AUDITORY DISPLAYS TO FACILITATE OBJECT TARGETING IN 3D SPACE

Keenan R. May, Briana Sobel, Jeff Wilson, and Bruce N. Walker

Georgia Institute of Technology
Atlanta, 30332
United States

{kmay,bsobel3}@gatech.edu, jeff@imtc.gatech.edu, bruce.walker@psych.gatech.edu

ABSTRACT

In both extreme and everyday situations, humans need to find nearby objects that cannot be located visually. In such situations, auditory display technology could be used to display information supporting object targeting. Unfortunately, spatial audio inadequately conveys sound source elevation, which is crucial for locating objects in 3D space. To address this, three auditory display concepts were developed and evaluated in the context of finding objects within a virtual room, in either low or no visibility conditions: (1) a one-time height-denoting "area cue," (2) ongoing "proximity feedback," or (3) both. All three led to improvements in performance and subjective workload compared to no sound. Displays (2) and (3) led to the largest improvements. This pattern was smaller, but still present, when visibility was low, compared to no visibility. These results indicate that persons who need to locate nearby objects in limited visibility conditions could benefit from the types of auditory displays considered here.

1. INTRODUCTION

There are a variety of situations in which humans need to navigate spaces with limited visual input. Auditory guidance systems, such as purpose-built navigation systems for visually impaired persons [1, 2, 3], or consumer navigation software, have tended to focus on guiding a person to locations of interest on a two-dimensional plane. However, supporting 2D navigation is only part of the solution. Many occupations and everyday tasks involve targeting nearby objects in 3D with limited visual input. For example, in first responder scenarios, personnel may need to quickly locate task-critical objects which are obscured by smoke or debris. Similarly, persons operating underwater or in other unique environments with limited visibility may need to locate tools or machinery. People with visual impairment must solve this problem to carry out everyday tasks. As visual-focused VR/MR (Virtual/Mixed Reality) systems become increasingly common and capable of operation in varied situations, research into the ability of auditory displays to assist with such tasks is needed.

Unassisted, targeting objects can be cumbersome without the use of vision. Searching a 3D space without full quality visual input can take a great deal of time, and be a frustrating experience. This type of task can be divided into two components: determining/recalling the right area to search, and targeting the object itself. Each of these task components could be supported by different types of information.

First, a person needs to know the general area within which a nearby object is likely to be located. For example, a firefighter might need to locate a control panel, and knows that these are typically mounted roughly at chest height. This component of the task can be considered a knowledge problem as much as a perceptual-motor problem. Information supporting the selection of the correct search area could be retrieved from a person's memory, or an MR system could communicate target information pulled from an object database [4] or inferred via machine vision.

After deciding on the correct search area, a person must then accurately move a hand or tool to their target. If high fidelity visual input is available, a visual search may be conducted to precisely locate the target, followed by a reaching motion that is guided by a visuo-motor feedback loop [5]. However, if sufficient visual input is not available, making precise motor movements to a specific location can be difficult, even if that location is known and serial tactile search is not required. This task component can be considered a perceptual-motor problem as well as a knowledge problem. Interventions might utilize machine vision or wearable sensors to provide precise nonvisual guidance that would assist the user in moving all the way to the target, effectively creating an 'audio-motor' feedback loop.

1.1. The Sound of Space

Sound can be used both to convey 3D location information and to guide movement. However, humans tend to be relatively poor at perceiving the elevation of sound sources. Planar localization can utilize multiple types of information derived from binaural disparities [6], but elevation perception must rely on subtle spectral information derived from the way sounds are occluded by the head, ears, and shoulders, depending their direction of origin.

For virtual sound sources rendered using spatial audio, this inherent difficulty is compounded by the fact that simulating spectral information with high fidelity is more difficult than simulating binaural disparities. Spectral changes can be synthesized using Head-Related Transfer Functions (HRTFs) [7]. HRTFs can be effective if customized to reflect the geometry of an individual's pinnae and head/shoulders, but this is rarely feasible. Generalized HRTFs, while functional, are often not effective enough for a listener to consistently resolve elevation [8]. As such, relying solely on spatial audio effects to represent the position of a target in 3D space is unlikely to be effective.

Some systems have instead utilized Text-To-Speech (TTS) to describe the location of nearby objects. Thakoor et al. [9] tested

https://doi.org/10.21785/icad2019.008

a system that provided TTS denoting the presence of objects recognized by a mobile camera in one of nine areas in front of the user (e.g. "upper right"). May et al. [3] suggested that brief TTS description of object manipulation information (e.g. "trash can, button on lid") could be appropriate in some cases. A system developed by Doush et al. [10] assisted participants with blindness in grasping a specific library book via TTS description.

