VIRTUAL AUDITORY CUEING REVISITED

Derek Brock, Brian McClimens, and Malcolm McCurry

U.S. Naval Research Laboratory,
4555 Overlook Ave., S.W.,
Washington, DC 20375 USA
derek.brock@nrl.navy.mil

ABSTRACT

In a 2002 dual-task experiment involving opposing screens, virtual auditory cueing significantly improved measures of performance and reduced the effort needed to pursue both tasks. An effort to model this result revealed that supplementary empirical information was needed and a new study, reported here, was subsequently carried out. In addition to focusing on modeling issues, the new study also investigated the contribution of both an augmented auditory reality (AAR) style of display and aurally based event identity information. The previously observed benefits of auditory cueing were replicated, but more importantly, neither AAR-based cueing nor the removal of aural identity information meaningfully impacted performance. These findings suggest that simpler auditory information designs for visual attention may, in fact, be preferable to richer designs, and that aural overlays on visual information are unnecessary, but not disadvantageous, in single-use auditory displays.

1. INTRODUCTION

The outcome of four manipulations in a 2002 human performance study involving a dual task displayed on opposing screens demonstrated that virtual (3D) auditory cues can both significantly improve performance and reduce the effort otherwise needed to carry out both tasks [1]. Critical measures in this study included decision response times for secondary task events and switches of attention between tasks. A subsequent effort to comparatively model the baseline and one of these manipulations in the EPIC cognitive architecture [2], using these and derived measures as criteria, revealed that further empirical knowledge is needed for a comprehensive account of how auditory cues assist performance in this operational paradigm. As a consequence, a new dual-task performance study was carried out in 2009 and is reported here.

In addition to addressing modeling issues, the new study specifically investigated the contribution of both an augmented auditory reality (AAR) style of display [3] and aurally conveyed event identity information (c.f., [4], [5], and [6]) through the incorporation of additional manipulations. Interest in both of these latter questions is motivated by the expectation that virtual auditory displays will eventually be routinely used for, or will supplement, more than one visual information and/or decision task at a time. Planned reductions in crew sizes on Navy platforms currently under development, for example, mean that future watchstanders and decision makers will oversee a broader array of concurrent tasks and, in many instances, manage more demanding workflows. Updated workstations designed for this purpose, now referred to as the "Common Display System" [7] (see Figure 1), will in fact feature more visual display space—and thus will likely present more simultaneous information and/or decision tasks—than an individual can monitor at once, which in turn will call for the design of effective strategies for guiding visual attention.

Virtual auditory cueing offers a recommended and reliable technique for this purpose [8], [1], but nevertheless remains subject to a range of human factors and design principles that in many cases have yet to be systematized [9]. The ideal, in what is already an aural information environment in Navy command settings because of voice communications and other uses of sound, is to indicate where visual attention is needed, but to do so in a way that is succinct and unambiguous and avoids overloading or confusing the operator.

2. BACKGROUND

2.1. Test bed

The dual task employed in the 2002 study [1] and in the work reported here (as well as in earlier studies [10], [11]), combines a challenging, continuous tracking activity with a series of rule-

Figure 1: Three-screen console configuration of the Common Display System, the new information workstation being acquired for the U.S. Navy's modernization program and next-generation surface ships.

based decision events. The first of these—the "tracking task"—is performed with a joystick and is presented to participants as their primary activity. The latter, which involves a procession of blips on a simulated radar screen—and is thus referred to in this paper as the "radar task," and also as the "secondary task"—requires participants to evaluate each item's behavior as it moves down the screen and to record their decisions on a numeric keypad after the blips change color. Task scenarios involve three types of blips that are numbered from 1 to 6 and move down the display in respectively distinct fashions that are easy to visually assess as hostile or neutral according to a predefined set of rules. Decision entries require participants to make two key presses, the first indicating their assessment and the second designating the assessed blip by its onscreen number.

Both tasks are visually demanding, and when they are placed on opposing right and left screens, in a manner that corresponds to the "outer" two of the three screens in the Common Display System design, a notable amount of mental and physical effort is needed to prosecute them at the same time [10], [1]. Much of this effort is a result of the distance between the outer screens as well as having to compensate for the loss of peripheral visual access to the opposing task that occurs when one looks to the left or right. In principle, however, a meaningful degree of these performance demands can be reduced by simply notifying the operator when the secondary task requires a response. In the previous dual-task studies cited above, 3D auditory cues have been shown to be a robust and effective technique for bringing this type of information to the user's attention.

