Audio Decision Support for Supervisory Control of Unmanned Vehicles

Literature Review

C.E. NEHME
M.L. CUMMINGS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY*

PREPARED FOR CHARLES RIVER ANALYTICS

HAL2006-06

JULY, 2006

http://halab.mit.edu e-mail: halab@mit.edu

*MIT Department of Aeronautics and Astronautics, Cambridge, MA 02139

Purpose of this literature review:

To survey scholarly articles, books and other sources (dissertations, conference proceedings) relevant to the use of the audio channel in a multimodal station for the supervisory control of unmanned vehicles.

Background:

The visual channel is one of high bandwidth [2]. However any attempt to extract or analyze more than rudimentary information from the field of view causes the visual field to perform as a limited capacity processor restricted to a small portion of the field of view at any one instant. It is therefore in these instances, when one is required to perform multiple supervisory tasks such as target tracking and recognition, that auditory support could be most useful.

One of the primary functions of the hearing system is the regulation of gaze (Heffner 1997; Heffner & Heffner 1992). Because the auditory field of regard extends beyond the visual field, it is possible to detect sound-producing objects outside the visual field and bring them into foveal vision by a series of head and eye movements. This natural ability suggests various applications to the unmanned vehicle (UV) operator, including visual search, and threat localization. Visual search is a paradigm in which a person is asked to find one item (a target) from a set of other items (distractors).

A term that is used often in the literature is that of "virtual audio cues". This term refers to the phenomenon of directing a human's field of view by providing an audio stimulus or cue. The ability to reproduce virtual audio cues in displays is afforded by advances in digital signal processing. True binaural spatial audio (3D audio), when presented over headphones, appears to come from a particular point in the space outside of the listener's head. This is different from ordinary 2D audio, which is generally restricted to a line between the ears when listened to with headphones. Sound localization tasks presented using this method resulted in localization accuracy for listeners that is reasonably similar to their performance in the free field (Wightman & Kistler 1989). A free field is an environment in which there are no reflections. This is unlikely to exist in most workstations. It is through headsets that we are able to simulate free field listening, and therefore eliminate the unrealistic need to create a relatively echo-free listening field. The literature will often compare performance of target localization in the free-field versus that using virtual spatial audio. The difference between the two is that in free-field audio, sound (possibly through speakers) originates from the same location as the target to be acquired whereas in the case of virtual spatial audio, the sound is simulated so as to appear as if it is coming from the location of the target in question.

Finally, the human's aural system is able to attend to sound emanating from one spatial location while masking other audio emanating from other locations. This phenomenon could prove useful to the UV operator that is required to deal with multiple aural channels aside from audio cues such as communications links between the UV operator and other team members.

Several defense laboratories have cultivated major research programs in the areas of sound localization and spatial audio displays, which include:

- National Aeronautics and Space Administration (NASA – United States)

- Air Force Research Laboratory (AFRL – United States)
- TNO Technische Menskunde (Netherlands)
- Defense Science & Technology Organization (DSTO - Australia)
- Defense Science & Technology Laboratory (DSTL - United Kingdom)
- Defense Research and Development Canada (DRDC - Canada)
- Advisory Group for Aerospace Research and Development (AGARD - France)

This diverse collection of laboratories has investigated the use of spatial audio technology for several applications, the most relevant of which are:

- sound localization (Martin, McAnally, & Senova 2001; Whightman & Kistler 1989)
- the use of spatial audio technology for the detection and localization of visual targets (Bolia, D'angelo & Mckinley 1999; Brokhorst, Veltman & van Breda 1996; Flanagan, McAnally, Martin, Meehan & Oldfield 1998; Nelson et al. 1998)
- the use of spatial audio technology for collision avoidance (Begault & Pittman 1996)
- the use of spatial audio technology for navigation (Moroney et al. 1999)

Although most of these applications have focused on manned cockpits, other work has addressed the appropriateness of the audio channel for support of operators of unmanned vehicles such as the Airborne Warning & Control System (Bolia 2003a, Bolia 2003b). The use of spatial audio technology to increase speech intelligibility as well as decrease communications workload for operators engaged in multichannel listening tasks has been the focus of such research.

