TO WHAT EXTENT DO LISTENERS USE AURAL INFORMATION WHEN IT IS PRESENT?

Derek Brock and Brian McClimens

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

ABSTRACT

In three of the manipulations in a 2009 dual-task performance study, virtual auditory cues were used to alert participants to the onset of three kinds of decision events in the secondary, but more demanding of the two tasks. Two aural parameters, the manner in which the cues were spatially presented and the level of task-related information they carried, were systematically altered to compare the impact of these factors on performance. In the work presented here, we focus on performance measures that can be correlated with participants' use of aurally-based task-related information in the study. Although secondary task decision response times were nominally the same in each manipulation, an analysis of head tracking data shows that, when they were cued, participants turned their attention from the primary to the secondary task significantly sooner when a single sound (always the same) was used to announce decision events. In contrast, when a different sound was used to signal each kind of decision event, participants, after being cued, spent less time (but not significantly so) examining the secondary task before entering their responses. The nature of this tradeoff and its implications for information design in auditory cueing is discussed.

1. INTRODUCTION

The use of sound to signal events is one of the most basic examples of an auditory display. This simple idea is not necessarily as simple as it sounds, though, because sound has the capacity communicate much more than just the onset of one thing or another. Naturally occurring sound is a by-product of physical interactions in the world and is usually understood by listeners to convey something about its source and its cause. Listeners come to know a great deal about the world through their aural perception of it, and this knowledge can be characterized in terms of its correspondence with a range of informational concerns. In addition to hearing when, what, and where, for example, listeners also make use of sound (or in some cases, its absence) to monitor and predict duration, rhythm, and process, discern many different kinds of state information, explore material structure, and infer a variety of consequences and relationships.

Our tool-based ability to exploit and manipulate sound, though, is what lies at the heart of our contemporary notion of auditory display. This practice allows us to take insights about how sound is naturally understood and appropriate these ideas for informational and semiotic purposes in activities involving information technology. Checked only by processing and rendering constraints, the programmed display of an aural cue can thus be made to convey, intuitively and/or by assignment, more than the fact that something has just occurred; it can also communicate what that something is, where it is located (or where the listener should look), and any number of additional informational matters that may be germane to the task at hand.

To some degree, the amount of information that can be packed into a particular sound is arbitrary and is only limited by the nature of the task the sound applies to. Much the same can be said for what kind information is included in a sound and what kind information is left out. Since what we hear is largely independent of the direction we are facing, sound is often used to draw attention to and reinforce visual information or information that is presented in some other modality. Moreover, redundant, co-occurring multimodal information, such as onset, cessation and other spatial and temporally governed factors, is known to enhance a listener's ability to perform visual search and other kinds of orienting tasks (e.g., [1]). But a given sound can also be used to relate information that may be otherwise unavailable, such as the identity, signature, or disposition of an event, and so on. Or the reverse can be true—a sound may be intentionally formatted to convey only some of the information needed to perform a task and leave the remainder to be acquired from another source or sources.

To be effective, however, the information capacities of an auditory display must be matched by a commensurate set of listener performance capacities, which entails both the ability to make any necessary aural discriminations the listening task requires and the capacity to process natural and assigned meanings. Informational differences among sounds can be encoded in many ways, but some are more difficult than others to discern. Some listeners—for instance, some musicians—have exceptional hearing skills, which can include the ability to identify harmonic patterns and to effortlessly recognize pitch relationships. For most, though, there are practical limits to what can be recognized and appreciated aurally, even with substantial training and experience. Kramer [2] refers to this sort of user limitation as the "aural equivalent of color blindness." There are also empirical bounds on listening performance that are subject to the nature and limitations of the human cognitive architecture [3]. Some of the more important constraints the architecture imposes are the nature of attention, the speed of perceptual processing, and the extent of what is often referred to as "working memory." Although a small number of auditory qualia can be perceived nearly instantaneously, many of the informational features that make one sound characteristically different from another are temporal in nature and thus require momentary if not sustained attention to be comprehended. And because aural attention, like other forms of perceptual attention, is a limited resource [4], even practiced listeners can be quickly overwhelmed by just a small number of sounds if the informational component of their task is aurally demanding as a result of its pace or complexity.