2.2. Auditory cueing

The radar task in these earlier studies was augmented with a set of three easily differentiated sounds—one for each type of blip—that signaled the onset of color changes and, thus, when blip decision responses were required. To give the auditory cues a deictic (or indexical) component [5], spatial information that could be intuitively indexed to the visual display was dynamically added, and the sounds were rendered binaurally in stereo headphones with a non-individualized head-related transfer function (HRTF). In the earliest auditory study with the dual task, each sound was spatially correlated with the visual location of its corresponding blip [11]. A simpler scheme indicating only the location of the radar task itself—rather than the location of each blip—was subsequently found to be equally effective in the dual-task study carried out in 2002 [1]. It is important to note, however, that the auditory cues in both of these spatialization schemes were not perceptually co-located with the visual display information they were designed to index.

2.3. Performance questions from the 2002 experiment

The 2002 dual-task experiment was conceived in part as an initial study of the notion of a "mixed-use" virtual auditory display— specifically, an auditory display in which information designated for more than one activity is sounded. A two factor design was employed that crossed two levels of auditory cueing for the tracking task with three levels of auditory cueing for the radar task. The first level in both factors was silence (no-sound), and the remaining levels were, respectively, an auditory cue for poor tracking performance, spatially indexed to the location of the tracking task, and the two spatialization schemes for the radar task that were described in the previous section (the blip location scheme and the simpler task location scheme), both employing the same set of three sounds. This resulted in a performance baseline, three manipulations with auditory cues for only one of the tasks, and two manipulations in which auditory cues were used for both tasks.

Three measures of performance were collected: tracking error and radar task response times (both recorded by the dual-task software), and counts of the number of attention shifts participants made during each exercise, which the experimenter recorded manually on a hand-held computer. With only minor qualifications, a similar pattern of significance emerged for each measure, which supported an encouraging overall result. Specifically, while almost no performance improvement was associated with the auditory cue for the tracking task (suggesting it may have been poorly conceived), the addition of this separately designated alert did not meaningfully impact the significant performance improvements that resulted across the board when auditory cueing was used for the radar task. Put another way, both of the virtual auditory cueing schemes for the radar task had significantly positive impacts on overall dual-task performance and, perhaps more importantly, these improvements persisted in the mixed-use manipulations.

The study, however, also left a number of underlying performance questions unanswered, which became readily apparent when an effort to model and explain the pattern of results with a cognitive architecture was undertaken [12]. Cognitive architectures are essentially theoretical computational frameworks for building explanatory models of human performance based on the constraints of human perceptual, cognitive, and motor processing. Several such architectures exist, but the EPIC architecture ("Executive Process-Interactive Control") was chosen for this endeavor in part because its framework for auditory processing is somewhat more complete than that of other architectures [2].

Cognitive modeling typically begins with a performance study, reduces the performance requirements to a theoretical sequence of goal-directed actions, and then evaluates the resulting model in terms of its correspondence with the observed data. Models of complex activities are often forced to make a number of conjectural assumptions due to gaps in underlying knowledge, and this proved to be the case in modeling the 2002 study. Key unknowns faced in the modeling work include:

- *The basis for switching attention between tasks in the absence of perceptual cues*. Specific questions include how decisions to switch attention are made and how time on task (dwell time) is allocated such that patterns in the hand-collected attention shift and (inferred) task dwell time data can be explained.
- *The radar screen inspection strategy*. The radar screen varies from being empty to showing several blips at once that may or may not have changed color (note that blips that have changed color require a response). How many and which blips are assessed before returning to the tracking task? Are blips assessed before they change color?
- *The blip assessment process*. Questions include how the relevant visual information needed to assess an individual blip is gathered, how long this takes, and whether this is done with a single "look," over multiple looks, or both.

- *Performance associated with auditory cueing*. Do auditory cues prompt immediate switches of attention or is some latency involved? Does the correspondence between aural identity and blip type speed the blip assessment process?

Carefully reasoned answers for these (and other) questions were explored and settled on, but it was also recognized that additional empirical measurements were needed. Accordingly, plans were made for a new dual-task study and the scope of the modeling effort was narrowed to providing an account of performance differences between the baseline (no-sound) condition and the manipulation involving only the spatially simpler of the two auditory cueing schemes for the radar task.

The resulting comparative cognitive model of dual-task performance in these two conditions incorporated a mix of parametric and theoretically plausible solutions, which in some cases (though not others) amounted to predictions that could be empirically tested. Switches of attention to the radar task were deemed to be prompted by knowledge of its status, characterized as the number of blips present, which, in turn, dictates time spent on the tracking task. Strategies for inspecting the radar screen and assessing blips were taken to be both subject to numerous individual differences and too opaque to characterize without eye-tracking studies. (A single-screen variant of the dual-task has subsequently been used with an eye tracker to examine these two issues [13].) As a consequence, solutions for these aspects of performance were parsimoniously modeled in algorithmic terms, and parameterized to balance the demands of both tasks; additionally, it was conjectured that, when possible, blips are assessed before they change color (that is, before a response is required). Finally, related empirical work at that time [14] suggested that responses to auditory cues entail a latency period of approximately 850 ms, and it was conjectured that auditory identity did not measurably facilitate blip assessments.

