that participants made use of the redundant position information in the egocentric decision sounds as a strategy for obviating visual checks of their decision responses in both conditions 2 and 5, we are forced to consider the possibility that in condition 5 the simultaneous sounding of both the tracking alert and sound loops corresponding to decision events on the right was sufficiently confusing to undermine participants' confidence in sound offsets as proof of the effect of their responses for these particular events. Since there was no chance of confusing the redundant position information carried by decision sounds presented on the left, meaning the offset strategy could be used for responses to these events, it would be reasonable to conclude that condition 5's keyboard-to-right and keyboard-to-left transition counts should fall somewhere between those of conditions 1 and 2, and in fact they do. Further consequences of condition 5's overlapping virtual sound positions will be discussed in Section 5.3.2.

5.2.5. Condition 6

In conditions 5 and 6, listeners performed the dual task with mixed-use auditory displays. The advantage they enjoyed in condition 6 was that sounds intended for each task were unambiguously separated from each other: sound loops for the decision task were heard to come from the monitor on the left and the tracking alert was heard to come from the monitor on the right. One result of this aural configuration is that all of this condition's transition counts are smaller than those in condition 4 and even in condition 1. Although its left-to-right and right-to-left transition counts are higher than the same counts in conditions 2, 3, and 5, its counts of transitions to and from the keyboard are notably smaller than all but those in condition 3. Clearly participants found it less necessary overall to look at the keyboard in conditions 3 and 6, but why? Part of the explanation maybe that the use of a single, left-handed location for presenting decision task sounds helped to secure or stabilize participants' sense of their left hand position, resulting in less of a need to watch their key presses.

However, a subtle consequence of the screen-centric scheme's reduction in auditory deixis is probably also at work here. Because of the many-to-one correspondence between the depiction of decision events and the location of their associated sounds in this scheme, it was arguably more difficult to anticipate specific sound onsets in the two conditions in which it was used, 3 and 6, than it was with the egocentric scheme used in conditions 2 and 5, where an operator could at times see, for instance, that a widely positioned decision event with an obvious classification was about to be sounded, and so could return to the tracking task, wait for the sound, and then turn directly to the keyboard. Slight differences in the respective percentages of right-to-left and right-to-keyboard turns between conditions 2 and 3, and 5 and 6, bear this out: in the screen-centric conditions, participants proportionally turned from the tracking task to inspect the decision display more often, and from the tracking task to address the keyboard less often. The effect of this bias is more pronounced between conditions 5 and 6, and we assume this is due to the same factors that explain condition 4, namely, that the tracking alert justifies more attention to the decision task, but its onsets also prompt more looks back. Add to this the weakness of the sound offset verification strategy in the screen-centric scheme, and it follows that participants in conditions 3 and 6 were forced to rely less on auditory monitoring strategies for decision information and more on inspecting the decision screen before and after their key presses, and so made fewer keyboard-to-right transitions (only 148 in condition 6).

5.3. Reaction times

Reaction times for decision events were measured in ms from the point at which icons in the decision task display first changed color to the point at which participants made the second of the two key presses required for decision responses (see Section 4.1). A two factor ANOVA of the resulting means showed that differences in the decision task sound manipulation and in its interaction with the tracking alert were significant (respectively, $F(2, 16) = 12.655$, $p < .001$ and $F(2, 16) = 5.714$, $p < .013$). As is the case with our other measures, there was no main effect for the tracking alert manipulation.

The performance advantage for this measure offered by both of the sound spatialization schemes used in this study is readily apparent in Fig. 4. Improvements over performance in condition 1 associated with the use of decision event sounds and no tracking alert were respectively 19.68% and 16.78% under the egocentric and screen-centric spatialization schemes. (The improvement observed in our previous study, with egocentric presentations, was 18.94%.) When the tracking alert was also used, the respective improvements were 13.52% and 20.06%. Although pairwise comparisons of the decision task sound treatments showed no significant difference between the egocentric and screen-centric manipulations, the trends in this data and the conspicuous interaction have, for the most part, quite intuitive explanations.

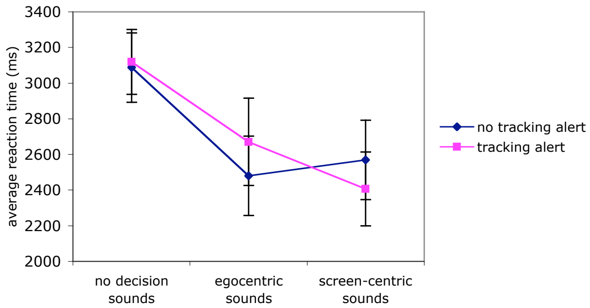

Figure 4. *Interaction plot of mean participant reaction times in ms in the study's six conditions.*

5.3.1. The Role of Auditory Deixis in Conditions 2 and 3

In Section 3.2, we noted that going into the present study, our premise regarding sound-assisted performance gains in the decision task was that their extent has been due in some part to the deictic information that is added to decision event sounds in the egocentric spatialization scheme. Thus, our purpose in devising the screen-centric spatialization scheme was to test this presumption explicitly. By design, the alternative scheme reduced deixis for all decision event sounds to a single, but relevant, task-directed dimension. If our premise was correct, listeners were using the corroborating position information in the egocentric presentations to speed their visual search for decision events, and without this information in the screen-centric mode their visual search would be unassisted. Even if this disadvantage were only on the order of 50 to 100 milliseconds per search, the cumulative effect of would be longer average reaction times [6]. Though it lacks statistical power, this surmise appears to be supported by the difference in average reaction times that can be seen between conditions 2 and 3 in Fig. 4. In fact, this difference is only 89.57 ms, which falls comfortably into the conservative range proposed here.