In short, sounds can readily be devised or adapted to serve any number of task-related purposes, but the utility of doing this is subject to the limits of listeners' abilities and their need to understand and act on what they hear in a given context. Furthermore, despite a great deal of interest and varied research and thinking in this area (e.g., [5], [6], [7]), it is fair to say that the implications of human audition for the design and application of task-based aural information remain under-characterized [8]. Thus, it is important to continue to explore how and under what circumstances listeners can or do make use of aural information when it is present, and whether some informational configurations are better than others, particularly when performance concerns are paramount.

2. BACKGROUND

The work presented here examines a specific and rather straightforward instance of aural information use in the context of a 2009 auditory cueing study reported in [9]. In three of the study's four major manipulations, virtual auditory cues were used to alert participants to the onset of three kinds of decision events in the secondary, but more cognitively complex of two visual tasks they were asked to carry out at the same time. The tasks—the primary and most attentionally demanding one requiring participants to continuously track an evasive target with a joystick, and the other requiring an intermittent procession of radar blips to be assessed and updated via a keyboard—were displayed on separate monitors, placed with their centers 90° apart from each other, to study attentional issues the wide visual span of the U.S. Navy's next generation of shipboard watchstations may pose for future watchstanders. An inherent consequence of using this much visual display space—a horizontal array of three high-definition monitors—is an inability to see everything when one turns to look at either the left or the right display.

In the three manipulations we are concerned with here, the use of auditory cues significantly improved participants' baseline measures of performance and demonstrably reduced the effort that was otherwise needed to pursue both tasks. This outcome replicated earlier findings with the same materials (see [10]), but more to the point, both of the informational differences in how auditory cueing was used in these manipulations, one involving location and the other identity, appeared to show that listeners made little or no use of supplementary task-related sound information when it was present—at least in this particular performance paradigm ($F(2, 34) = 0.84, p > 0.1$). To be specific, the use of head tracking and spatial audio technology in [9] made it possible to manipulate how aural location information was virtually displayed. And likewise, the use of one sound vs. three in contrasting treatments allowed the utility of assigning sound identities to decision events to be evaluated in the same manner.

Elaborating on the location manipulation first, in the conditions designated as AAR3 and AAR1 in [9], sounds were rendered as an instance of aurally augmented reality, meaning that each auditory cue seemed to come directly from the secondary task itself. In contrast, in the condition designated as NAAR3, this pairing between real and virtual locations was uncoupled. All of the auditory cues in the latter condition came, instead, from a virtual location that reflected the orientation of the listener's head in the familiar way un-instrumented stereo headphones create a personal listening space that is unaligned with, and independent of, the listener's real-world environment. As a consequence, the apparent location of each auditory cue in the NAAR3 condition had to be conceptually mapped to the secondary task display—at least, that is, if this deictic aspect of the cue was deemed by the listener to be useful—and this additional mapping step, if present, should theoretically take time. As Figure 1 shows, however (and a similar pattern emerged for the other performance measures in [9]), participants' mean secondary task response time in the NAAR3 condition was essentially the same as in the two conditions where all auditory cues appeared to come directly from the secondary task.

Analogously, in the manipulation involving the use of aural identity information, each of the three kinds of decision events in the secondary task were cued by the same sound in the AAR1 condition and by identifiably different sounds in NAAR3 and AAR3 (the same three sound-to-event pairings were used in both of these conditions). Giving each kind of decision event its own aural identity conceivably could prime the operator's selection process, both for the kind of decision event (radar blip) to look for and for which assessment rules to apply and, therefore, could speed these aspects of the task. But listening for and registering this information should also take time if it were something the operator decided to do before carrying out the rest of the response procedure. Yet, as Figure 1 shows, again, there were no meaningful performance differences across all three audio conditions.