Literature Review:

This section contains a summary of those papers deemed most relevant to audio considerations in supervisory control of unmanned vehicles. See Figure 1 (at the end) to see the relationships between the most prominent papers that were reviewed. The primary resources used in this literature review include:

- International community for auditory display (ICAD)
- International Journal of Aviation Psychology Journal
- Human Factors Journal
- International Journal of Occupational Safety and Ergonomics
- Air Force Research Laboratory studies
- International Symposium on Aviation Psychology Proceedings
- Journal of the Audio Engineering Society
- Proceedings of the Annual Human Factors and Ergonomic Society Symposiums

Citation: [1]	Bolia, R.S., D'Angelo,W.R., and McKinley, R.L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors, 41(4),* 664-669.
Summary:	This paper describes the experimental protocol and results of an experiment that was carried out in order to evaluate the effectiveness of spatial audio displays on target acquisition performance. Participants performed a visual search task with and without the aid of a spatial audio display. The independent variables included: - the number of visual distractors present - the spatial audio condition (no spatial audio, free-field spatial audio, virtual spatial audio) Results indicated that both free-field and virtual audio cues produced a significant decrease in search times without a corresponding decrease in the percentage of correct responses. Potential applications of this research include the design of spatial audio displays for UV operators that under high workload circumstances can benefit from reduced search times.

Citation: [2]	Cisneros, J., D'Angelo,W.R., McKinley, R.L., and Perrott, D.R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors, 38,* 702-715.
Summary:	The paper examined the minimum latency required to locate and identify a visual target (visual search) in a two-alternative forced-choice paradigm in which the visual target could appear from any azimuth (0 degree to 360 degrees) and from a broad range of elevations (from 90 degrees above to 70 degrees below the horizon) relative to a person's initial line of gaze. Seven people were tested in six conditions: unaided search, three aurally aided search conditions, and two visually aided search conditions. The unaided search condition involved spatially uncorrelated sound in which the auditory stimulus cued the participant that a visual target was present but provided no spatial information as to where to find it. The aurally aided conditions were: - Spatially correlated actual sound in which the auditory stimulus cued the participant to look in a certain direction (real sound) - Simulated 3D sound using a 3D virtual sound display - Simulated 2D sound where only information regarding the azimuth of the source is provided The visually aided conditions were: - a video monitor located in the center of the visual field was used to stimulate a 2D radar display indicating the relative location of the visual target - same as above condition except that the monitor was located to the participant's right and well below the initial field of view Aurally aided search with both actual and virtual sound localization cues proved to be superior to unaided and visually guided search.

Citation: [3]	Bronkhorst AW, Veltman JA, van Breda L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors, 38(1),* 23-33.
Summary:	The effectiveness of a three-dimensional auditory display in conveying directional information was investigated. In all conditions, the participants viewed an outside cockpit computer-generated image on which they could spot the target only when it was in close range and a three dimensional tactical display indicating the target position. Additional displays were a bird's-eye-view radar display and a three dimensional auditory display which generated a warning sound from the relative direction of the target. The independent variables were: - the type of 3D display used to indicate the target's position (no display, an auditory display, a visual display, or both) - and the azimuth and elevation of the position where the target emerged The dependent variables were: - search time - subjective workload Results showed that the radar and auditory displays caused about the same significant reduction in search time in comparison with the tactical display only. A further reduction was found when the two additional displays were presented simultaneously.