However, there may be other factors at work in this result that could be confounds for our test. Since we have no eye tracking data for this task, our notions about assisted and unassisted visual search must continue to be viewed through the lens of our data. If we return to our discussion in Section 5.2.2, though, about why tracking performance and counts of head turns were better in condition 3 than in condition 2, you will recall that we noted there that onsets of wide-to-the-right sounds in condition 2's egocentric presentation scheme may have had the unanticipated consequence of prompting unintended looks to the right. The deeper implication of this possibility is that there is an inherent, if subtle, perceptual processing cost for listeners in our egocentric mapping scheme. It is well established that goal-directed saccades of the visual system to spatially coincident auditory-visual targets are 10 to 50 ms faster than saccades to strictly visual or auditory stimuli and that this improvement diminishes "as the spatial and temporal separation of the stimuli increases"[5]. Since our egocentric mapping scheme effectively enlarges or magnifies the visual field of decision events, which is about 8 inches wide, into a roughly 150° arc of virtual positions in front of the listener, it is clear that a significant percentage of sounds presented in this way in a given dual-task exercise are spatially separated from the icons to which they are intended to correspond. Listeners in the egocentric setting not only have to map wide-of-the-mark sounds back into the visual display, but also have to cope with the innate impulse to look in the direction that these sounds appear to be coming from. Conversely, in the screen-centric setting, all sound presentations are approximately coincident with their visual counterparts. Hence, in our study, responses in condition 2's egocentric setting probably do not enjoy a pronounced auditory-visual benefit, but their visual search probably benefits from the spatialization scheme's additional deictic information. In contrast, saccades to the decision task in condition 3's screen-centric setting may be consistently faster, but responses are probably sufficiently slowed by the lack of fine-grained position information to make their subsequent visual search relatively expensive, especially when multiple events are onscreen. If this analysis is correct, the result of our test, then, is that while a benefit of virtual auditory deixis does appear to be demonstrated, the measures of average reaction time in these two conditions are closer to each other than we might initially expect.

5.3.2. The Tracking Alert Interaction

Participants' average reaction times in conditions 1 and 4 are notably close to being equal (see Fig. 4). Although we have already seen in Section 5.2.3 that the introduction of the tracking alert in condition 4 resulted in the study's grossest tracking error and its largest counts of head movements, it seems not to have affected participants' performance in the non-auditory-cued version of the decision task. This is an interesting result for two reasons. First it suggests that participants' non-cued visual search times for decision task events were stable or about the same in both conditions, and second, it suggests that participants had no trouble segregating the purpose of the tracking alert from the decision task. In spite of the fact that the tracking alert appears to have been a poor design, this second inference mirrors the apparent ability of participants to segregate the screen-centric spatialization scheme from the non-cued tracking task in condition 3.

As we noted earlier in Sections 3.2 and 5.2.4, though, the mixed-use auditory display participants heard in condition 5 proved to have a presentation conflict that we overlooked in our study's design phase. An illustration of the overlapping use of the listening space that caused this conflict is shown in Fig. 5.

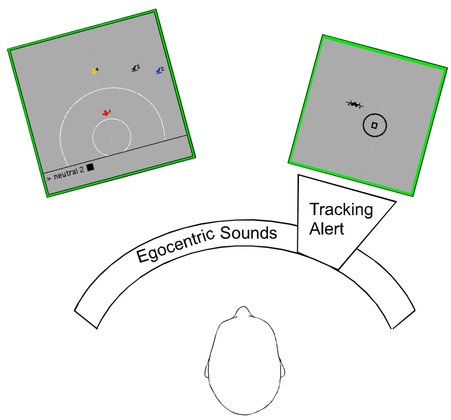

Figure 5. *A schematic illustration of the overlap in virtual auditory positions used by the egocentric decision event sound scheme and the tracking alert.*

In Section 5.3.1, we argued that a significant percentage of egocentrically presented sounds add a small perceptual processing cost to the listener's job of translating auditory information into the visual locations of decision events in the form of inefficient saccades. Onsets of the tracking alert in the middle of the right arc of the egocentric presentation scheme in condition 5, though, very likely added a much greater processing load. Because of the overlapping use of the listening space in this condition's auditory display, participants not only had to adapt to the slightly counterintuitive mapping of decision event sounds that were wide-of-the-mark, but also had to decide which task the sounds presented on their right were intended for. Since this additional disambiguation problem was not so much perceptual as deliberative, its cost could easily have resulted in the longer average reaction time that was observed in condition 5 as compared to the cost of visual search incurred in condition 3. Would this have resulted in more head movements and greater tracking error? Probably not, unless listeners decided to abandon their ears, or rather, their auditory cognition, and resort to visual inspections to resolve their uncertainties—a strategy that intuitively seems more effortful because of the head movements involved than just thinking quickly. In fact, judging from how the data points for these measures in condition 5 compare with condition 2's in Figs. 2 and 3, it is apparent that under the circumstances, participants in condition 5 actually coped quite well: they may have had the least improved average reaction time, but they countered this with relatively stable tracking performance and very little additional effort.