However, TTS description of object position can be relatively slow and cumbersome. It is also inappropriate in the myriad of situations in which a person's auditory environment is not conducive to TTS comprehension, or, conversely, in those in which a person's capacity to comprehend incoming speech should not be disrupted by TTS. In this study, we instead consider two approaches to utilizing *nonspeech* audio to either (1) quickly convey initial search-limiting location information, or (2) continuously and precisely guide motor movements.

1.1.1. Area Cueing Approach

One form a nonspeech targeting display could take is a discrete, informational *area cue* that informs the user in which area to search for the target object. Such a system could be implemented using information retrieved from a database about expected object locations, or in response to one-time machine-vision recognition of a target object. Systems of this nature could reduce target acquisition times by reducing the space that must be searched. However, they would not assist with the second stage of targeting in which the object must be precisely located and targeted.

Several area cueing systems have been considered in prior work. Chincha and Tian [11] developed a system in which users issued voice commands to initiate machine-vision search for target objects. If the object was in the camera's field of view, a sound confirmed its presence. Schauerte et al. [12] developed a machine-vision-based "lost object finding" system. That system sonified objects within the upper camera-viewable area with higher pitched tones, and lower-area objects with lower pitched tones. Tempo was mapped to object location confidence, and left-right location was represented through sound panning. Users gave the system generally positive ratings.

The effectiveness of an area cue may depend on its ability to swiftly and correctly communicate spatial information, to allow a user to immediately begin moving their extremity toward the target area without waiting to interpret more elaborate TTS or nonspeech displays. As such, choosing sounds that match expectations is crucial to optimizing this information transfer. It has consistently been found that higher pitched sounds tend to be associated with more highly elevated objects, and that lower pitched sounds tend to be associated with lower objects [13,14,15]. This pitch-elevation mapping reflects a statistical regularity of acoustic scenes [16, 17]. Thus, for the area cue sonification evaluated in this study, cue pitch was used to quickly communicate the elevation of the area in which the target resided.

1.1.2. Proximity Feedback Approach

Alternatively, a system could guide the entire process of object targeting by displaying an ongoing sonification of the user's hand position relative to the target. This would allow the user to target the object in 3D space solely through the sonification. While a system of this nature could represent relative position in three dimensions, in this study we considered a simpler, unidimensional display that provided continuous *proximity feedback*. The proximity feedback paradigm is similar to the real-time sonification of human movement, which has been shown to be effective for athletes and others endeavoring to carry out complex, precise movements, even when visual feedback was also available [18]. Unlike area cueing, proximity feedback supports the entire targeting task. However, it could also become distracting in environments with some visibility, and has significant technical requirements such as wearable sensors or cameras.

Displaying proximity feedback entails representing a dynamically changing variable: the current distance from the user's hand to the target. Higher pitch and tempo tend to be conceptually associated with closer proximity, as well as the related property of urgency [19, 20]. As such, in the proximity feedback design tested in this study, pitch and tempo communicated the proximity of the participant's hand to the target as it moved about in 3D space.

1.2. Current Study

The goal of this study was to investigate the effectiveness of area cues and proximity feedback in facilitating object targeting in local space. Participants were asked to walk around a virtual kitchen (Figure 1), and physically reach for target objects, with assistance from either an area cueing display, a proximity feedback display, both displays at once, or without assistance, in either a low visibility or a no visibility environment.

2. METHOD

2.1. Participants

There were 40 participants, with a mean age of 21 (SD = 3.23). 27 were male, 10 were female, and 3 elected not to specify. All were undergraduates at a technical university in the southeast United States. Participants reported normal/corrected vision and hearing, and had sufficient mobility/ dexterity to complete study tasks.

2.2. Materials

2.2.1. Virtual Environment

The experiment took place in a virtual environment created in Unity[1]. SteamVR[2] was used to support an HTC Vive VR system. The Unity scene ran on a control computer, which streamed video and audio to the Vive head-mounted display, as well as haptics to a handheld controller. This controller was also used to track the participant's hand position, and accept button-press responses from the participant. Audio was spatialized using the Steam Audio Unity asset, which provides real-time blended HRTF and acoustic simulation effects[3]. The software automatically recorded performance data. The rendered environment consisted of a kitchen-like room approximately 3×3 meters in size (Figure 1). There were two drawer-countertop-cupboard "stacks" along each of the four walls, making for a total of eight possible 2D locations.