3. OBJECTIVES OF THE CURRENT STUDY

The study reported below was developed to gather new dual-task performance measures and test several of the model-based predictions outlined above, and also to investigate additional design issues that are thought to be relevant to the successful implementation of mixed-use auditory displays in future Navy decision environments. In particular, the utility of auditory cueing in such settings will largely depend on the ability of deictic sounds to reliably facilitate the performance of concurrent information tasks when these sounds are used in conjunction with the virtual presentation of multiple channels of voice communications (see [15]). This context for auditory design ultimately involves balancing aural attention at more than one level: balancing competing auditory functions that are intended for operators individually, such as auditory cues and voice communications, and balancing competition between the individual's auditory display and sounds in the public setting, such as face-to-face conversation among team members, intercoms, shipboard alerts, ambient noise, and so on.

To minimize the potential for confusion among auditory sources and their informational meanings, it can be argued that virtual auditory displays in this context should be simple (i.e., no more elaborate than necessary) and should function as a fixed aural overlay on the individual operator's visual environment. Simulating the manner in which sounds in the real world are ordinarily perceived as co-located with their apparent sources, regardless of the orientation of the listener's head, is the function of an augmented auditory reality (AAR) display. An important virtue of this type of rendering is that "attaching" or "fixing" a sound to or at a meaningful visual location effectively makes any deictic function the sound is intended to have unambiguous because no perceptual mapping is involved—the sound appears to arise and persist for its duration at the place the listener is intended to look.

Using AAR to simplify auditory deixis in this way is consistent with the broader contention made in the previous paragraph, that auditory displays should be, in principle, no more elaborate than the performance context of any corresponding task calls for. Sounds can be designed to support a multiplicity of information functions—deixis, onset, identity, and disposition, to name a few—but it may well be the case that operators only make use of the information functions present in a particular instance of sound that are the most effective for the purpose at hand. If so, they can be said to adhere to a principle of "least aural effort," implying that any additional task-related auditory information that is superfluous or more readily acted upon from another cognitive or perceptual source will be ignored, if possible. A corollary to this conjecture is that excessive elaboration may be counterproductive.

An immediate test of this notion of least aural effort in the present dual task is the question of whether the correspondence between aural identity and blip type appreciably facilitates the blip assessment process. Another test is whether the unambiguous deixis AAR provides is measurably better than the spatially relative deixis that is provided by a non-augmented (auditory) reality (NAAR) style of virtual auditory display. Positive differences, if seen in both tests in the same context, could be taken as evidence in support of this proposal, as could a lack of differences, if there is evidence that other, more readily exploited, task information is also available.

Consequently, the new experiment was designed in part to be a replication of the two the manipulations from the 2002 study that were modeled in EPIC—the baseline condition and the condition in which only spatially simplified auditory cues for radar task were used—and in part to investigate the two comparative design questions posed above—the use of an AAR vs. an NAAR display and the relative importance of auditory identity information—in preparation for follow-on studies with a new test bed that will explore other issues for mixed-used displays such as overlapping use of listening space and temporal competition.

The conduct of the experiment also presented a related opportunity to measure the total time required for radar blip assessments, which was not adequately known at the start of the modeling work and had to be partially inferred [12]. The time course of this process in the EPIC model assumes that blips are acquired by the eyes, assessed in some way, and then responded to. Since the time required for this sequence of actions can be measured directly with an appropriate variation of the radar task that displays blips one at a time, scenarios with and without auditory cues were developed and added to the study.

Finally, state information that bears on a number of the performance questions that were confronted in the modeling work was captured in the study. Among the issues this data will eventually help to empirically evaluate are the radar screen inspection strategy, blip assessments, and the relationship between time given to the current task and the status of the unattended task.

4. METHOD AND APPARATUS

4.1. Setup

During the period in which the baseline condition and the simpler radar cueing manipulation from the 2002 experiment were being modeled, the dual-task software was revised to run natively under the current Macintosh operating system. The software was then further modified to communicate with a new virtual audio server and to record state information that can be used to reconstruct scenarios in future analyses of performance data. A separate software package, run under the Windows operating system, was developed to present the auditory cues and utilize an inertial head tracker. As before, the audio component of the study was rendered binaurally in stereo headphones with the same non-individualized HRTF employed in the earlier study. Two flat panel monitors facing the operator on opposite sides, respectively, at 45° angles, were used to display the visual components of the experiment. The radar task was shown on the left, and blip decision responses were entered on a numeric keypad positioned below the monitor. The tracking task, which shows a rapidly moving aircraft silhouette as seen from behind, was presented on the right, and participants controlled the movement of its circular cursor with a Hall effect joystick.