Citation: [4]	Flanagan, P., McAnally, K., Martin, R., Meehan, J. and Oldfield, S. (1998). Aurally & visually guided visual search in a virtual environment. *Human Factors, 40(3),* 461-468.
Summary:	This paper investigated the time it took subjects to perform a visual search task for targets outside the visual field using a helmet-mounted display. The experiment was run under six different conditions: - nonspatial audio - nonspatial audio with a visual aid - transient spatial audio - transient spatial audio with a visual aid - updating spatial audio - updating spatial audio with a visual aid The visual aid consisted of a dynamic arrow that indicated the direction and angular distance from the instantaneous head position to the target. The transient spatial display was a triplet of noise bursts (50 ms) recorded at the target azimuth and elevation. The updating auditory display was a train of single noise bursts at a rate of approximately 8/s. Both visual and auditory spatial cues reduced search time dramatically compared with unaided search. The updating audio cue was more effective than the transient audio cue and was equally as effective as the visual cue in reducing search time. These data show that both spatial auditory and visual cues can markedly

	improve visual search performance.
Citation: [5]	Parker, Smith, Stephan, Martin, and McAnally (2004). Effect of supplementing head-down displays with 3-D audio during visual target acquisition. *International Journal of aviation psychology, 14(3),* 277-295.
Summary:	This study investigated the effectiveness of supplementing head-down displays (HDDs) with high-fidelity three–dimensional audio using a flight simulation task in which participants were required to visually acquire the image of a target aircraft. There were 3 conditions: a visual HDD providing azimuth information combined with a nonspatial audio cue, a visual HDD providing azimuth and elevation information combined with a nonspatial audio cue, and a visual HDD providing azimuth information only combined with a 3–D audio cue which provided elevation info as well as azimuth info. The independent variable was the type of display (within), and the dependent variables were the: - localization error - mean percentage of front-back confusions (occur when true and perceived locations of the sound source are in different front-back hemifields) - visual acquisition time - the frequency of display scanning by the subject - NASA-TLX subjective workload measure - Subjective measure of situational awareness When 3–D audio was presented, the visual acquisition time was faster, perceived workload was reduced, and perceived situational awareness was improved. This performance improvement was attributed to the fact that participants were often able to perform the task without the need to refer to the HDD.

Citation: [6]	Tannen, Nelson, Bolia, Warm and Dember (2004). Evaluating adaptive multisensory displays for target localization in a flight task. *International Journal of Aviation Psychology, 14(3),* 297-312.
Summary:	This study was designed to determine the effectiveness of providing target location information via head-coupled visual and spatial audio displays presented in adaptive and nonadaptive configurations. In the adaptive configurations, presentation of information was dependent on the location of the target in relation to the orientation of the pilot's head, whereas in the non-adaptive configurations, the information was always present. For example, In the adaptive auditory + nonadaptive visual display condition, the use of spatial audio cueing was determined by target locations in excess of ±15° from the center of the pilot's head orientation, whereas the visual cueing display was always present. Independent variables: Seven interface conditions were combined factorially with two target types (ground and air) and two initial target location conditions (within and beyond ±15° of the participant's

	instantaneous head orientation) in a completely within-subjects design. Dependent variables: - designation accuracy - designation time - subjective workload The integration of visual displays with spatial audio cueing enhanced performance efficiency, especially when targets were most difficult to detect. Several of the interface conditions were also associated with lower ratings of perceived mental workload. The benefits associated with multisensory cueing were equivalent in both adaptive and nonadaptive configurations.

Citation: [7]	Bolia, R.S. and McKinley, R.L. (2000). The effects of hearing protectors on auditory localization: Evidence from audio-visual target acquisition. *International Journal of Occupational Safety and Ergonomics*, 6, 309-319.
Summary:	Response times (RT) in an audio-visual target acquisition task were collected from 3 participants while wearing either circumaural earmuffs, foam earplugs, or no hearing protection. Analyses revealed that participants took significantly longer to locate and identify an audio-visual target in both hearing protector conditions than they did in the unoccluded condition, suggesting a disturbance of the cues used by listeners to localize sounds in space.

