SPATIAL AUDITORY DISPLAYS: SUBSTITUTION AND COMPLEMENTARITY TO VISUAL DISPLAYS

Elizabeth M. Wenzel

Advanced Controls and Displays Group,
NASA Ames Research Center,
Moffett Field, CA, 94035, USA
Elizabeth.M.Wenzel@nasa.gov

Martine Godfroy-Cooper, Joel D. Miller

San Jose State University Foundation,
Advanced Controls and Displays Group,
NASA Ames Research Center,
Moffett Field, CA, 94035, USA
Martine.Godfroy-1@nasa.gov
Joel.D.Miller@nasa.gov

ABSTRACT

The primary goal of this research was to compare the performance in localization of stationary targets during a simulated extra-vehicular exploration of a planetary surface. Three different types of displays were tested for aiding orientation and localization: a 3D spatial auditory display, a 2D North-up visual map, and the combination of the two in a bimodal display. Localization performance was compared under four different environmental conditions combining high and low levels of visibility and ambiguity. In a separate experiment using a similar protocol, the impact of visual workload on performance was also investigated contrasting high (Dual-Task paradigm) and low workload (Single Orientation task). A synergistic presentation of the visual and auditory information (bimodal display) lead to a significant improvement in performance (higher percent correct orientation and localization, shorter decision and localization times) compared to either unimodal condition, in particular when the visual environmental conditions were degraded. Preliminary data using the dual-task paradigm suggest that the performance with displays utilizing auditory cues was less affected by the extra demands of additional visual workload than a visual-only display.

1. INTRODUCTION

During extra vehicular activities (EVA) on surface, astronauts must maintain situational awareness of a number of spatially distributed "targets": other team members (both human and robotic), rovers, habitats and other critical resources. These targets are often outside the astronaut's immediate field of view (FOV) or are too distant to be visible from current location. Further, since visual resources may be needed for other task demands during EVA, alternate modalities such as auditory displays will be advantageous for supplementing the visual channel and acting as a critical backup if visual systems fail.

The use of acoustic cues for orienting, navigation, and way finding, particularly under conditions that are hazardous and/or lack adequate visual information can be found in the literature since at least the 1930s. Recently, Simpson et al. [1], [2] described the use of acoustic orientation beacons in the cockpit, particularly during emergency situations causing pilot spatial disorientation that are exacerbated by the lack of a visual horizon. Wijngaarden et al. [3] described a similar beacon display for evacuation during a fire under limited visibility conditions. Loomis et al. [4] investigated the use of spatial auditory information displayed interactively in real-time for navigation, particularly for the visually impaired. Other work includes navigation displays for the military [5] and the SWAN system for the visually impaired [6], [7]. All have proposed incorporating directional information from GPS or other location-tracking devices in such systems.

Previous efforts at NASA Ames demonstrated a number of “orientation beacon” displays developed for situation awareness (SA) during EVA [8], [9]. For example, one auditory display prototype simulated an augmented reality auditory display for an astronaut conducting an EVA on the moon (using the *slab3d* spatial audio rendering software; http://slab3d.sonisphere.com). Three auditory beacons assisted the astronaut in locating a rover, the lander, and another astronaut ("partner"). Voice commands were used to interact with the display.

The work presented here focuses on the experimental evaluation of a revised beacon display prototype, an audio-visual simulation of a spatial audio augmented-reality display for telerobotic planetary exploration on Mars. A primary goal of the evaluation studies was to compare performance with different types of display modalities for aiding orientation during exploration.

In Study 1 (Single Orientation Task), the effects of visual degradation and visual ambiguity were investigated. Here, the task was performed under four different environmental visual conditions defined by their visibility level (low vs. high) and/or their ambiguity level (low vs. high). (Some aspects of Study 1 have been previously reported in [10].)

In Study 2 (Dual Task), the effect of visual workload was tested in a dual task scenario in a high visibility/ low ambiguity condition where the participants were required to monitor and respond to meters displaying visual information about the levels of EVA mission consumables (carbon dioxide, oxygen, water, and battery), while simultaneously performing the orientation task.

Performance defined by the percentage of correct orientation (forced-choice left/right paradigm), left/right decision time (LRDT), percentage of correct localization and localization time (LT) was estimated with three different types of navigation aids (NavAids): a 3D spatial auditory display (A), a 2D North-Up visual map (V), and the combination of the two in a bimodal display (B). The different metrics were expected to reveal how the nature of the information available would impact the two phases of the task, i.e., decision and localization.

The decision task was designed to evaluate the ability of the operators to represent their own position as well as the position of the different entities (such as rovers, other astronauts or habitat) on the surface, and to maintain and update this representation as they were following a predetermined path. The localization task was intended to test the usability of an "auditory localizer", i.e., a localization aid functioning like a Geiger counter, available only in the A and the B conditions.

In the A condition, in which the soundscape (auditory scene) was continuously presented, the reference frame was egocentric (craniocentric [11]), i.e., locations were represented with respect to the particular perspective of the perceiver. In this condition, called 3D directional mapping, the reference direction and the operator's heading merge and the operator's localization capabilities depend essentially on the spatial resolution of the auditory system. Auditory spatial acuity is poorer by up to two orders of magnitude (minimum audible angle [MAA]: 1° to 2° for the frontal position, 6° to 7° at the rear [12], [13] compared to the visual domain (1 min of angle [14]). However, the auditory environment has the advantage of extending in all directions around the observer, while the visual environment is necessarily restricted to frontal regions.