4.2. Recording Switches of Attention

The critical augmentation in the setup for the new study was the addition of a head tracking system, which is necessary for implementing an AAR display but also allows head orientation data to be logged automatically, in contrast to the manual technique that was used before to track shifts of attention between the two tasks. The hand-held computer used for this purpose in the previous study enabled the experimenter's observations to be time stamped, and this, in addition to providing a both a record of attentional transitions and a measure of task switching effort, allowed cumulative distributions of time-on-task between attention shifts (dwell times) in each condition to be developed for the modeling work (allowing for experimenter errors and a one sec. resolution for manual input).

The right-skewed patterns exhibited in these distributions for both tasks yielded a number of important explanatory insights and were among the key criteria the modeling work aspired to account for. For example, differences between the baseline distribution of tracking task dwell times and the corresponding distribution in the (modeled) auditory display condition revealed that most of the significantly greater number of attention switches participants made to the radar task in the absence of auditory cues were associated with very short episodes of tracking. Since all attention to the radar task in the baseline condition was unprompted, the dominance of short tracking dwells indicates that participants were forced to look at the secondary task early and often to maintain sufficient awareness of its status. As noted above, the model's account for this data predicts that short periods of attention to the tracking task correspond to phases in which relatively high numbers of blips are present on the radar screen. In contrast, the smaller numbers of short tracking task dwells in the sound condition demonstrates that the correspondence of auditory cues with blip color changes affords longer periods of attention to the tracking task by reducing the need to see when blip responses are required.

Gathering attention shift data in the new study by automated means is not expected to refute the insights gained from the previous study's manually collected data on the basis of greater accuracy, but, instead, is expected to provide the means for evaluating the analysis of this earlier data, realized as modeling predictions, and, somewhat less importantly, to provide more objective counts of attention switches and a better temporal resolution of dwell times.

4.3. Experimental Design

Twenty NRL staff members volunteered to participate in the experiment. Of these, two individuals had to be dropped due to anomalous attention switching performance, resulting in a group of 6 women and 12 men, ranging in age from 19 to 49 with a mean of 30. Over the course of two days, participants trained to perform the two tasks separately and together, were familiarized with the sounds used in the study, and then carried out the main experiment, which was composed of four dual-task exercises under different treatments in a single-factor, repeated measures design. Treatments were given to participants in counter-balanced order, and independently of this, each exercise was successively scripted by a different radar task scenario involving 65 blip decision events. After completing the main experiment, participants were given two further exercises involving only an altered version of the radar task. A summary of all of the exercises participants were assigned is given in Table 1.

The four treatments in the main experiment entailed a baseline exercise with no sound, designated as NS below, and three manipulations, respectively designated as NAAR3, AAR3, and AAR1, in which the radar task was augmented by progressively different virtual auditory cueing designs. The first of these, used in NAAR3, was an auditory display with three easily differentiated auditory cues (one for each type of radar blip) that were localized in the same manner as the simpler of the two spatialization schemes used in the 2002 study. As its designation implies, this display was an NAAR listening space, meaning that the correspondence between the radar task and the virtual source of the auditory cues—nominally located forward and 45° to the left in the listener's auditory field—was relative to the direction the listener was facing. NAAR3 replicated the aurally-cued manipulation that was modeled in EPIC. The next auditory manipulation, AAR3, was like NAAR3 in all respects except that it used an AAR listening space. Thus, in this second auditory cueing design, the virtual source of all three sounds appeared to be co-located with the radar task regardless of the orientation of the listener's head. The final manipulation, AAR1, used the study's third auditory cueing design, which, like AAR3, was also an AAR display that used a single virtual sound source co-located with the radar task. However, in this final auditory cueing design, only one sound was used instead of three, and its aural identity was different from the three sounds used in the NAAR3 and AAR3 treatments.

The auditory materials used to augment the radar task are short audio files of warning sounds that are played as sound

a) Main Experiment

Condition	Description
NS	Baseline dual task exercise with **no sound** (i.e., no auditory cueing was used)
NAAR3	Dual-task exercise with a **non-augmented auditory reality** display using **3** auditory cues to signal radar blip color changes - each blip type signaled by an *identifying* sound - one virtual source for all three sounds - correspondence of radar task to perceived location of sounds is relative to orientation of listener's head
AAR3	Dual-task exercise with an **augmented auditory reality** display using **3** auditory cues to signal radar blip color changes - each blip type signaled by an *identifying* sound - one virtual source for all three sounds - radar task and perceived location of sounds are co-located
AAR1	Dual-task exercise with an **augmented auditory reality** display using **1** auditory cue to signal blip color changes - all three blip types signaled by the *same* sound - one virtual source for all three sounds - radar task and perceived location of sounds are co-located

b) Blip-Assessment-Time Study

Condition	Description
BA-NS	**Blip assessment** exercise—**no sound**
BA-S	**Blip assessment** exercise with **sound**