Within each of the eight kitchen stacks, a target could exist within three elevation areas: low (in one of the drawers), middle (on the counter), or high (on a shelf within the cabinet, see Figure 2). Each of these areas was populated with 2–5 distractor objects near the target. Distractor objects were plates, coffee mugs, bowls,

[1] https://unity3d.com/

[2] https://steamcommunity.com/steamvr

[3] https://valvesoftware.github.io/steam-audio/

Figure 1: Virtual study environment.

glasses, and white cylinders. At the start of each trial, one of the white cylinders was replaced by a white capsule (Figure 2), which was the target object. Thus, the distractors were all visually similar to the target, to the point where participants in the low visibility conditions would need to move their head close to the objects to tell if the target was present in that area, and/or which object was the target. While participants could complete the task in this way, they could also elect to use the auditory displays to determine the target object's general location or guide targeting.

Figure 2: A kitchen stack, with the target (capsule, center) and distractors (plates, glasses, bowls, mugs, and cylinders).

2.2.2. *Auditory and Haptic Displays*

The 2D navigation beacon was a tone that was spatialized to "point" in the direction of the target stack. Its tempo increased as the participant approached the target stack, similar to [1].

The area cue was a brief sound played just after the participant entered the capture radius of the target stack. One of three variations was played depending on whether the target was in the middle elevation area (countertop), the high elevation area (cupboard) or the low elevation area (drawers).

The area cue was designed to strike a balance between clarity, brevity, and appropriate continuity with the 2D navigation experience. As such, each cue variant was constructed as a composite of several copies of the 2D beacon sound. Some of these copies were pitch shifted up or down, with the original 2D navigation sound always included. This produced a "chord," including the 2D beacon sound as the highest, middle, or lowest comprising note. For the middle elevation area cue, the 2D beacon sound was played alongside components both one octave higher and one octave lower. For the high area cue, components were added that were pitch-shifted upward by up to two octaves. For the low area cue, components were added that were pitch-shifted down by up to two octaves. The higher or lower pitched components faded in gradually over the course of a half second, creating a transition between the 2D beacon and area cue.

The proximity feedback was implemented as a repeating tone whose pitch and tempo changed depending on the proximity of the hand to the target. At maximum range, the tone played approximately once a second, and at minimum range it played approximately 10 times per second, and was one octave higher in pitch. Thus, increasing proximity was displayed via rising pitch and increasing tempo.

In the conditions with both the area cue and the proximity feedback, the area cue played once upon capture radius entry, and then the proximity feedback began playing normally. In order to simulate the ability of a person to search for an object by feeling object contours, the Vive's haptic feedback capabilities were utilized. When the handheld controller (Figure 3) was inside an object, the participant felt a continuous vibration. This vibration was given one of three strengths, depending on the type of virtual object that the participant's hand was inside of.

If the participant's hand was inside of a wall or kitchen structure such as a cabinet or drawer, they felt a weak vibration. If it was inside of a distractor object, they felt a medium vibration. Finally, if the participant's hand was inside of the target object, they felt a strong vibration. Vibration strengths were different enough to be clearly discriminable to a person with typical tactile acuity. In the No Visibility + No Sound condition, participants relied entirely on this haptic information.

The two visibility levels were created using Unity post-processing effects. In conditions with no visibility, post-processing was activated to make the scene completely dark. However, participants were able to see a blue box representing the floor, and a blue wireframe representing the virtual safety boundary. In low visibility conditions, a depth of field effect was applied in order to simulate generic limited visibility conditions. This effect caused objects to appear too blurry for a viewer to resolve precise form at most ranges. From a distance, participants could see the contours of the cabinets, and perceive that objects were present, but could not discriminate between targets and distractors without leaning in closer. Objects only resolved completely when viewed within a distance of approximately 15cm. Instead of leaning closer, which was physically effortful, participants were also able to utilize haptic feedback, or the auditory targeting displays, or could repeatedly guess.

2.3. Procedure

Upon consenting to participate, participants were fitted with the virtual reality headset and instructed in the task. Participants first practiced completing the task in a full visibility training mode. Each condition consisted of a set of object targeting trials. After each condition, participants were given an iPad, which they used to complete the NASA TLX, which assesses subjective workload associated with a task [21]. After completing all eight conditions, participants filled out a demographic questionnaire.

Each trial consisted of two stages. First, the participant used the 2D auditory beacon to walk to the kitchen stack that contained

the target. This procedure was included to increase the validity of the targeting task, and the 'virtual room' paradigm. Upon entering the 0.75-meter capture radius of the target stack, the 2D navigation beacon ceased.

During the second stage, the participant was instructed to find the target object as quickly and accurately as possible, using the different 3D assistance sounds (Figure 3). Doing this required moving the handheld controller, so that it was within the target object, and depressing the trigger on the controller to simulate grasping the target. The three sound types provided different forms of assistance during this second stage of each trial.