Neither of these outcomes were fully expected, but as we have hinted above, the explanation for this may have more to do with the experimental paradigm than with the possibility that listeners only require aural onset information in multitask settings. What was evaluated in [9] was performance with a single-purpose auditory display in the context of a demanding dual task; only one of the concurrent activities was aurally cued and the pace of both tasks required participants to maintain a high degree of situation awareness and remain physically and visually attentive to both of the task displays. With no advantage to be gained in turning away from the combined activity, participants' physical knowledge of the respective locations of each display would be essentially automatic after a minimal amount of experience (much as, for instance on a different scale, one's fingers can learn to physically "know" the pattern of a qwerty-style keyboard). Consequently, with only one task being cued, little or no deliberative thought would be needed

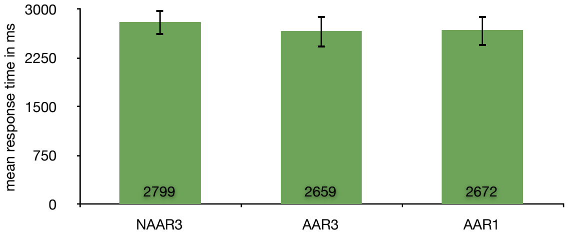

Figure 1: Mean response times (ms) in the secondary task in the three aurally cued manipulations reported in [9]. Error bars show the standard error of the mean (s.e.m.). The task required participants to decide whether instances of three kinds of radar blips were hostile or neutral based on different rules associated with each type. Measures shown are for the second of two key strokes participants were required to make to record each decision. Condition naming key: N=non, AAR=augmented auditory reality, 3=three auditory cues used, 1=one auditory cue used. For reference, the mean and s.e.m. for this measure in the baseline (NS=no-sound) condition were 3353 and 193 ms.

to know precisely where each cue was indicating to look, which, in turn, would obviate most if not all of any additional benefit aurally encoded location information could provide. Assuming this explanation is right, monaural or diotic cueing arguably would have been as effective. Similarly, because it was to the participants' advantage to maintain a high degree of secondary task awareness, any informative benefit and/or cognitive priming the aural event identity information in the NAAR3 and AAR3 conditions could provide may have already been effectively achieved by other, presumably visual, means. In effect, it could be the case that all that virtual auditory cueing really brings to the table in this particular multitask paradigm is event onset information. In the material presented below, however, we take a closer look at the head tracking data that was collected in [9], and this more detailed analysis appears to tell a somewhat different story about participants' use of aural information.

Both of the performance findings the previous paragraph's after-the-fact account attempts to explain were construed in [9] as examples of a more general conjecture that was characterized there as a "principle of least aural effort." The claim is that listeners tend to use only the information in a particular instance of sound that is the most effective for their present purposes and will essentially disregard information that is superfluous or that can be more readily acquired and acted upon from other cognitive or perceptual resources. This is not to assert that listeners always can or do fully tune out unneeded aural information—sound *per se* can be hard to ignore—nor that the division between sources of task information available to performers is an all-or-nothing proposition, but only that listeners simply tend to be cognitively and perceptually efficient. If this premise is correct, it should be possible to see how listeners compensate when equivalent task related information can also (or must) be acquired non-aurally; it also suggests that an auditory display design need be no more elaborate than its performance context requires.

3. ANALYSIS OF HEAD TRACKING DATA

The head tracking data generated in the 2009 study reported in [9] was collected, in part, so that certain aspects of attention switching in the dual task could be examined more extensively than previous experiments have allowed (e.g., [10]). Because of our present focus on the degree of participants' use of aurally encoded information, and because the principle of least aural effort argues that listeners will tend to do what is easiest and most effective for them, our objective in the analyses presented in this section—which assess the 2009 head tracking data in greater detail than in [9]—is to evaluate patterns of performance that are specifically associated with aurally cued shifts of attention.

Empirically, competent performance of the dual task on opposing displays entails a large number of looks back and forth between the two tasks, and this turns out to be the case even when auditory cues are employed, in spite of the meaningful performance improvements that attend with their use (see [9] for the mean number of attention shifts in each condition). So even though all of the secondary task decision events in the three auditory manipulations in [9] were aurally cued, only a subset of the events—albeit more than half—happen to have occurred when participants were looking at the primary task. Thus, while the nominal purpose of auditory cueing in the study was to convey relevant information about activity on the secondary task display requiring the operator's immediate attention and a response, this purpose was only really served—in the sense that it led to an observable shift of attention—by the cues in this subset.