Table 1: A summary of a) the four experimental conditions in the main experiment, showing their coded designations, and b) the two additional exercises conducted to measure blip assessment times. All exercises were assigned to participants in counter-balanced order.

loops. Loops start when each event's color assignment is made and end when decisions are entered, but are only sounded one at a time and always correspond to the oldest unacknowledged event whenever overlaps occur. The sounds used in the NAAR3 and AAR3 manipulations are a police siren, an air-raid siren, and a diesel truck horn, and the sound used in the AAR1 manipulation is a low frequency pulse alert. Unspatialized examples of each of the auditory cues are given in the audio files accompanying this paper, which are listed below (these files are also available by email from the first author as .wav or .mp3 files).

[SIREN.WAV]
[AIRRAID.WAV]
[HORN.WAV]
[PULSE.WAV]

The two radar-task-only exercises that followed the main experiment were designed to explicitly measure how much time radar blip assessments take. These exercises were conducted as an ancillary study to develop parameters for future modeling work and are analyzed here as a single factor, repeated measures study with two levels, designated as BA-NS and BA-S. In each exercise, a scripted sequence of 72 individual blips was displayed by a version of the radar task that was altered to present a black screen with a red dot corresponding to the center of the radar display before each moving blip was shown. Participants were asked to focus on the red dot and then look at the displayed blip, assess it as they would in the dual task, and enter their decision on the numeric keypad. In the exercise designated BA-NS, participants assessed blips without auditory cues; the BA-S exercise was augmented by the auditory display used in the AAR3 manipulation in the main experiment. A different script was consistently used for each exercise and the manipulations were assigned to participants in alternating order.

4.4. Data and Planned Analyses

As in the 2002 study, three primary measures of performance were collected in the main experiment: tracking error, radar task response times, and counts of the number of attention shifts participants made during each exercise. Based on the previous findings, a correlated pattern of significant differences among the treatment means for each of these measures was expected to be found. Also, because the auditory design questions the main experiment addresses are progressive, the manipulations were specifically ordered to allow planned orthogonal contrasts to be made. A significance level of .05 is used for all analyses.

Only preliminary progress has been made on the more detailed, secondary analyses that are expected to shed light on model-related questions. These results will be reported at a later date. However, implications of the present analyses for the modeling work are covered below, as well as the measures resulting from the two blip assessment exercises.

5. RESULTS

The treatment means for the primary measures in the main experiment are shown in the plots in Figures 2, 3, and 4 (error bars in all of the plots show the standard error of the mean). A consistent pattern of performance differences is present, and a one-way, repeated measures ANOVA for each measure was significant (see Table 2). As in the 2002 study, tracking error data was normalized to compensate for individual differences by subtracting each participant's mean tracking error in their final tracking-only training exercise from their mean tracking error in each manipulation and dividing these differences by the standard deviation of the tracking-only mean. Radar task response times were measured in ms from the point at which blips first change color to the point at which participants made the second of the two key presses required for decision responses (see Section 2.1). The means for these two measures in the NS and NAAR3 conditions are relatively close to the respective values in the corresponding manipulations in the 2002 study: tracking errors are slightly lower and blip response times are a little over 200 ms higher than their earlier counterparts. The mean number of attention switches in the NS and NAAR3 manipulations, though, at 299.5 and 224.4, are

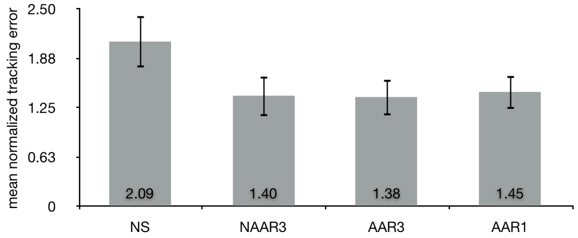

Figure 2: Mean normalized tracking error in the main experiment. The method of normalization is given in the text (Section 5).

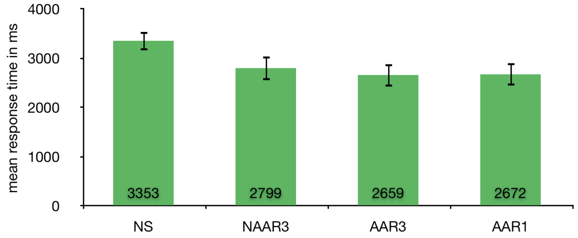

Figure 3: Mean blip response time for the radar task in the main experiment. Measures shown are for the second of the two key presses participants were required to make to record each decision (see Section 2.1).