In the V condition, the spatial arrangement of the scene was coded in an allocentric reference frame, where the entities are represented within a framework external to the holder of the representation. The 2D North-Up visual map contains both a locational representation, conveying the location of the entities in space and a heading representation, conveying the heading of the operator in space in a North-up allocentric reference frame.

Specific predictions for the decision task were that the performance in the V condition would be affected by the mental transformation associated with the differences between the reference frame and the operator's heading while performance in the A condition would only be affected by the perceptual limitations of audition. Further, it was expected that the B condition would result in a compromise between performance in the A and V conditions. To test these hypotheses, analyses of the percentage of correct orientation and of the LRDTs were performed. For this part of the task, no effect of the visual environment was expected, except a potential effect of the type of symbology used on the 2D map, iconic vs. orthographic.

For the localization task, it was expected that LTs in the V condition would be more affected by the degradation/s of the visual environment than in the A condition. The magnitude of the performance drop in the V condition should be inversely proportional to the level of degradation of the visual environment (low visibility alone, high ambiguity alone or combined low visibility and high ambiguity), while conversely, the gain (multisensory response enhancement, MRE) in the B condition should also be inversely proportional to the level of degradation of the visual environment, as stated by the principle of inverse effectiveness [15].

In Study 2, the hypothesis was that increased visual workload would lead to longer RTs, with a greater impact expected in the V condition.

Traditional analyses based on mean RTs do not take the distribution's shape into account and, for that reason, may obscure some aspects of the performance. An ex-Gaussian (ExG) decomposition of the RT distributions using a 3-parameter model based on the convolution of a normal distribution (*mu*: average performance and *sigma:* variability in performance) with an exponential distribution (*tau:* extremes in performance) can inform whether latent variables influence the different parameters [16]. Here, the hypothesis was that the distributions of the LRDTs obtained for the A, V and B conditions of presentation of the display would provide some insight into the cognitive processes that may produce differences in the parameter means. One of the most widely used distinctions is between more stimulus-driven automatic (non-analytic) processes (Gaussian component) and more central, attentional (analytic) processes (exponential component) [17], [18]. However, there is no real consensus on the exact meaning of these parameters and some studies have proposed an opposite interpretation of *mu* and *tau* [19], [20]. Consequently, no specific prediction was made in terms of an explanation for potential differences in the parameter values as a function of the modality of presentation of the display.

2. METHOD: STUDY 1

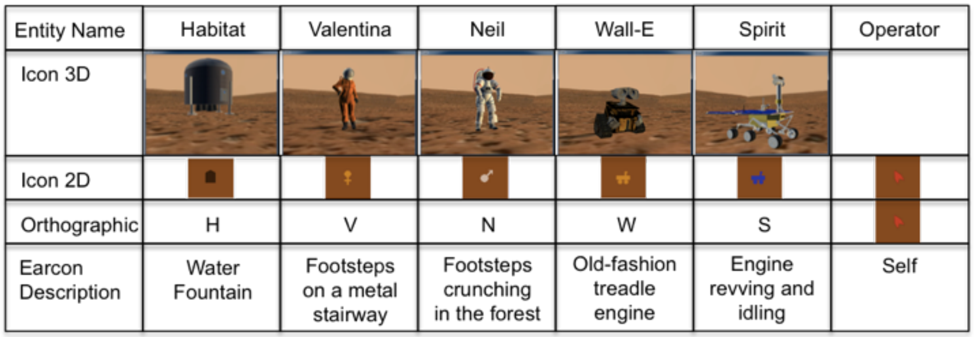
Using joystick control, participants steered a tele-robotically controlled rover in a 3D graphical simulation of a Mars plateau (100m x100m) presented on a flat Dell 22" display (resolution 1680 x 1050) in a full screen mode with a vertical FOV spanning 45° and a horizontal FOV of ~60° (Figure 2).

The scenario assumed that the participant was an astronaut controller remote from the physical site, but close enough that control latency was not a problem; e.g., the astronaut was elsewhere on the planetary surface. The participant saw and heard from the point of view of the remote rover. Five entities, two rovers, two astronauts, and a habitat populated the plateau (Figure 1). The 3D models of entities were from the Google 3D Warehouse and the Mars plateau was derived from NASA JPL images taken by the Mars Exploration Rover Spirit ("Hills Over Yonder"). All entities were stationary during the simulation scenarios. User motion in the 3D world was constrained to 2 degrees of freedom (DOF). The forward/backward variable speed of the participant driver/rover was a maximum of two eye-heights per second, i.e., 1.6 m/s. The joystick also controlled the yaw of the tele-robotically controlled rover at a variable rate, depending on joystick motion, with a maximum of 90°/s for steering, orientation and selection.

The characteristics of the three types of NavAids are summarized in Table 1. In the V and B display conditions, participants saw a 2D North-up navigation display/map representing the entire plateau, superimposed on the primary view of the virtual world in the bottom, left corner of the screen (see Figures 2 and 5).

Table 1: Navigation aid display characteristics.