Figure 3: Tracked space and participant view during full visibility training.

Typically, humans make goal-directed movements in two parts. First, a large, rapid movement is undertaken that often falls short of the target. Second, after a moment of information uptake, a smaller and slower movement is undertaken to refine the limb position and reach the target [22, 23]. In this study, if the area cue was present, participants could first make a rapid, imprecise movement into the vicinity of the countertop, cupboard, or drawers, as specified by the area cue. Whether or not they heard the area cue, participants ultimately had to determine which of the objects was in fact the target, and guide the controller precisely to it. The proximity feedback assisted with this by providing a continuous sonification of the controller's distance from the target as the controller moved.

In the No Visibility + Area Cue and No Visibility + No Sound conditions, the nature of the targeting task was qualitatively different. Because visibility was zero, participants needed to use the haptic information to determine the layout of the stack and/or to disambiguate targets from distractors. During pilot testing, participants were capable of completing the task in these two conditions, but found it frustrating and time consuming. In response, a 'timeout' procedure was implemented. If a participant took over a minute to complete a trial, the system moved on to the next trial and recorded a 'timeout.' Data were not analyzed for these timed-out trials.

Upon pulling the controller's trigger while it occupied the same virtual space as the target, participants heard a confirmation sound and the next trial began. In the case of a timeout, the next trial began without the confirmation sound.

The target was placed in each of the 8 stacks, 3 times (one each for low, medium or high areas), for a total of 24 trials per condition. The order of trials was randomized. To avoid confusion, participants never had to navigate to the same stack twice in a row, nor to either of the immediately adjacent stacks.

2.4. Experiment Design

There were two independent variables, Sound Type and Visibility Level. Visibility Level could be either no visibility or low visibility (Table 1). Sound Type could be either no sound, area cue, proximity feedback, or the area cue with subsequent proximity feedback (AC+PF). Each participant experienced all of the resulting eight experimental conditions in a single session. The order of conditions was counterbalanced.

		Visibility Level	
		No Visibility	Low Visibility
Sound Type	No Sound	No Sound + No Visibility	No Sound + Low Visibility
	Area Cue	Area Cue + No Visibility	Area Cue + Low Visibility
	Proximity Feedback	Prox. Feed + No Visibility	Prox. Feed + Low Visibility
	Area Cue + Proximity Feedback (AC+PF)	AC+PF + No Visibility	AC+PF + Low Visibility

Table 1: Conditions experienced by each participant.

2.4.1. Dependent Variables

Six dependent variables were measured. Task time was measured as the elapsed time from the moment the trial began to the moment the participant found the target. Hand travel distance was measured as the distance the participant's hand traveled from the start of the targeting task, to when it reached the target. A shorter hand travel distance indicated that participants had moved their hand to the target more efficiently. The number of timeouts reflected the number of cases a participant took more than a minute to complete a task, generally reflecting the participant becoming lost or giving up. The number of errors was measured as a tally of instances in which the participant pulled the trigger on the handheld controller without it being within the target.

Although there was always sufficient information to avoid such errors, participants could "guess" by moving the controller and pulling the trigger without waiting to confirm if it was within a target. As such, this error count reflects frustration or impatience more than targeting accuracy. Finally, to assess subjective workload, a NASA TLX composite score was generated.

2.4.2. Hypotheses and Analyses

It was hypothesized that the sound types would have different effects depending on the level of visibility.

When no visibility was present, it was expected that the sound types that conveyed the most information about location of the target would perform better, with the AC+PF condition leading to the highest performance, followed by the proximity feedback, area cue, and then no sound conditions.

In the low visibility conditions, it was expected that the area cue would lead to the highest performance, due to the fact that it could provide helpful information without interrupting the task flow of participants who elected to target using the visuals.

Finally, it was hypothesized that all sound types would lead to decreased workload, relative to no sounds, and that these differences would be largest in the no visibility conditions.

For each dependent variable, a two-way Hyunh-Feldt repeated measures ANOVA was conducted, followed, when appropriate, by post-hoc paired Bonferroni t-tests. Post-hoc comparisons between the no visibility and low visibility conditions within each sound type showed significant differences in all cases, and are omitted for brevity. Test statistics for other post-hoc t-tests (represented in results tables) are also omitted.

3. RESULTS

3.1. Visibility Level

Across all dependent variables, participants performed significantly better in the low visibility conditions, compared to the no visibility conditions (Table 2).