3.1. Criteria

Each of the aurally prompted shifts of attention in [9] can be thought of as entailing two stages: *Listening*: the time from the cue's onset until the listener leaves the primary task, and *Responding*: the time it takes the listener to find and respond to the sounding event after turning to the secondary task. The pattern of this sequence of actions is shown schematically in Figure 2.

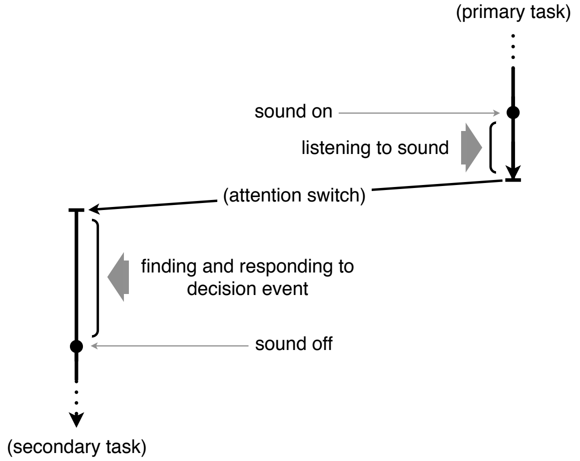

Figure 2: Schematic illustration of the pattern of actions exhibited by participants in the auditory manipulations in [9] when an auditory cue is sounded while they are looking at the primary task. (Time moves downward in the figure.) A short period of time is required to encode the auditory cue after its onset. Visual attention is then switched from the primary task on the right to the secondary task on the left, after which, the sounding decision event must be located and the appropriate response entered.

These corresponding *Listening* and *Responding* stages were identified in the head tracking data in the following way. First, all instances in which participants were attending to the primary task when the onset of an auditory cue occurred were identified. Next, the duration of the corresponding *Listening* stage was determined using the midpoint between the two task displays as the determining point for a switch of attention from the primary to the secondary task. The same midpoint was then used as the starting point for the subsequent *Responding* stage, whose duration was determined by the second of the two key presses participants were required to make to correctly assess each event.

Additionally, the mean number of aurally cued switches of attention that occurred in each of the corresponding conditions was compared, and as Figure 3 shows, there are no significant differences among the three groups $(F(2, 34) = 1.08, p > 0.1)$.

3.2. Results

As an aside before turning to the outcome of the duration data for the *Listening* and *Responding* stages, it should be noted that in carrying out the dual task, participants in [9] did not always turn to the secondary task to make their decision responses. Instead, in

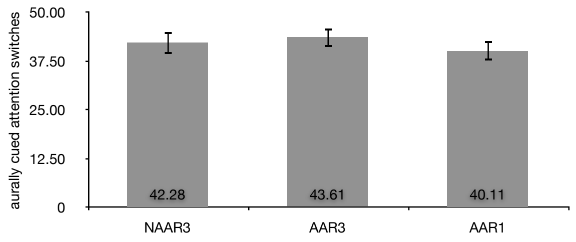

Figure 3: Mean counts of aurally cued switches of attention from the primary task to the secondary task in the three aurally cued conditions in [9]. Error bars show the s.e.m.

roughly 10% of the overall number of decisions, participants apparently were able to plan their answer before the decision event's onset, return to the primary task, and, without looking back, enter their response after the onset occurred, thus markedly improving their response time for that particular event. As a consequence, the mean counts shown in Figure 3 do not reflect all of the auditory cues that occurred while participants were looking at the primary task, but only those that prompted participants to turn to the secondary task display to make their response.