	Auditory	Visual	Bimodal
3D simulation	100m x100m Mars plateau	100m x100m Mars plateau	100m x100m Mars plateau
NavAid	Spatial auditory display presented through stereo headphones	73mm x 73mm 2D North-up navigation display/map	A + V
Targets Symbology	Earcons (see Figure 1)	2D Visual icons (see Figure 1)	A + V
Reference frame	Egocentric	North-up Allocentric	A + V
Spatial Resolution	1°-2° frontal, 6° to 7° at the rear	1 min of visual angle	A + V
Audio Locator	Yes	No	Yes



Entity Name	Habitat	Valentina	Neil	Wall-E	Spirit	Operator
Icon 3D						
Icon 2D	■	■	■	■	■	▶
Orthographic	H	V	N	W	S	▶
Earcon Description	Water Fountain	Footsteps on a metal stairway	Footsteps crunching in the forest	Old-fashion treadle engine	Engine revving and idling	Self

Figure 1: Representation of the target entities in the 3D simulation.

The display showed the real-time location and orientation of the participant's rover and the location of the other entities in the scenarios using small visual icons (Figure 1). In visually ambiguous conditions, the unique visual icons were replaced with letters corresponding to the first initial of an entity's name.

In the A and B conditions, the participants heard a spatial auditory display in which each of the five entities emitted its unique sound icon ("auditory icon" or "earcons") played continuously and simultaneously in a kind of background soundscape.

The sounds were chosen so that they were easily discriminated from one another and reasonably pleasant to listen to on a continuous basis. The sounds also had a conceptual connection to each of the entities so that the association between the two was easy to learn.

The sounds were spatialized and looped in real time using the *slab3d* spatial audio rendering software [21] with non-individualized head-related transfer functions (HRTFs). Presented through circumaural stereo headphones (Sennheiser HD 595), the sounds appeared to emanate from the directions of each entity. The gain of the sounds decreased with distance according to the inverse-distance law for sound sources in the free field and was adjusted to provide approximately the same loudness when the listener was a nominal 35 m from each entity. The overall level of the sounds at the headphones was adjusted to a comfortable listening level.

In the localization phase of the task the participants were also provided with an "Audio Locator" presented spatially with its emitter placed at the same location as the target, at a constant volume, regardless of target-listener distance. The Audio Locator had two angular thresholds, one for white noise bursts which started playing at ±30° relative to the target's azimuth and one for a 400 Hz pure tone starting at ±5° (noise disabled). It was silent outside ±30°. Between ±30° and ±5°, the locator's noise burst period decreased linearly from 1000 to 200 ms (the burst fade-in/out time was 30 ms with a total duration of 80 ms).

System latency measurements indicated that this multimodal system performance was quite responsive compared to human perceptual thresholds: auditory display latency was 53.7±18 ms, visual display latency 117.5±19 ms and the visual-auditory latency was 63.9±4 ms.

2.1. Environmental Conditions

Four different environmental conditions were created to assess the effect of visual degradation on performance.

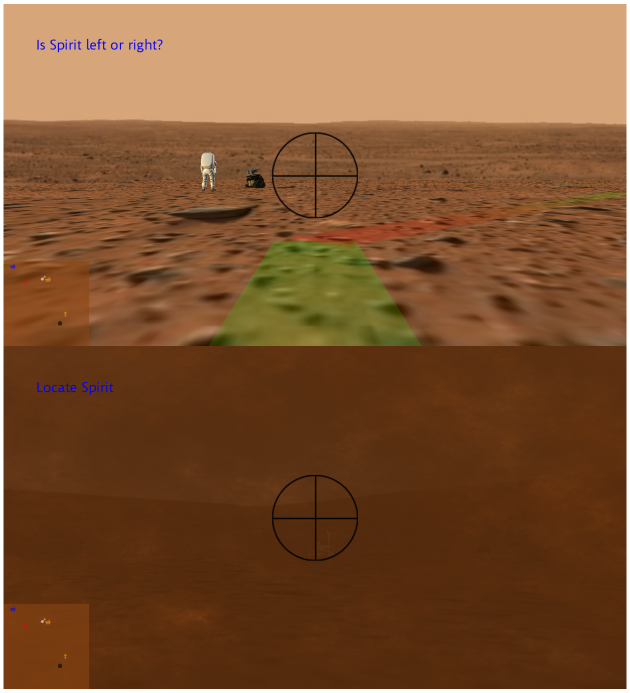
In the low visibility (LV) condition (Figure 2, bottom), a "dynamic" sand storm (randomly selected fractal cloud overlays refreshed every 1 Hz) partially occluded the outside view. In the high ambiguity (HA) condition, all the different entities looked like the Wall-E rover. The LV/HA condition combined both degradations.

2.2. Participants

A total of 48 paid volunteer males and females (aged 18 to 64) were recruited to participate in Study 1. They were randomly assigned to 4 groups of 12, with each group assigned to one of the 4 visual environment conditions (HVLA, LVLA, HVHA and LVHA). For Study 2, 6 additional (of 12 planned) male and female participants (aged 18 to 35) were assigned to the HVLA condition.

2.3. Experimental Task

The participant's task was to follow a virtual path consisting of six linear segments (~15 to 21 m/segment), with angular turns that ranged randomly from approximately 45° to 90°, superimposed over the primary visual display (Figures 2 and 5). Target events (trials) occurred 3 times during each path scenario and were triggered at pre-specified locations along 3 of the 6 possible path segments (maximum 1 trial per segment). For a given trial, a text message appeared on the visual display indicating the entity to locate, always outside the participant's FOV.


Is Spirit left or right?

Locate Spirit

Figure 2: Screenshots of the participant's (rover) viewpoint with the targeting reticle. Top: Decision task, High Visibility condition. Bottom: Localization task, Low Visibility condition. Note the 2D visual map on the bottom left of the screen displayed in the V and B conditions.