	ANOVA Result	No Vis *M,* (*SD*)	Low Vis M, (SD)
Task Times (seconds)	$F(1, 25) = 280.47$, $p < .001$, $\eta_p^2 = .92$	34.50s (9.15)	8.84s (3.06)
Hand Travel Distance (decimeters)	$F(1, 25) = 150.20$, $p < .001$, $\eta_p^2 = .857$	12.90 dm (5.28)	2.45 dm (0.75)
Number of Timeouts	$F(1, 25) = 66.13$, $p < .001$, $\eta_p^2 = .726$	12.90 (5.27)	2.45 (0.36)
Number of Errors	$F(1, 25) = 91.89$, $p < .001$, $\eta_p^2 = .786$	25.60 (13.89)	1.91 (1.57)
Subjective Workload (0-100 Score)	$F(1, 31) = 91.07$, $p < .001$, $\eta_p^2 = .75$	42.64 (14.60)	23.77 (11.34)

Table 2: Results by Visibility Level.

3.2. Sound Type

Sound Type had an impact on targeting task times, $F(2.56, 63.99) = 70.10, p < .001, \eta_p^2 = .74$. As shown in Table 3, participants were substantially faster with all three types of sounds, compared to when no sounds were present. They took the shortest time when they heard either the proximity feedback or AC+PF. However, task times did not differ between the two conditions with proximity feedback, suggesting that participants did not receive meaningful benefits from the area cue when the proximity feedback was also present.

	No Sound	Area Cue	Prox. Feed.	AC+PF
Mean Time (*SD*)	28.01s (13.17)	25.16s (4.67)	17.06s (5.85)	16.46s (5.30)
Differs from:	Area Cue, Prox. Feed., AC+PF	No Sound, Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

Table 3: Task time (seconds) by Sound Type.

The distance that participants moved their hand to reach targets was affected by the type of sound that they heard, $F(1, 53.34) = 29.72, p < .001, \eta_p^2 = .535$. Table 4 shows that participants in the two proximity feedback conditions were twice as efficient with their movements toward the target, compared to the area cue and no sound conditions. However, as with other dependent variables, proximity feedback and AC+PF led to equatable performance. Hand travel distance was not different between the area cue and no sound conditions, perhaps reflecting the fact that area cued participants still had to do a significant amount of effortful haptic and/or low visibility visual search to precisely locate the targets.

	No Sound	Area Cue	Prox. Feed.	AC+PF
Mean Distance (*SD*)	11.33 dm (6.01)	9.44 dm (3.79)	4.91 dm (2.19)	5.01 dm (3.60)
Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

Table 4: Hand travel distance (decimeters) by Sound Type.

The number of times that participants timed out and failed to find the target was affected by the type of sound they heard, $F(2, 50.06) = 51.34, p < .001, \eta_p^2 = .673$. Table 5 shows that the two conditions containing proximity feedback both led to fewer timeouts than the no sound and area cue conditions. However, performance was not different between these two conditions.

	No Sound	Area Cue	Prox. Feed.	AC+PF
Mean Timeouts (*SD*)	6.96 (3.39)	5.89 (3.94)	1.40 (2.63)	1.48 (3.12)
Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

Table 5: Hand travel distance (decimeters) by Sound Type.

The number of errors made by participants was impacted by the type of sounds they heard, $F(1.78, 44.39) = 51.34, p < .001, \eta_p^2 = .715$. Table 6 shows that, when participants heard either the proximity feedback alone, or AC+PF, they committed fewer errors than when they heard either the area cue or no sound. It was observed that participants tended to "guess" more often in the no sound and area cue conditions, thus increasing error count.

	No Sound	Area Cue	Prox. Feed.	AC+PF
Mean Errors (*SD*)	26.98 (15.44)	21.33 (12.49)	3.03 (10.64)	3.70 (11.66)
Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

Table 6: Number of errors (count per trial) by Sound Type.

Subjective workload was impacted by the type of sound that participants heard $F(2.23, 69.02) = 19.13, p < .001, \eta_p^2 = .382$. Table 7 shows that, when participants heard either proximity feedback or AC+PF, they reported lower workload, compared to when they heard the area cue or no sound. However, when participants heard the area cue only, they reported the same level of workload as when they heard no sound. This suggests that utilizing the area cue to limit subsequent search area was less impactful on perceived workload compared to the difficulty of carrying out the subsequent targeting movement without assistance from the proximity feedback.

	No Sound	Area Cue	Prox. Feed.	AC+PF
Mean Score (*SD*)	37.80 (15.19)	37.61 (14.04)	28.53 (10.82)	28.88 (12.24)
Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

Table 7: Subjective workload (NASA TLX, 0-100) by Sound Type.