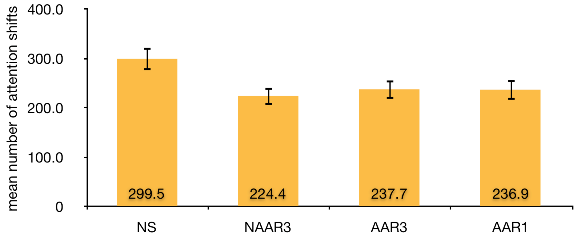

Figure 4: Mean number of attention switches between tasks in the main experiment derived from head tracking data. See the text (Section 5) for additional information about the calculation of these counts.

notably lower than the respective counts of 411.2 and 295.2 that were obtained by hand in the previous experiment, and may, in fact, underreport the number of attention switches participants actually made. The counts published here are a function of the underlying head orientation data collected in each exercise. This measure, which was logged at rate of 20 Hz, proved to be much noisier and subject to individual differences than expected. Although unambiguous shifts from right to left and back again are present in much of the data, many instances where it is unclear whether a genuine change in orientation occurred are also present. To smooth this directional jitter, lower sample rates and a series of distance thresholds were methodically explored. A sample rate of 4 Hz in combination with five thresholds ranging in even steps from 0.02 to 0.1 radians (1.15 to 5.73 degrees) resulted in a stable series of progressively decreasing counts in each of the four treatments. The numbers reported here correspond to the largest threshold and are the most conservative set of the group. However, any of the other thresholds could have been reported with no impact on the significance of the main effect for this measure. Although the empirical counts in Figure 4 potentially challenge the targets for this measure in the modeling work, the ratio of NS to NAAR3, at 1.33, (as well as this ratio for the lower thresholds described above) is quite close to the corresponding ratio of 1.39 in the earlier study.

a) Normalized Tracking Error

Comparison	Test
main effect	$F(3, 51) = 6.9, p < .001$*
NS with (NAAR3+AAR3+AAR1)/3	$F(1, 17) = 21.1, p < .001$*
NAAR3 with (AAR3+AAR1)/2	$F(1, 17) = 0.006, p > .05$
AAR3 with AAR1	$F(1, 17) = 0.43, p > .05$

b) Blip Response Time

Comparison	Test
main effect	$F(3, 51) = 10.14, p < .001$*
NS with (NAAR3+AAR3+AAR1)/3	$F(1, 17) = 17.62, p < .001$*
NAAR3 with (AAR3+AAR1)/2	$F(1, 17) = 1.07, p > .05$
AAR3 with AAR1	$F(1, 17) = 0.029, p > .05$

c) Attention Shifts

Comparison	Test
main effect	$F(3, 51) = 12.36, p < .001$*
NS with (NAAR3+AAR3+AAR1)/3	$F(1, 17) = 20.38, p < .001$*
NAAR3 with (AAR3+AAR1)/2	$F(1, 17) = 1.81, p > .05$
AAR3 with AAR1	$F(1, 17) = 0.006, p > .05$

Table 2: Summary of statistical analyses of the primary performance measures in the main experiment: a) normalized tracking error, b) blip response time, and c) number of attention shifts between tasks. Tests marked with an asterisk are significant.

Planned comparisons among the means for each of the primary measures are also shown in Table 2. These linear contrasts progressively compare a) performance in the baseline condition to the mean of the three auditory display conditions, b) performance in the NAAR design to the mean of the two AAR designs, and last, c) performance with three auditory cues to performance with just one. The first of these is significant for all three measures, and thus provides clear evidence that the auditory treatments meaningfully helped participants carry out the competing tasks at the same time. None of the contrasts comparing the three auditory display designs amongst themselves reached significance, though, and this is an important result that will be considered in greater detail below.

The two blip-assessment exercises with a modified version of the radar task that followed the main experiment yielded a substantial amount of information that will be useful for additional modeling refinements. The scripts for these exercises required participants to decide whether blips from all three of the type categories were hostile or neutral, both before and after they changed color; instances of each of the color assignments (which have not been covered in this paper) were also included, thus giving a balanced set of measures for the different configurations of visual information participants dealt with in the main study. Although comparisons of these breakdowns are not presented here, the means for both treatments, BA-NS and BA-S, are shown in Figure 5. The times shown are for the first of the two key presses participants made for each decision. This measure affords the most straightforward way to use performance constraints to infer the amount of time an operator spends in the overall assessment procedure gazing at a blip to encode its criterial information (see Section 3). Specifically, the time required to acquire each blip visually and the time required to execute the appropriate first key press can be calculated on the basis of standard results in the human performance literature. These intermediate values, which "frame" the core measure of interest, can then be deducted from the gross measure to extrapolate the time spent studying the blip.