The task was separated into two phases. First, in the decision task, the operator had to decide whether the target was located to the left or to the right of his position ("Is Spirit left or right?", Figure 2, top) and to press a corresponding button on the joystick. In the localization task ("Locate Spirit", Figure 2, bottom), the participant was instructed to bring the reticle over the target before pushing a trigger to validate the response.

The objective dependent measures were: (1) % correct orientation, (2) LRDT to choose this initial direction, (3) % correct localization and (4) LT to find the target entity.

2.4. Experimental Conditions

The experiment was a three-factor design measuring performance for the display type (repeated measures factor) under 4 visual environment conditions (2 between subjects factors). First, the participants learned the appearance of the entities in the virtual scene, the corresponding visual icons on the V display, and the corresponding auditory earcons in the A display (Figure 1). During the training session, the participants were exposed to each of the display presentation modalities, i.e., A, V and B (5 trials/modality, 3 target events/trial). The presentation order of the trials was counterbalanced according to modality across subjects. During the experimental session, all participants were presented the same 6 blocks of trials (2 per modality, 5 unique trials per block, for a total of 90 experimental events) in a unique randomized order within modality. The duration of an experimental session lasted less than an hour (breaks were allowed at the participants' request).

2.5. Data Analysis

Percentage of correct responses for the Decision task was compared for the three presentation modalities. The responses categorized as "incorrect" and/ or "incongruent" (correct direction chosen but localization motion was opposite to the decision) were excluded from the dataset. The data were then filtered such that responses with mean LRDTs >= 8 s for each participant and each condition were removed, resulting in 4.56% of trials being excluded in total. The Gaussian (*mu*, *sigma*), and the exponential components of the distributions (*tau*) were computed for each subject. The mean ($\bar{x} = \mu + \tau$) and standard deviation ($s = \sqrt{\sigma^2 + \tau^2}$) of the ex-Gaussian Probability Density Function were derived from these initial parameters [23]. In addition to the initial data-filtering for LRDTs, LT values < 1s and >= 10 s were also excluded, corresponding to an additional 0.45% of the data. LRDTs and LTs were analyzed with factors of visibility level (HV, LV), ambiguity level (LA, HA) and modality (A, V, B). For the analysis of the redundant signal effects (RSE), the magnitude of the effect (i.e., the redundancy gain) was calculated by subtracting the bimodal RT from the RT for the fastest unimodal condition. These objective measures were analyzed via factorial and repeated measures ANOVAs, paired *t*-tests and post-hoc Bonferroni-Dunn tests.

3. RESULTS: STUDY 1

3.1. Percent Correct Orientation

Overall, the percentage of correct orientation judgments (Table 2) was very high (89.76%). Performance was best for the B condition (91.17%), and was significantly lower in the A condition (87.36%) when compared to either the B or V conditions (90.89%) (A, B: X_1^2=10.92, *p*=.009; A, V: X_1^2=9.24, *p*=.002; V, B: X_1^2=.07, *p*=.78. Not surprisingly, the likelihood of congruent motor responses was significantly higher when the L/R decision responses were also correct (X_2^2=529.28, *p*<.0001; correct=99.76%, wrong=84.38%).

When the left/right response was correct, the proportion of congruent motor response was equivalent between modalities (X_2^2 =.65, *p*=.72; A=99.8%, B=99.7%, V=99.8%).

However, when the response was wrong, the percentage of incongruent motor responses was significantly higher in the V and in the B conditions than in the A conditions (X_1^2=27.90, *p*<.0001; A=4.9%, B=18.9%, V=26%).