Although a few such "eyes-free" responses were anticipated in the aurally cued manipulations, somewhat surprisingly, they also occurred (albeit, a bit less frequently) in the sound-free, baseline condition, which means that in a relatively small number of cases, participants were able to visually anticipate when a decision response was likely to occur, return from the secondary task, and execute their response as they resumed the tracking activity! Overall, the mean number of eyes-free responses ranged from a low of 4.3 in the NS condition to a high of 8.2 in NAAR3; the corresponding means in the AAR3 and AAR1 conditions were 6.4. and 6.6. The differences between these counts among the four conditions was not significant $(F(3, 51) = 1.21, p > 0.1)$, but, as can be seen in Figure 4, there was a conspicuous difference between mean response time in the NS manipulation and the corresponding means in the three audio manipulations $(F(3, 51) = 3.27, p < 0.05)$; the differences between the values in the latter conditions are not significant $(F(2, 34) = 0.84, p > 0.1)$.

The pattern of eyes-free response times across the four conditions shown in Figure 4 closely mirrors the pattern that was observed for each of the major performance measures in [9], namely, overall secondary task response time, tracking error, and the number of attention switches needed to carry out the dual task. That is, all of the manipulations involving sound in the study resulted in a uniform improvement over performance in the baseline condition. A comparison of the measures in Figure 4 with those in Figure 1, shows that when listeners were able to execute this eyes-free strategy, they were able to exploit the signal meaning of the sounds to better their response times by well over a second. A benefit for making eyes-free responses is also present in the NS condition (note, the mean secondary task response time in the NS condition is given in Figure 1's caption), but is not as pronounced by half, which can be attributed to the difficulty of carrying out an anticipatory move without any perceptual confirmation.

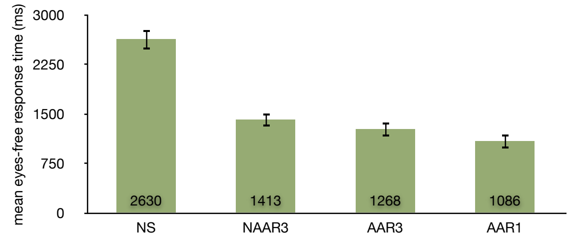

Figure 4: Mean "eyes-free" response times in the four conditions in [9]. These data correspond to secondary task decision events for which the onset and response occurred while participants maintained their attention on the primary task. Error bars show the s.e.m.

Returning to the analysis of participants' aurally cued switches of attention in [9], the mean durations for the *Listening* and *Responding* stages of this class of decision responses, which are combined in Figure 5, show that participants in the NAAR3 and AAR3 conditions, where each of the three kinds of secondary task decision events were cued by a correspondingly different sound, were slightly more than a sixth of a second slower to leave the primary task than they were in the AAR1 condition, where the same sound was used to cue all of its decision events. The pattern of differences in the *Listening* stage across the three manipulations turns out to be significant $(F(2, 34) = 3.76, p < 0.05)$, but the corresponding pattern for the *Responding* stage is not, even though the mean duration for this stage in the AAR1 condition does trend a bit higher than the corresponding data in the NAAR3 and AAR3 conditions by about a seventh of a second $(F(2, 34) = 0.73, p > 0.1)$.

This outcome, and the fact that none of the differences in the overall response times shown in Figure 1 are significant, indicates that participants were able to make an effective tradeoff between using their ears and using their eyes to determine which of the three kinds of decision events required a response at a given point in the dual task. In other words, listening for event identity information, when it was available both aurally and visually, was as efficient as having to look for it when it was not present in the auditory cues.

It should be emphasized that determining which of the three kinds of decision events is being cued—or, alternatively, confirming that the sounding event is the one the operator expects it to be—is essential to executing the secondary task correctly, and doing this inherently takes time, whether it is done by sound or by sight. Thus, even though the summary measures in Figure 1 are effectively the same in each of the aural conditions, the differences between the respective component measures in Figure 5 reflect real impacts the auditory manipulations imposed on how listeners had to acquire secondary task event identity information. Since, by design, the auditory cues in the AAR1 condition did not convey this

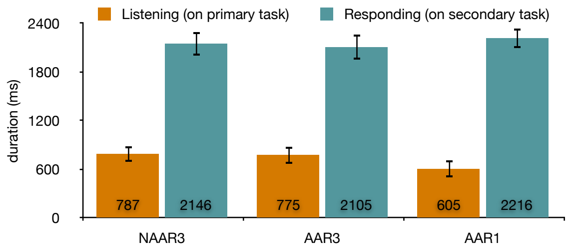

Figure 5: Mean durations of the *Listening* and *Responding* stages of aurally cued switches of attention between the primary and secondary tasks in the three auditory manipulations in [9]. Error bars show the s.e.m.

information aurally, it could only acquired by taking time to visually examine the secondary task display. Consequently, the slightly longer duration for the *Responding* stage in the AAR1 condition in Figure 5 can be taken at face value despite its lack of significance.