3.3. Interaction Effects

For all dependent variables, the effect of Sound Type depended on Visibility Level. Overall, Sound Type was more impactful in the no visibility conditions. This was likely because these participants tended to rely on the sounds, in particular the proximity feedback or AC+PF. However, the sounds still led to some performance benefits in the low visibility conditions.

The effect of Sound Type on task times depended on Visibility Level, $F(2.47, 61.63) = 28.18, p < .001, \eta_p^2 = .530$, see Table 8. Sound Type impacted task times in the no visibility conditions, but not in the low visibility conditions.

		No Sound	Area Cue	Prox. Feed.	AC+PF
No Vis	Mean Time (*SD*)	45.85s (8.73)	40.96s (12.40)	25.41s (9.91)	26.26s (11.80)
	Differs from:	Area Cue, Prox. Feed., AC+PF	No Sound, Prox. Feed., AC+PF	No Sound	No Sound

Table 8: Task times (seconds) by Sound Type by Visibility Level.

The effect of Sound Type on hand travel distance depended on Visibility Level, $F(2.25, 56.34) = 21.54, p < .001, \eta_p^2 = .463$, see Table 9.

The effect of Sound Type on timeout count depended on Visibility Level, $F(2.14, 53.57) = 46.61, p < .001, \eta_p^2 = .651$. The number of timeouts differed in the proximity feedback and AC+PF conditions compared to the no sound and area cue conditions when where was no visibility, but when low visibility was present, there were no significant differences (Table 10).

As shown in Table 11, the effect of Sound Type on the number of errors depended on Visibility Level, $F(1.73, 43.16) = 46.78, p < .001, \eta_p^2 = .652$.

		No Sound	Area Cue	Prox. Feed.	AC+PF
No Vis	Mean Distance (*SD*)	18.85 dm (9.81)	16.20 dm (6.26)	7.31 dm (3.48)	7.75 dm (5.97)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound	No Sound
Low Vis	Mean Distance (*SD*)	3.03 dm (1.05)	2.60 dm (1.10)	2.12 dm (0.55)	1.94 dm (0.42)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound	No Sound

Table 9: Travel distance (decimeters) by Sound Type by Visibility Level.

		No Sound	Area Cue	Proximity Feedback	AC+PF
No Vis	Mean Timeouts (*SD*)	13.60 (6.08)	11.37 (7.10)	2.50 (4.50)	3.53 (5.97)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

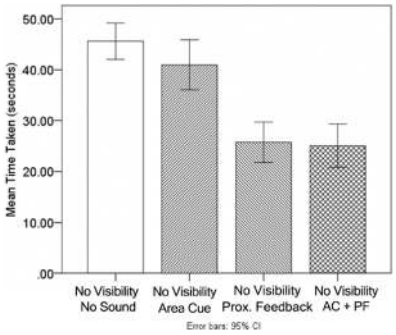
Table 10: Timeouts by Sound Type by Visibility Level.

The effect of Sound Type on subjective workload also depended on Visibility Level, $F(2.54, 78.88) = 4.79, p = .006, \eta_p^2 = .134$. When visibility was low, there were fewer significant differences between conditions (Table 12).

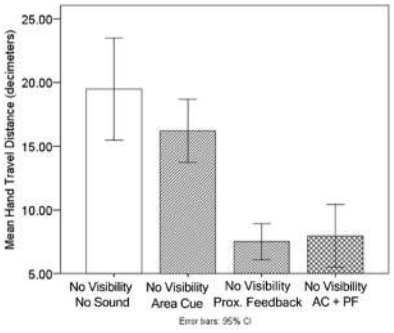
4. DISCUSSION

In this study, three auditory display approaches were evaluated in terms of their ability to assist with finding nearby objects in limited visibility conditions. Using a VR targeting task, the proximity feedback display was found to be most effective at increasing performance and improving the subjective experience of object targeting with limited visibility. The area cue was less effective at achieving these goals, and notably did not lower subjective workload, but did improve performance via several metrics. When both sound types were used in tandem (AC+PF), results were the same as when proximity feedback was used exclusively, indicating that area cue displays may have limited utility when continuous audiomotor feedback can be provided. This pattern of results was similar for both levels of visibility, but less pronounced in the low visibility conditions.