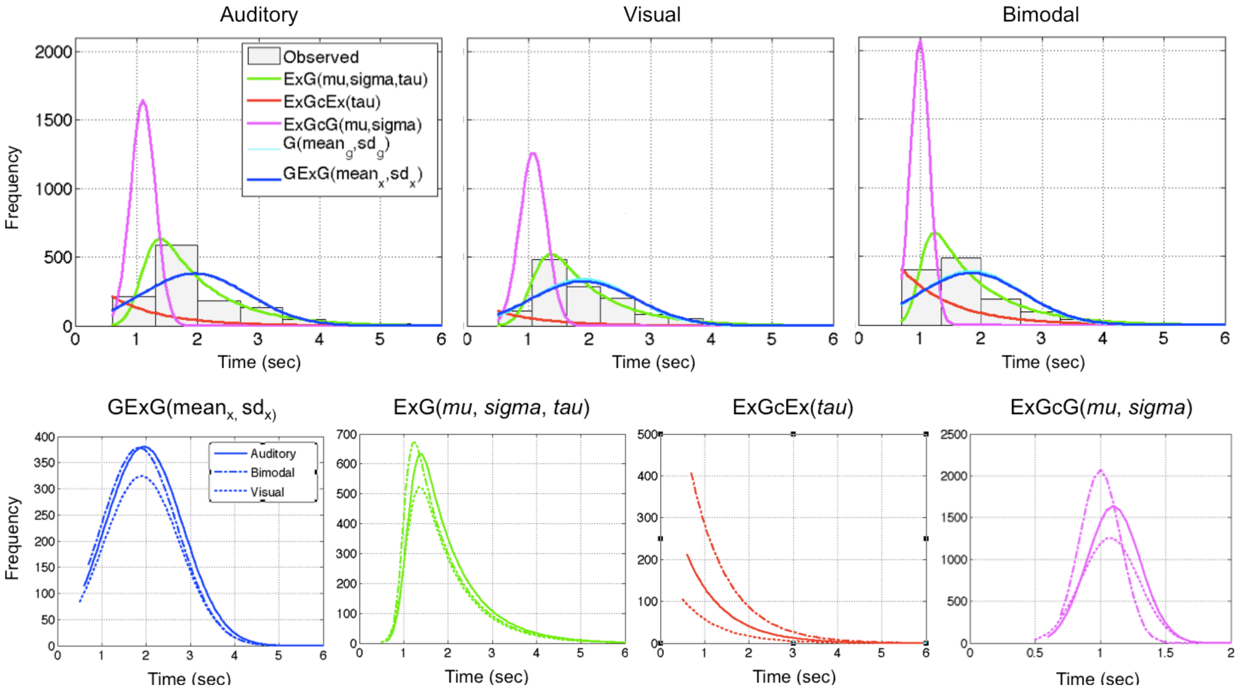


Figure 3: Distribution parameters (mu, tau and sigma), mean (meanx) and standard deviation (sdx) for the three modalities of presentation of the display. Meang ($\bar{x}$) and sdg (s) represent the parameters of the Gaussian distribution.

To assess a potential effect of practice or fatigue, the percentage of correct responses was computed as a function of the block order. There was no systematic variation in the rate of correct responses as a function of blocks order, suggesting a minimal impact of practice or fatigue (HVLA: X_5^2=5.77, *p*=.32; LVLA: X_5^2=5.51, *p*=.35; HVHA: X_5^2=12.74, *p*=.02; LVHA: X_5^2=10.24, *p*=.06).

3.2. Decision Time: Analysis of Distribution parameters

The data for one participant were disregarded based on extreme variance. Repeated measure ANOVAs of the ex-Gaussian distribution parameters (Figure 3 and Table 2 for $\bar{x}$) revealed that the B distribution was a compromise between the A and the V distributions.

Table 2: Percentage of correct responses and left/right decision times ($\bar{x}$ (sec) as a function of the environmental conditions.

	Auditory		Visual		Bimodal		Total
	% Correct	LRDT	% Correct	LRDT	% Correct	LRDT	% Correct
HVLA	86.7%	1.83 (.45)	86.4%	1.96 (.66)	86.7%	1.72 (.50)	86.6%
LVLA	90.6%	2.00 (.64)	93.1%	2.05 (.95)	92.2%	2.02 (.85)	91.9%
LVHA	88.6%	1.92 (.58)	90.6%	1.79 (.64)	93.6%	1.73 (.59)	90.9%
LVHA	83.6%	2.14 (.73)	93.6%	1.89 (.45)	92.2%	1.88 (.51)	89.8%
Total	87.36%	1.97 (.60)	90.89%	1.92 (.68)	91.17%	1.84 (.62)	89.7%

The overall mean LRDTs were statistically equivalent in the two unimodal conditions (Modality: $F_{(2,87)}$=3.74, *p*=.02; A, V: $t_{(46)}$=-.91, *p*=.36) and significantly shorter in the B than in either unimodal condition (A, B: $t_{(46)}$=2.62, *p*=.01; B, V: $t_{(46)}$=-2.04, *p*=.04). The effect of visual environment was not significant ($F_{(1,44)}$=.30, *p*=.82).

The effect of the block presentation order was not significant (HVLA: $F_{(5,1079)}$=1.66, *p*=.14; LVLA: $F_{(5,1079)}$=1.18, *p*=.31; HVHA: $F_{(5,1079)}$=1.02, *p*=.40; $F_{(5,1079)}$=1.94, *p*=.08).

In the B condition, both the *mu* and the *tau* values were statistically equivalent to their counterpart in the best unimodal condition, A for *mu* (*p*=.55) and V for *tau* (*p*=.17). For *sigma*, no significant difference between modalities was observed.

3.3. Percent Correct localization

To provide an estimate of localization accuracy (i.e., localization response and target location are spatially congruent), incorrect localization responses were disregarded (10.2%).

Table 3: Correct localization (%) as a function of display modality for the 2 levels of visibility and ambiguity.

	Visibility				Ambiguity				Total
	High	Low	X_1^2	*p*	Low	High	X_1^2	*p*	
Auditory	98.8%	97.7%	2.71	.09	98.7%	97.9%	1.52	.21	98.3%
Visual	98.1%	96.6%	3.35	.06	98.4%	96.4%	6.24	.01	97.4%
Bimodal	99.7%	98.3%	7.21	.007	99.3%	98.7%	1.15	.28	99.0%
Total	98.9%	97.6%	11.41	.0007	98.8%	97.7%	8.48	.003	