Two additional points about this data are worth noting. The first concerns the manipulation of the number of sounds participants heard: only one sound was utilized in the AAR1 condition, but listeners always heard one of three (with equal probability) in the NAAR3 and AAR3 conditions, so the response task in the aural manipulations can therefore be likened to one in which the number of alternative stimuli is varied across conditions. A standing result in the human performance literature posits that reaction time in tasks of this nature is a logarithmic function of the number of choices that are presented,

$$RT = a + b \log_2 N, \tag{1}$$

where a is the base reaction time for a single stimuli, b is a constant, and N is the number of alternatives. Unlike the single-task paradigm this generalization is derived from, though, participants in the dual task employed in [9] were pursuing an entirely different activity—the primary task—when the majority of the stimuli they were expected to react to were sounded; so the value that would correspond to a in Equation 1, namely, the mean duration of the *Listening* stage in the AAR1 condition, includes not just the listener's reaction time, but also the time it takes the listener to recognize the interruption and disengage from the primary task. Presumably, this additional time would also be included in the corresponding measures in the other auditory manipulations. Using 605 ms for a, then, and the average of the *Listening* stage durations in the NARR3 and AAR3 conditions (781 ms) as the value of RT corresponding to $\log_2(N = 3)$ results in an estimate of 111 ms for b, which turns out to be consistent with the range of empirical values that have been reported for this constant [11].

The remaining point to consider about the data shown in Figure 5 concerns the mean duration for the *Responding* stage in the AAR1 manipulation. Since participants were unable to infer directly from the auditory cues in this condition which of the three kinds of decision events required their attention, it was generally necessary for them to turn away from the primary task and acquire this information visually on the secondary task display before making their response—unless they were prepared (or very nearly prepared) to respond in the eyes-free manner discussed earlier. Consequently, if the *Listening* stage of participants' AAR1 responses is simply regarded as the preliminary time it took to move to the secondary task, the duration of the *Responding* stage can be interpreted as a measure of decision event response effort (i.e., finding the active event, categorizing it, assessing it—or, if the listener has already done this earlier, recalling his or her decision—and entering the response) that began when the listener arrived at the secondary task display.

Given this characterization of the AAR1 *Responding* stage, it is worthwhile to compare its mean duration with another class of secondary task response times in [9] that are functionally similar, specifically, responses that were prompted by the onset of a decision event while the participant was already attending to the secondary task. The mean response times for this class of decisions and the corresponding mean number of these responses in each of the four manipulations are shown in Figure 6. The extent of the error bars shown in the figure and the relatively narrow spread among the means (approximately 250 ms) suggest that participants' ability to do the secondary task while they were attending to it was relatively constant in all four conditions, despite what may be a small benefit from the added value of auditory cues in the aural manipulations. Indeed, there is no main effect $(F(3, 51) = 1.40, p > 0.1)$ if the counts of these data are presumed to be equally distributed; in fact, however, they are not $(F(3, 51) = 7.82, p < 0.001)$ which can be inferred from the pattern of mean counts in Figure 6.

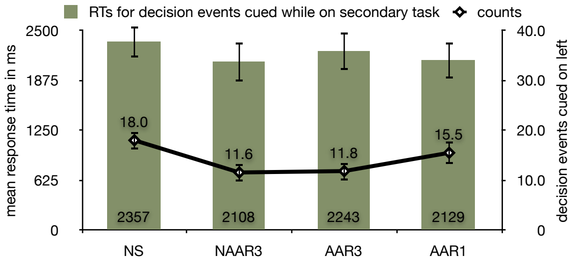

Figure 6: Column series/axis on left: mean response times (RTs) for decision events in [9] whose onset was cued while participants were already attending to the secondary task display (on their left). Point-and-line series/axis on right: corresponding mean number (counts) of these events in each manipulation. (Secondary task decision events were visually cued in all four of the manipulations and were also aurally cued in the auditory manipulations.) Error bars show the s.e.m.