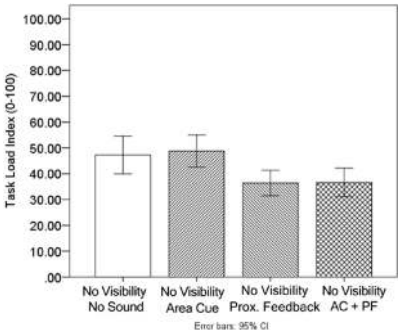
In the no visibility conditions, proximity feedback and AC+PF both led to large improvements across dependent variables (Figures 4a, 4b, and 4c). Notably, the proximity feedback led to a tenfold decrease in errors, indicating that participants were less likely to adopt a "guessing" strategy. Decreases in hand travel distance and task times indicate that, overall, participants were able to utilize the proximity feedback to move more efficiently to the target. The area cue was also effective at increasing targeting performance, but less so than expected, and not via all metrics. Notably, the area cue did not lead to a reduction in workload (Figure



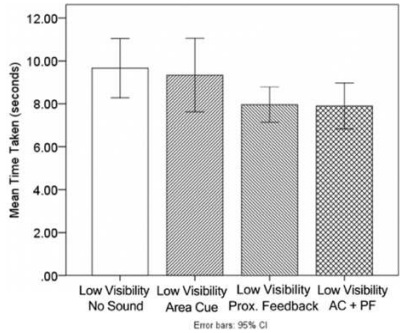
(4a) Mean task times for each sound type, no visibility conditions.



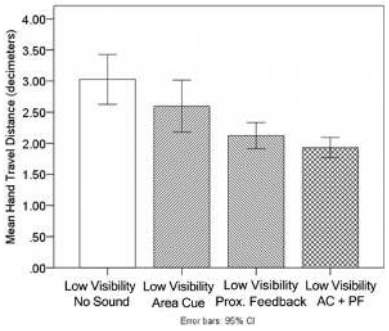
(4b) Mean hand travel distance for each sound type, no visibility conditions.



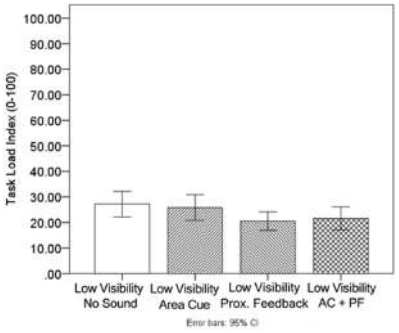
(4c) Subjective workload for each sound type, no visibility conditions.



(5a) Mean task times for each sound type, low visibility conditions.



(5b) Mean hand travel distance for each sound type, low visibility conditions.



(5c) Subjective workload for each sound type, low visibility conditions.

		No Sound	Area Cue	Prox. Feed.	AC+PF
No Vis	Mean Errors (*SD*)	52.24 (28.8)	40.12 (22.56)	5.83 (6.55)	6.39 (7.27)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue
Low Vis	Mean Errors (*SD*)	3.37 (2.16)	2.63 (1.92)	1.02 (1.44)	0.83 (1.92)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue

Table 11: Errors by Sound Type by Visibility Level.

		No Sound	Area Cue	Prox. Feed.	AC+PF
No Vis	Mean Score (*SD*)	47.25 (21.21)	48.78 (18.05)	36.44 (12.25)	36.63 (16.07)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed., AC+PF	No Sound, Area Cue	No Sound, Area Cue
Low Vis	Mean Score (*SD*)	27.26 (14.66)	25.84 (14.62)	20.57 (10.69)	21.60 (13.19)
	Differs from:	Prox. Feed., AC+PF	Prox. Feed.	No Sound, Area Cue	No Sound

Table 12: Workload (NASA TLX, 0-100) by Sound Type by Visibility Level.

4c). While the area cue should have reduced the amount of effort required by a full two thirds, these results suggest that the primary determiner of both subjective workload and task performance was whether or not the participant had to perform the laborious task of object targeting using only tactile information.

In the low visibility conditions, a similar pattern was present: benefits were observed for all sound types compared to no sound. However, the magnitude of the advantage, as well as differences between the displays, was less pronounced compared to when there was no visibility (Figures 5a, 5b, and 5c). This compression of differences suggests that participants utilized visual input when it was available. However, there were still significant performance benefits when the auditory displays were active, as well as a reduction in subjective workload associated with the proximity feedback and AC+PF conditions (Figure 5c). This indicates that persons who are able to complete a targeting task with limited but usable visual input can still be expected to benefit from the pres-

ence of auditory targeting displays, in terms of both performance and workload.

4.1. Conclusion

The three auditory displays evaluated in this study were effective at increasing object targeting performance, and should be incorporated into virtual or mixed reality systems that endeavor to assist humans in limited visibility conditions, depending on the technical abilities of each system and needs of the task. Providing proximity feedback with which motor movements can be guided should be considered when feasible, rather than solely utilizing area cueing displays. Incorporating auditory targeting displays of the types discussed here into future systems could increase the usability of everyday environments without visual input, and support task performance in a variety of low visibility situations.