In contrast, the tracking alert's effect on reaction times in condition 6 resulted in the best average decision task reaction time in the study. The explanation for this appears to be that participants in this condition made proportionally more looks, in both directions, between the decision task and the keyboard than in any of the other conditions, which demonstrates a greater degree of attention to the decision task. But what motivated this and how does it square with our explanation of average reaction time in condition 3, the other condition in which listeners heard screen-centric decision event sounds? As we concluded in Section 5.2.5, participants in this condition were probably less able to rely on the range of auditory monitoring strategies that seem to have been used under other circumstances, particularly the use of offsets, and so, were forced into a pattern of inspecting the decision window more frequently both before and after looks to the keyboard. In addition, the proportion of looks from the tracking task to the decision display is greater in this condition than any but condition 1. Thus, participants were motivated to be acutely aware of the state of the decision display and were better prepared to resolve ambiguous, screen-centric onsets of decision event sounds than in condition 3, where, counterintuitively, the lack of a tracking alert resulted in fewer screen-to-screen head movements, and consequently, better tracking performance.

6. CONCLUSIONS

With this third study in our ongoing program of auditory display research at NRL, we set out to address several important questions about the design, role, and use of auditory information as a technique for managing attention and reducing effort in single-operator, multi-task decision environments. Using a complex dual task that involves continuous tracking and intermittent decisions with a mixed-use auditory display, we looked at the following issues: Can tracking performance be improved with a varying auditory alert tied to error? To what degree do listeners use virtual auditory deixis as a cue for improving decision reaction times? Can a previous finding involving participants' use of sound offsets be repeated? And, last, are there performance trade-offs when auditory displays for separate tasks are combined? In summary, we found first, that tracking performance as measured by RMS error was not improved and was apparently negatively affected by the use of our auditory design; next, that listener's use of even limited virtual auditory deixis is robust, but it is probably also sensitive to the degree it is coincident with the location of corresponding visual stimuli in the task environment; third, that on the basis of manually collected head movement data, listeners apparently do make opportunistic use of sound offsets, especially those whose implicit position information corroborates the intended effect; and last, a significant interaction, as measured by average participant reaction time, was observed between the auditory display used for one task and the manipulation of the degree of auditory deixis encoded in the auditory display used for the other task in our paradigm. The circumstances of this interaction suggest that competitive, overlapping use of the virtual listening space in the design of mixed-use auditory displays should be avoided.

This work was sponsored by the Cognitive, Neural, and Social Science & Technology Division of the Office of Naval Research under Work Order N0001404WX20235.

7. REFERENCES

[1] J.A. Ballas, "Delivery of information through sound," in G. Kramer (Ed.), Auditory display: Sonification, audification, and auditory interfaces," Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, MA, 1994.

[2] J.A. Ballas, D. Kieras, D. Meyer, D. Brock, and, J.L. Stroup, "Cueing of display objects by 3-D audio to reduce automation deficit," in *Proceedings of the 4th Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment*, Patuxent River, MD, 1999.

[3] D. Brock, J.L. Stroup, and J.A. Ballas, "Effects of 3D auditory cueing on dual task performance in a simulated multiscreen watchstation environment," in *Proceedings of the Human Factors and Ergonomics Society 46th annual Meeting*, Baltimore, MD, 2002.

[4] D. Brock, J.L. Stroup, and J.A. Ballas, "Using an auditory display to manage attention in a dual task, multiscreen environment," in *Proceedings of the 8th International Conference on Auditory Display*, Kyoto, Japan, 2002.

[5] B.D. Corneil, M. Van Wanrooij, D.P. Munoz, and A.J. Van Opstal, "Auditory-visual interactions subserving goal-directed saccades in a complex scene," *Journal of Neurophysiology*, vol. 88, pp.438-454, 2002.

[6] W.D. Gray and D.A. Boehm-Davis, "Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior," *Journal of Experiment Psychology: Applied*, vol. 6, no.4, pp. 322-335, 2000.

[7] G. Osga, "21st century workstations: Active partners in accomplishing task goals," in *Proceedings of the Human Factors and Ergonomics Society 44th annual Meeting*, San Diego, CA, 2000.

---

[1] We have found that the error rate for manual coding of head movements is low enough (~0.5%) to counter the much greater cost of gleaning the same information from an automated source such as a head-tracking device.