More to the point, though, is the fact that the mean duration of the AAR1 *Responding* stage in Figure 5 falls comfortably within the range of mean decision event response times participants turned in when they were fully focused on the secondary task. This is the case even if the measure for the NS condition is removed from this group, but it is included in Figure 6 because of its superficial similarity to the characterization of the AAR1 *Responding* stage, made just above, as a predominantly visual task. That is, when decision events occurred in the NS condition while participants were looking at the secondary task, the only cue they received was a visual change in the radar blip that required a response. Somewhat similarly, in the AAR1 *Responding* stage, although participants had been aurally prompted to turn from the primary task, when they arrived on the secondary task display, their only recourse was to look for the same kind of visual status information (a changed radar blip) that would indicate a decision event that was already in progress. Participants also had to do much the same in response to the events in the AAR1 condition in Figure 6, too, because the corresponding aural cues only signaled the onset of decision events, but not what kind of events they were. Yet in all three of these situations, participants were able to use their eyes to do what was needed to carry out the response task in approximately the same amount of time as when a critical part of the necessary information was also available aurally in the NAAR3 and AAR3 conditions.

4. DISCUSSION

The motivation for incorporating and manipulating different types and degrees of supplementary, task-related aural information in the study reported in [9] was to investigate the performance impact (good or bad) of this additional information on participants' use of auditory cues as an aid for managing their attention to a demanding pair of concurrent tasks presented on opposing displays.

Although auditory cueing, *per se*, has been clearly shown to significantly improve critical aspects of operator performance in settings that entail multitask display configurations of this sort (e.g., [10]), little additional research has systematically investigated the kinds of informational questions that were targeted in [9], which are thought to be important for the design of mixed-use auditory displays.

How locational information is conveyed to listeners relative to the task environment, for instance, may be of consequence when auditory cues are employed for different purposes on separated displays. If, for some of the reasons offered in the second-to-last paragraph of Section 2 above, the head tracking and response time data in [9] appear to confirm that listeners effectively ignored both types of locational cueing in the study, this outcome nevertheless demonstrates that there is no performance downside for the use of AAR techniques, which are ultimately expected to benefit listeners in operational environments where attentional distractions are likely to be frequent, because AAR is able to approximate the natural manner in which most correlated aural and visual information is ordinarily coupled. It can also be construed as straightforward evidence for the principle of least aural effort.

Similarly, in the right circumstances, the use of identifiably different sounds to convey specific kinds of task-relevant information intuitively has the potential to help listeners quickly identify and/or sort through the particulars of a given activity or context, and thus facilitate timely performance. No such benefit, however, was demonstrated by the manipulations in [9]. Instead, for reasons considered more fully below, participants demonstrated what could be interpreted as a preference for aurally conveyed information, but also manifested an equally efficient ability to substitute one form of perceptual information gathering (using their eyes) for another (their ears) when there was no other choice. Thus, as a potential design principle, it is unclear to what extent the use of different sounds to characterize the pivotal or diverse elements of an intermittent task is necessary if this information can be easily made known in another way. Moreover, if competing tasks are cued by different sets of corresponding sounds, there may be performance decrements or even listener confusions depending on the number of choices each task presents and the nature of the auditory materials, despite any discriminatory benefits aural locational cues may provide. These broad questions for mixed-use auditory displays will need to be addressed in future studies.

In any event, some additional consideration must be given to the implications of the data shown in Figure 5. If, as it appears, listeners in the AAR1 condition were able to substitute a “visual approach” for determining secondary task event-kind information for the “aural approach” they ostensibly adopted and preferred in the two conditions that associated an aural identity with each kind of decision event, what does this say about the principle of least aural effort? In other words, if the visual approach is equally as effective as the aural approach, what is the motivation for bothering to listen to which kind of secondary task decision event was being sounded, especially if it would be more efficient, response-wise, to ignore this supplementary aural information and go directly to the secondary task display?