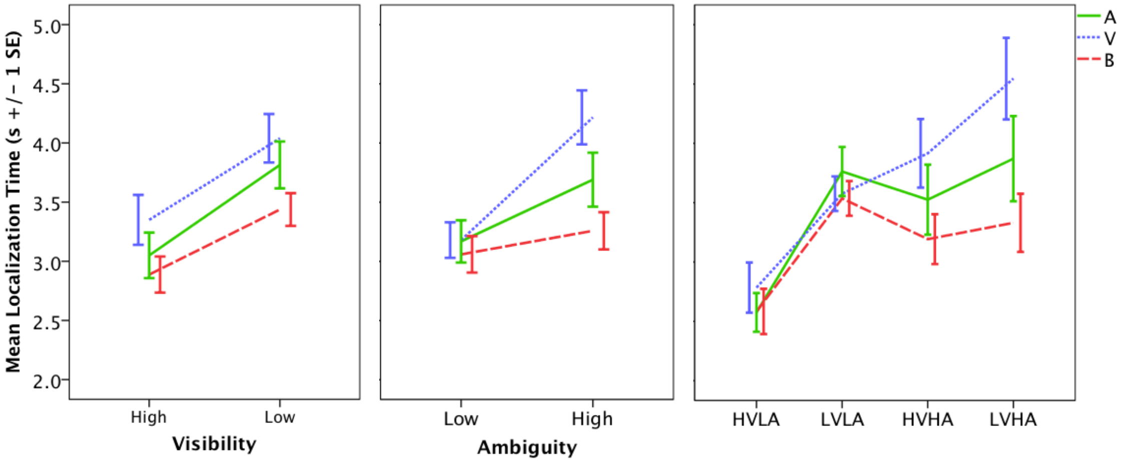


Figure 4: Localization Time: Interactions between Modality, Visibility and Ambiguity.

The threshold for mislocalization response accuracy (i.e., an entity other than the target was localized) was equivalent to the overall RMSE + 1 SD (2.13º+6.27º=8.4º). Overall, performance in the V condition was significantly lower than in the A and B conditions (A, V: X_1^2=7.73, p=.005; V, B: X_1^2=10.79, p=.001), as seen in Table 3. There was no difference between the A and the B conditions (X_1^2=.25, p=.61).

The LV level was associated with a decrease in correct localization rate for all modalities (marginally significant for A and V). Meanwhile, the HA level only impacted performance in the V condition, leading to a greater rate of incorrect responses.

3.4. Localization Time

Mean LTs were shorter in the B condition than in either of the two unimodal conditions (paired t tests: B, A: $t_{(46)}$=3.90, p<.0001; B, V: $t_{(46)}$=6.13, p<.0001). Comparison of the two unimodal conditions showed that localization times were shorter in the A than in the V condition (A, V: $t_{(46)}$=-3.36, p<.0001). The LV condition contributed to a significant increase in LT for the three NavAid conditions ($F_{(1,45}$=7.38, p=.009), as seen in Figure 4, left. Conversely, the HA condition impacted only the V condition ($F_{(1,45}$=7.46, p=.009) (Figure 4, middle). In the V condition, the LTs were the longest when LV and HA levels were combined (Figure 4, right), suggesting some form of additivity, although the difference was significant only with the LV condition (LVLA, LVHA: $t_{(12)}$=-.87, p=.03).

3.5. Analysis of Redundant Signal Effect

To test for the RSE, we determined the unimodal condition associated with the faster RTs for each subject and for each environmental condition.

For 28 out of 48 participants (58.3%) the mean LRDTs turned out to be faster in the V than in the A condition. In 50% of the cases, LRDTs in the B condition were significantly faster than in the best unimodal condition ($F_{(1,20)}$=29.98, p<.0001) by an average of 6.5% (≈11ms advantage). There was no effect of visibility ($F_{(1,20)}$=.03, p=.84) or ambiguity ($F_{(1,20)}$=.04, p=.34) and all the interactions proved insignificant. Conversely, the auditory LTs were significantly faster for 62.5% of the participants. Ambiguity level significantly affected this distribution (X_1^2=5.68, p=.01; A faster: LA=45.8%, HA: 63.3%). In the most degraded visual condition (LVHA), auditory LTs were significantly faster in 100% of the cases. LTs in the B condition were significantly faster than in the best unimodal condition in 60.4% of the cases ($F_{(1,25)}$=21.17, p<.0001) by on average 11% (≈38ms advantage).

4. STUDY 2: DUAL-TASK PARADIGM

The Dual Task Study was designed to evaluate the effect of increased visual workload more compatible with real situational conditions that astronauts may experience during tele-operations or surface operations. The monitoring and control of four consumable meters in a head up display (HUD) configuration was introduced as a Dual Task, i.e., a task of equivalent priority to the Single Orientation Task study.

4.1. Dual Task Method

The experimental protocol was the same as in Study 1 except for the control and monitoring of four meters representing the levels of EVA mission consumables (carbon dioxide, oxygen, water, and battery) superimposed on the visual scene at the top left of the display (Figure 5). The meters had randomly occurring depletion rates that led eventually to a critical state, changing from green to red after crossing a marked threshold. When the meter(s) turned red, the participants were instructed to touch the corresponding meter(s) on a touch screen display to bring it back to a nominal state. During a block, a failure always occurred a minimum of ~5 sec prior to each orientation target. In addition to these "regular" failures, periodically additional "dummy" meter failures were inserted in the scenario at random times during the localization phase to lessen the predictability of the meter failures. All meters reset to nominal values at end of a localization trial.

To approximate terrain motion effects, camera spatial disturbances were modeled by offsetting camera pitch and roll by two 5-sine Sum-of-Sines (SOS) functions. The pitch SOS used a fundamental frequency of 0.5 Hz with a maximum disturbance of 5 degrees, while the roll SOS used 1 Hz and 2.5 degrees. Both SOSs used the 1st, 3rd, 5th, 7th, and 11th harmonics scaled by the reciprocal of the harmonic number (1/1, 1/3, 1/5, etc.). All sine phases were randomized at the beginning of each trial. To date, preliminary data from 6 participants has been collected under non-degraded visual conditions. Additional participants and experimental conditions will be tested under normal (HVLA) and degraded (LVLA) visual environments.