5. REFERENCES

[1] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert. 2007. "SWAN: System for Wearable Audio Navigation." In *Wearable Computers, 2007 11th IEEE International Symposium on Wearable Computers*, IEEE, Boston, MA. 491–98.

[2] J. Loomis, R. D. Golledge, and R. L. Klatzky. 2001. "GPS-based navigation systems for the visually impaired." In *Barfield W, Caudell T, editors. Fundamentals of wearable computers and augmented reality*. Lawrence Erlbaum. Mahway, NJ. 429–446.

[3] K. R. May, X. Ma, P. Roberts, and B. N. Walker. (Under Review). "Spotlights and Soundscapes: Participatory Design of Mixed Reality Auditory Environments for Persons with Visual Impairment."

[4] R. Yaagoubi, T. Badard, and G. Edwards. 2009. "Standards and Spatial Data Infrastructures to help the navigation of blind pedestrian in urban areas." *Urban and Regional Data Management* (February 2009).

[5] C. Prablanc, J.E. Echallier, M. Jeannerod, and E. Komilis. 1979. "Optimal response of eye and hand motor systems in pointing at a visual target." *Biological Cybernetics* 35, 3 (1979), 183–187.

[6] J. C. Middlebrooks. 1991. "Sound Localization by Human Listeners." *Annual Review of Psychology* 42, 1 (January 1991), 135–159.

[7] D. R. Begault, E. M. Wenzel, and M.R. Anderson. 2001. "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source." *Journal of the Audio Engineering Society*, 49(10), 904–916.

[8] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. 1993. "Localization using nonindividualized head-related transfer functions." *The Journal of the Acoustical Society of America*. 94, 1, 111–123.

[9] K. Thakoor, N. Mante, C. Zhang, C. Siagan, J. Weiland, L. Itti, and G. Medioni. 2015. "A System for Assisting the Visually Impaired in Localization and Grasp of Desired Objects." *Computer Vision—ECCV 2014 Workshops Lecture Notes in Computer Science* (2015), 643–657.

[10] I. A. Doush, S. Alshatnawi, A. Al-Tamimi, B. Alhasan, and S. Hamasha. 2016. "ISAB: Integrated Indoor Navigation System for the Blind." *Interacting with Computers* (2016).

[11] R. Chincha and Y. Tian. 2011. Finding objects for blind people based on SURF features. *2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)* (2011).

[12] B. Schauerte, M. Martinez, A. Constantinescu, and R. Stiefelhagen. 2012. "An Assistive Vision System for the Blind That Helps Find Lost Things." *Lecture Notes in Computer Science Computers Helping People with Special Needs* (2012), 566–572.

[13] C.C. Pratt. 1930. "The spatial character of high and low tones." *Journal of Experimental Psychology* 13, 3 (1930), 278–285.

[14] K. Evans and A. Treisman. 2011. "Natural cross-modal mappings between visual and auditory features." *Journal of Vision* 10, 1 (June 2011), 6–6.

[15] K. Pisanski, S. G. Isenstein, K. J. Montano, J.M. O'Connor, and D. R. Feinberg. 2017. "Low is large: spatial location and pitch interact in voice-based body size estimation." *Attention, Perception, & Psychophysics* 79, 4 (2017), 1239–1251.

[16] C. V. Parise and C. Spence. 2009. "'When Birds of a Feather Flock Together': Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes."

[17] C. V. Cesare V. Parise, K. Knorre, and M. O. Ernst. 2014. "Natural auditory scene statistics shapes human spatial hearing." Proceedings of the National Academy of Sciences 111, 16 (July 2014), 6104–6108.

[18] A. O. Effenberg. 2005. "Movement sonification: Effects on perception and action." *IEEE Multimedia*, 12(2), 53–59.

[19] B. N. Walker. 2007. "Consistency of magnitude estimations with conceptual data dimensions used for sonification." *Applied Cognitive Psych*. 21, 5. 579–599.

[20] J. Edworthy, E. J. Hellier, & R. Hards. 1995. "The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals." *Ergonomics*, 38(11), 2341–2361.

[21] J. Lyons, S. Hansen, S. Hurding, and D. Elliott. 2006. "Optimizing rapid aiming behaviour: movement kinematics depend on the cost of corrective modifications." *Experimental Brain Research* 174, 1 (2006), 95–100.

[22] M.D. Byrne, M.K. O'malley, M.A. Gallagher, S.N. Purkayastha, N. Howie, and J.C. Huegel. 2010. "A preliminary ACT-R model of a continuous motor task." *PsycEXTRA Dataset* (2010).

[23] S. G. Hart and L. E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." *Advances in Psychology Human Mental Workload* (1988), 139–183.