The most plausible answer is that listeners’ use of the aural identity information in the NAAR3 and AAR3 conditions allowed them to spend a small amount of additional time on the primary task and, so, to try do a bit better at what was a decidedly demanding activity. The evidence for this claim is not borne out by statistically meaningful results, but only by a small, relatively consistent pattern of trends in the relevant measures. Specifically, participants spent a slightly smaller percentage of their time on the secondary task ($F(2,34) = 2.97, p > 0.05$) and turned in very slightly better tracking scores ($F(2,34) = 0.07, p > 0.1$) in the NAAR3 and AAR3 conditions, as can be seen in the combined data shown in Figure 7. Given such thin evidence of any motivating benefit, even if this explanation correct, it will have to be revisited. There are other possible explanations, too. It may be that the principle of least aural effort proposed in [9] is wrong to the extent that operators prefer to utilize more than one perceptual resource when they can or are inherently slower to respond in any context in which more than one informational sound occurs, regardless of utility. An unlikely, but conceivable auditory cueing design that employs more than one sound for the same purpose, as well as for no purpose, in the context of a larger information task, within a new multitask environment that the authors are developing to study mixed-use auditory displays, will be used to address these questions.

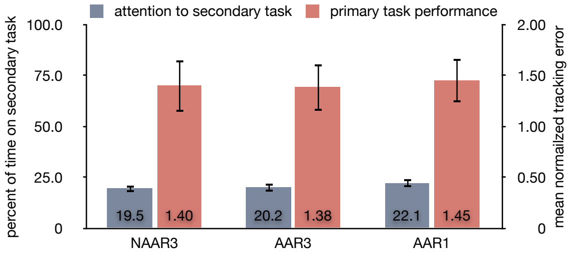

Figure 7: Side by side comparisons of (left axis) mean time spent on the secondary task, expressed as a percentage of total time on the dual task and (right axis) mean performance on the primary task, expressed as normalized tracking error (lower is better) in [9]. Error bars show the s.e.m.

5. ACKNOWLEDGMENT

The authors would like to thank Christina Wasylyshyn and Dennis Perzanowski for comments on drafts of this paper. This work was supported by the Office of Naval Research under work request N0001410WX20448.

6. REFERENCES

[1] H. Colonius and P. Arndt, “A two-stage model for visual-auditory interaction in saccadic latencies,” *Percept Psychophys*, vol. 63, no. 1, pp. 126 147, 2001.

[2] G. Kramer, “An introduction to auditory display, in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed., Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, MA, pp. 1-77, 1994.

[3] A. Newell, *Unified Theories of Cognition*, Harvard University Press, 1994.

[4] B. Shinn-Cunningham, ”Object-based auditory and visual attention,” *Trends in Cognitive Sciences*, vol. 12, no. 5, pp. 182-186, 2008.

[5] P. Ulfvengren, “Design of Natural Warning Sounds in Human-Machine Systems,” doctoral thesis, KTH, Royal Institute of Technology, Stockholm, Sweden, 2003.

[6] M-S. Mustonen, “A review-based conceptual analysis of auditory signs and their design,” in *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, June 24-27, 2008.

[7] S. Garzonis, S. Jones, T. Jay, and E. ONeill, “Auditory icon and earcon mobile service notifications: Intuitiveness, learnability, memorability and preference,” in *Proceedings of CHI 2009 Conference on Human Factors in Computing Systems*, Boston, USA, pp. 1513-1522, April 4-9 2009.

[8] B. N. Walker and G. Kramer, “Mappings and metaphors in auditory displays: An experimental assessment,” *ACM Transactions on Applied Perception*, vol. 2, no. 4, pp. 407-412, 2005.

[9] D. Brock, B. McClimens, and M. McCurry, “Virtual auditory cueing revisited,” in *Proceedings of the 16th International Conference on Auditory Display*, Washington, DC, June 9-15, 2010.

[10] 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.

[11] S. W. Keele, “Motor control.” in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kauffman, and J. P. Thomas, Eds., John Wiley and Sons, New York, 1986.