4.2. Dual Task Data Analysis and Results

The preliminary data (6 of 12 subjects) were compared to their counterpart in Study 1, i.e., the non-degraded HVLA condition. Analyses were performed only for LRDTs, since the effect of dual task wasn't systematically investigated during the localization task.

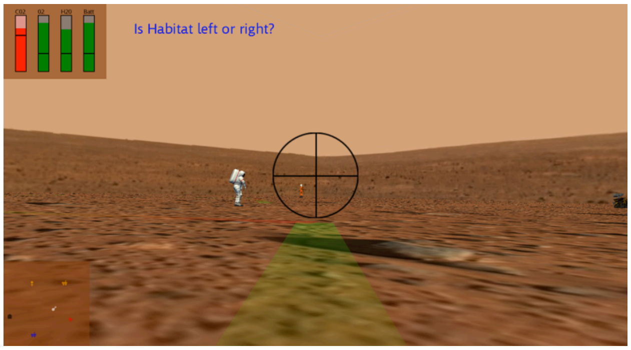


Figure 5: Dual Task condition. The consumables states are displayed on the top left of the screen. Red color-coding is associated with a critical level (green is nominal).

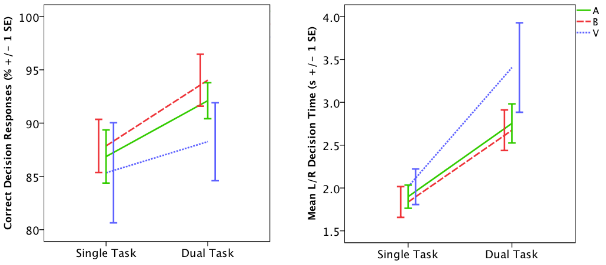


Figure 6: Left: Percentage of correct orientation responses. Right: Mean LRDTs as a function of the display modality in Single and Dual Task conditions.

It can be seen from Table 4 that the percentage of correct responses in the DT condition was lower in the V than in either the A or B conditions, although this difference didn't reach significance (A, V: X^2_1=2.70, p=.1; V, B: X^2_1=2.70, p=.1; A, B: X^2_1=0).

For LRDTs, we observed a significant effect of the Task (between subject effect: $F_{(1,16)}$=10.51, p=.005), a significant effect of Modality (within subject effect: $F_{(2,32)}$=6.78, p=.004), as well as a marginally significant interaction between Modality and Task ($F_{(2,32)}$=2.79, p=.07). As seen in Figure 6, the mean LRDTs in the V condition were much more sensitive to the effect of the dual task than in either the A or B conditions.

Table 4: Percentage of correct orientation responses and left/right decision times ($\bar{x}$ + s, in seconds) in Study 1 (single task, ST) and Study 2 (dual task, DT).

	Auditory		Visual		Bimodal		Total
	% Correct	LRDT	% Correct	LRDT	% Correct	LRDT	% Correct
ST	86.7%	1.89 (.86)	86.4%	1.80 (.92)	86.7%	1.99 (.94)	86.6%
DT	93.3%	2.75 (1.30)	88.3%	3.39 (1.69)	93.3%	2.69 (1.18)	91.7%

5. DISCUSSION

The present studies were designed to assess the benefits and usability of different types of displays in an orientation/ localization task in a simulated exploration environment under different levels of degradation of the visual environment. In particular, we investigated how the spatial dimension of sound could be used to provide veridical reproduction of auditory percepts over space and time, carrying not only the "What," but also the "Where" of information. To this end, five auditory earcons (sounds that reflect the meaning of the event) were used to create an auditory virtual scene matching the visual information from the simulation (operator's FOV) and from the 2D visual map. Two modes for the use of (spatialized) sound were tested: *substitution*, in which one modality replaces another (A alone) and *complementarity*, which is the case where congruent inputs from the different sensory channels are combined (i.e., A + V).

The auditory NavAid proved to be a very reliable source of information for both orientation and localization. The inherent inferiority of audition in terms of spatial resolution (less accurate in the front than vision) was offset by faster access to the information, as evidenced by higher percentages of correct orientation and faster decision times than in the V condition (no need for mental transformation).

Taken together, the experiment demonstrates that the ex-Gaussian convolution analysis provides a good description of the LRDT distributions and that the parameters of the convolution analysis behave differentially as a function of the NavAid modality. For the decision times, *mu* was lower in the A than in the V condition, while *tau* was lower in the V than in the A condition. In the B condition, the mean *mu* and *tau* values were statistically identical to the best unimodal estimate, i.e., A for *mu* and V for *tau*, utilizing the best of both worlds.

For localization, the audio locator was undoubtedly useful, in particular when the low visibility and high ambiguity conditions were combined, leading to both shorter localization times and greater accuracy. A major concern in the development of multimodal virtual interfaces is the reference frame in which the spatial information is displayed (egocentric for audition, allocentric for the 2D map, with operator's heading reported in a North-Up coordinate system). The use of different reference frames prevented a complete three-dimensional pairing for visual cues outside the FOV. For these reasons, we expected that the performance with a bimodal display would not reveal a real integrative process but rather reflect some form of statistical facilitation. In general, both decision time and localization time benefited from a combined presentation of the information.

In the Dual Task condition, the preliminary results suggest that the performance with displays utilizing auditory cues (both the A and B conditions) was less affected by the extra demands of additional visual workload than the visual-only display. At this point of the analysis, there is no significant difference between the performance in the B condition and the best unimodal condition, A. In future work, additional conditions will test the combined effect of a degraded visual environment and high visual workload on performance.