Finally, it is interesting to note that while the small, 2.6 percent difference between these means is not significant, *F*(1, 17) = 1.09, $p > .05$, it is nevertheless in the direction that is typically seen when auditory cues accompany visual information. The difference is slightly larger in the same direction, at 3.8 percent, for the mean of the second key presses in these exercises, which are 2476 and 2382, respectively. These latter numbers are essentially measures of distraction-free responses, so it is useful to compare them with the mean blip response times shown in Figure 3 as a way of understanding the impact of the operational paradigm on decision making. In the absence of auditory cues, the presence of an additional task (i.e., tracking) and the distance between the task displays adds 877 ms (nearly a second) to decision responses. And even with auditory cueing, a difference of 277 ms with the measure in the AAR3 condition (the type of display used for the BA-S treatment and the lowest in the study) is still present.

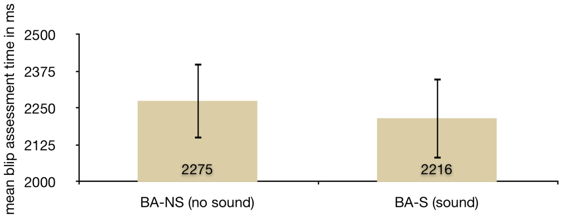

Figure 5: Mean response times from the two blip assessment exercises with a modified version of the radar task that followed the main experiment. Measures shown are for the first of the two key presses participants were required to make to record each decision (see Section 2.1 for the radar task response procedure).

6. DISCUSSION AND CONCLUSION

Two important purposes were met in the design and implementation of this study. The first was to revisit the manipulations from the 2002 dual task experiment that were modeled within the framework of the EPIC cognitive architecture, with the intent of examining specific attentional performance predictions and inferred parameters that came to light in this work. The outcome of this goal was a replication of the main finding of the earlier study, namely, that virtual auditory cueing can meaningfully improve the performance of widely separated concurrent tasks, in large part by significantly reducing the degree of attention switching (taken to be a measure of effort) that is needed to maintain adequate awareness of both tasks. The logging of state information, not gathered in the 2002 experiment, which can be used to reconstruct the status of the dual task at key points, allowing questions about the relationship between courses of action and specific situational patterns to be studied, is expected to be a useful asset for further explanatory and predictive modeling of concurrent tasks involving visual and auditory information.

The other major purpose served by the experiment was the methodical investigation of progressively different auditory display designs involving elements that are thought to be important for the composition and use of much richer auditory information displays than the relatively straightforward, single-purpose application that was evaluated here. While the lack of meaningful differences among these treatments may seem puzzling, it is nevertheless a valuable and encouraging result.

None of the prior series of dual-task studies involving auditory cues have utilized an AAR listening space. Yet it seems unlikely that any NAAR design could have reliable utility in real-world settings in which operators must regularly interact with multiple team members, turn to face large, team-oriented displays, and maintain a general awareness of a complex information environment that is likely to include public uses of sound. Disparate uses of virtual audio that had to be mapped to more than one task would potentially invite confusion unless operators were required to remain perpetually oriented toward their workstations. In principle, however, AAR would directly address this concern, particularly in the context of a mixed-use auditory display, by allowing auditory cues and other sound information to be virtually co-located with, and so inherently draw attention to, the different tasks they correspond to, regardless of where the listener might be looking.

On the basis of this reasoning, it is unlikely that the AAR3 treatment would have been in some way inferior to the NAAR3 treatment, and the fact that performance in both AAR treatments was effectively no different than in the NAAR3 manipulation can be taken as persuasive evidence that this is indeed the case. But this finding does suggest that virtual aural overlays on visual information are probably unnecessary—though certainly not disadvantageous—in relatively simple, single-use auditory display applications (e.g., the radar task in the present study), especially when the pace of the environment requires the operator to maintain a high degree of situation awareness and remain oriented toward the performance context. More to the point, it is entirely likely that adding any form spatial information to auditory cues is unnecessary in visually circumscribed, single-purpose applications because operators can readily intuit the import of the sounds.

In an indirect, but principled way, support for this last assertion is arguably provided by the contrasts between the AAR3 and AAR1 treatments, which show, for the purpose of executing the dual task, that the removal of aural blip identity information had no meaningful impact on performance, that is, one auditory cue was as good as three. This outcome implies that simpler auditory information designs for visual attention can be in some cases as good as, or even preferable to, more information-laden designs, which, in turn, may be a particularly useful finding for the design of mixed use auditory displays.

The principle of use that unifies these two outcomes is the notion of least aural effort that was proposed in Section 3, which asserts that, on the whole, listeners only make use of the information functions present in a particular instance of sound that are the most effective for the purpose at hand. (c.f. the "principle of least effort" in [16]). The evidence from the contrasts of auditory treatments in the study is that, beyond the onset function of the auditory cues, listeners were indifferent to the manipulation of two kinds of additional task-relevant information: identity and locational deixis. The most plausible explanation for this indifference is that participants were able to more efficiently gather and process these essential pieces of information for performing the radar task from other sources, one being cognition (where is the task?) and the other being the visual display (what must be decided?). This is not to say that the augmentary aural information could not have been used, only that it appears to have been superfluous in the specific context of the dual task as employed here.