There is a need to investigate further how 2D visual maps (or even 3D) should be configured to minimize the effect of mental transformation. An egocentric representation of the scene would provide a case where the A and the V spatial information use the same frame of reference, potentially leading to a greater multisensory performance enhancement in the B condition. Other display characteristics, such as animated icons on the 2D map synchronized with the auditory display, could be used to increase the likelihood of enhanced multisensory gain.

Continued work is planned that addresses the utility of 3D audio in a multi-tasking context as outlined here. Other future work should explore the likely advantages of auditory displays for more complex scenarios such as those in which entities in the environment may be in motion, entities may be occluded, much longer and more complicated distances must be navigated, or where communication latency is a factor. Future EVA display designs should also include caution, warning, and emergency cueing for off-nominal situations (e.g., injured astronaut, loss of signal/communications).

6. ACKNOWLEDGEMENTS

Work supported by the Space Human Factors Engineering Project within NASA's Human Research Program at NASA Ames Research Center.

7. REFERENCES

[1] B. D. Simpson, D. S. Brungart, R. C. Dallman, R. J. Yasky, G. D. Romigh, and J. F. Raquet, "In-flight navigation using head-coupled and aircraft-coupled spatial audio cues," *Proc. Hum. Fact. Ergon. Soc.,* Baltimore, MD, 2007, pp.1341-1344.

[2] B. D. Simpson, D. S. Brungart, R. C. Dallman, R. J. Yasky, and G. D. Romigh, "Flying by ear: Blind flight with a music-based artificial horizon," *Proc. Hum. Fact. Ergon. Soc.,* 2008, pp. 6-10.

[3] S. J. Van Wijngaarden, A. W. Bronkhorst, and L. C. Boer, "Auditory evacuation beacons," *J. Aud. Eng. Soc.,* vol. 53, pp. 44-53, 2005.

[4] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "GPS-based navigation systems for the visually impaired," In W. Barfield and T. Caudell, (Eds.), *Fundamentals of Wearable Computers and Augmented Reality*, (pp. 429-446). Mahwah, NJ: Lawrence Erlbaum, 2001.

[5] T. V. Tran, T. Letowski, and K. S. Abouchacra, "Evaluation of acoustic beacon characteristics for navigation tasks," *Ergonomics,* vol. 43, pp. 807-827, 2000.

[6] B. N. Walker and J. Lindsay, "Using virtual reality to prototype auditory navigation displays," *Assist. Tech. J.* vol. 17(1), pp. 72-81, 2005.

[7] Walker, B. N. and J. Lindsay, "Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice." *Hum. Fact.,* vol. 48, pp. 265-278, 2006.

[8] E. M. Wenzel and M. Godfroy, "Spatial auditory displays to enhance situational awareness during remote exploration," *4th IEEE Int. Conf. Space Miss. Challenges Info. Tech.,* Palo Alto, CA, August 2011.

[9] M. Godfroy and E. M. Wenzel, "Human dimensions in multimodal wearable virtual simulators for extra vehicular activities," *Proc. NATO Workshop Hum. Dimens. Embedded Virt. Sims,* Orlando, FL, October, 2009.

[10] E. M. Wenzel, M. Godfroy, and J. D. Miller, "Prototype spatial auditory display for remote planetary exploration," *Proc. Aud. Eng. Soc.,* San Francisco, CA, October 2012.

[11] C. F. Altmann, S. Getzmann, and J. Lewald, "Allocentric or craniocentric representation of acoustic space: An electrotomography study using mismatch negativity," PLoS ONE 7(7): e41872, doi:10.1371/journal.pone.0041872, 2012.

[12] A.W. Mills, "On the minimum audible angle," *J. Acoust. Soc. Amer.,* vol. 30, pp. 237-246, 1958.

[13] J. Blauert, "*Spatial Hearing."* Cambridge, MA: MIT Press (Original edition *Räumliches Hören.* Stuttgart, Germany: S. Hirzel Verlag, 1974), 1983.

[14] P. Buser and M. Imbert, "*Vision. Neurophysiologie fonctionnelle IV.*" Paris, Hermann, 1987.

[15] M. A. Meredith and B. E. Stein, "Spatial determinants of multisensory integration in cat superior colliculus neurons," *J Neurophys.*, vol. 75, pp. 1843–1857, 1986.

[16] R. Ratcliff, "Group reaction time distributions and an analysis of distribution statistics," *Psych. Bull.*, vol. 86, pp. 446-461, 1979.

[17] D. A. Balota and D. H. Spieler, "Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency," *J. Exp. Psych.: Gen.*, vol. 128(1), p. 32, 1999.

[18] F. Schmiedek, K. Oberauer, O. Wilhelm, H. M. Süß, and W. W. Wittmann, "Individual differences in components of reaction time distributions and their relations to working memory and intelligence," *J. Exp. Psych.: Gen.*, vol. 136(3), p. 414, 2007.

[19] W. J. McGill, "Stochastic latency mechanisms," In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology.* (Vol. 1, pp. 309-360). New York: Wiley, 1963.

[20] L. M. Reder, H. Park, and P. D. Kiefabber, "Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding," *Psych. Bull.*, vol. 135(1), pp. 23–49, 2009.

[21] E. M. Wenzel, J. D. Miller, and J. S. Abel, "Sound Lab: A real-time, software-based system for the study of spatial hearing," *Proc. Aud. Eng. Soc.,* Paris, France, preprint 5140, February 2000.