With only a secondary task requiring intermittent attention and all of the criterial information for blip assessments readily available to the eyes, the dual task presents little or no opportunity for listeners to make timely use of the two categories of auditory information that were manipulated in this study. But this circumstance is unlikely to hold where mixed-uses of auditory cues are required. In Navy operations, watchstanders already attend to opposing chat and tactical situation displays and are subject to documented lapses of attention [17]. Virtual auditory cueing is being studied as a strategy for ameliorating this concern, and it is difficult to argue that performance in the absence of aural identities and deixis for these and other tasks in this type of setting will serve the operator well, precisely because these functions index a specific task among several. Additional aural elaboration, though, may be unneeded or counterproductive unless it can be exploited more readily than other sources of task-relevant information.

7. ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under work request N0001410WX20448.

8. REFERENCES

[1] D. Brock, J. A. Ballas, J. L. Stroup, and B. McClimens, "The design of mixed-use, virtual auditory displays: Recent findings with a dual-task paradigm," in *Proc. of the 10th Int. Conf. on Auditory Display (ICAD),* Sydney, Australia, July 6-9, 2004.

[2] D. Kieras and D. Meyer, "An overview of the EPIC architecture for cognition and performance with application to human-computer interaction," in *Human Computer Interaction*, 12, 391-438, 1997.

[3] H. Fouad, J. A. Ballas, and D. Brock, "An extensible toolkit for creating virtual sonic environments," in *Proc. of the 6th Int. Conf. on Auditory Display (ICAD),* Atlanta, USA, April 2-5, 2004.

[4] W. W. Gaver, "Using and creating auditory icons," in *Proc. of the 1st Int. Conf. on Auditory Display (ICAD),* Santa Fe, USA, October, 1992.

[5] J. A. Ballas, "Delivery of information through sound," in *Proc. of the 1st Int. Conf. on Auditory Display (ICAD),* Santa Fe, USA, October, 1992.

[6] N. A. Stanton and J. Edworthy, "Auditory warning affordances," in N. A. Stanton and J. Edworthy (Eds.), *Human Factors in Auditory Warnings*, Ashgate, Aldershot, UK, 1999.

[7] General Dynamics Advanced Information Systems, "Common display system," at http://www.gd-ais.com/index.cfm?acronym=cds

[8] G. A. Osga, "Human-centered shipboard systems and operations," in H. R. Booher (Ed.), *Handbook of Human Systems Integration*, Wiley, Hoboken, USA, 2003.

[9] S.C. Peres, V. Best, D. Brock, C. Frauenberger, T. Hermann, J. Neuhoff, L.V. Nickerson, B. Shinn-Cunningham, and A. Stockman, "Auditory interfaces," in P. Kortum (Ed.), *HCI Beyond the GUI.* Morgan Kaufman, San Francisco, USA, 2008.

[10] 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 *Proc. of the Human Factors and Ergonomics Soc. 46th Ann. Meeting*, Baltimore, MD, 2002.

[11] 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 *Proc. of the 4th Ann. Symp. and Exhibition on Situational Awareness in the Tactical Air Environment*., Patuxent River, MD: Warfare Center Aircraft Division, 1999.

[12] D. Brock, B. McClimens, A. Hornof, and T. Halvorson, "Cognitive models of the effect of audio cuing on attentional shifts in a complex multimodal dual-display dual-task," in *Proc. 28th Ann. Meeting of the Cognitive Science Soc*., 2006.

[13] A.J. Hornof, Y. Zhang, and T. Halverson, "Knowing where and when to look in a time-critical multimodal dual task, to appear in *ACM CHI 2010: Conference on Human Factors in Computing Systems,* New York: ACM, 2010.

[14] A.J. Hornof, T. Halverson, A. Issacson, and E. Brown, "Transforming object locations on a 2D visual display into cued locations in 3D auditory space," in *Proc. of the 52nd Ann. Meeting of the Human Factors and Ergonomics Soc.,* 2008, pp. 1170-1174.

[15] D. Brock, B. McClimens, J.G. Trafton, M. McCurry, and D. Perzanowski, "Evaluating listeners' attention to and comprehension of spatialized concurrent and serial talkers at normal and a synthetically faster rate of speech," in *Proc. of the 14th Int. Conf. on Auditory Display (ICAD)*. Paris, France, June 24-27, 2008.

[16] H.H. Clark, *Using language*, Cambridge University Press, New York, 1996.

[17] J. Cantanzaro, M. Risser, J. Gwynne and D. Manes, *Facilitating critical event detection in chat communications*, (Technical Report). San Diego, CA: Pacific Science & Engineering Group, Inc., 2006.