Citation: [8]	Nelson WT, Hettinger LJ, Cunningham JA, Brickman BJ, Haas MW, McKinley RL. (1998). Effects of localized auditory information on visual performance using a helmet-mounted display. *U.S. Air Force Research Laboratory, Wright-Patterson Air Force Base, 40 (3)*, 452-460.
Summary:	An experiment was conducted to evaluate the effects of localized auditory information on visual target detection performance. Visual targets were presented on either a wide field-of-view dome display or a helmet-mounted display and were accompanied by either localized, nonlocalized, or no auditory information. The addition of localized auditory information resulted in significant increases in target detection performance and significant reductions in workload ratings as compared with conditions in which auditory information was either nonlocalized or absent. Qualitative and quantitative analyses of participants' head motions revealed that the addition of localized auditory information resulted in extremely efficient and consistent search strategies. Implications for the development and design of multisensory virtual environments are discussed. Actual or potential applications of this research include the use of spatial auditory displays to augment visual information presented in helmet-mounted displays, thereby leading to increases in performance efficiency, reductions in physical and mental workload, and enhanced spatial awareness of

	objects in the environment.
Citation: [9]	Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors, 33(4),* 389-400.
Summary:	Visual search performance was examined under two different conditions. The visual targets were either presented concurrently with a sound located at the same position as the visual target or presented in silence. Both the number of distractor visual figures and the distinctness of the target relative to the distractors were considered. Under all conditions, visual search latencies were reduced when spatially correlated sounds were present. Aurally guided search was particularly enhanced when the visual target was located in the peripheral regions of the central visual field and when a larger number of distractor images were present. These results indicate that spatially correlated sounds may have considerable utility in high information environments.

Citation: [10]	Bolia, R. S.,& Nelson,W. T. (2003). Spatial audio displays for target acquisition and speech communication. In L. J. Hettinger &M.W. Haas (Eds.), *Virtual and adaptive environments: Applications, implications, and human performance issues,* 187–197. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Summary:	This book chapter presents some basic research in the areas of: - spatial audio for target acquisition - spatial audio for speech communications It also discusses the form that such displays may take as well as some implementation issues. The potential for interaction among these two areas is also addressed along with the possibility of dynamically switching between the two interface types according to the state of the operator, of the environment or both. Also, the role of these displays as components of multisensory adaptive interfaces is discussed.

Citation: [11]	Begault, D. R. (1993). Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors, 35*, 707–717.
Summary:	This paper describes a preliminary experiment designed to measure and compare the acquisition time for capturing visual targets under two auditory conditions: - standard one-earpiece presentation - two-earpiece three dimensional audio presentation Twelve commercial airline crews were tested under full mission simulation conditions at the NASA-Ames Man-Vehicle Systems Research Facility

	advanced concepts flight simulator. Scenario software generated visual targets corresponding to aircraft that activate a traffic collision system aural advisory. The spatial auditory position was linked to the visual position with 3D audio presentation. Results showed that crew members using a 3D auditory display acquired targets approximately 2.2 s faster than did crew members who used one ear-piece headset, but there was no significant difference in the number of targets acquired.

Citation: [12]	Begault, D. R., & Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors, 35,* 361–376.
Summary:	In this study, 11 inexperienced subjects judged the apparent spatial location of headphone-presented speech stimuli filtered with non-individualized head-related transfer functions (HRTFs). About half of the subjects leaned towards either the median or the lateral vertical planes when judging the spatial location of the stimuli, and estimated were almost always elevated. Individual differences were pronounced for the distance judgments; 15% to 46% of stimuli were heard inside the head, with the shortest estimates near the median plane. The results suggest that most listeners can obtain useful azimuth information from speech stimuli filtered by nonindividualized HRTFs. Measurements of localization error and reversal rates are comparable with a previous study that used broadband noise stimuli.

Citation: [13]	Begault, D. R., & Wenzel, E. M. (1992). Techniques and applications for binaural sound manipulation in human-machine interfaces. *International Journal of Aviation Psychology, **2,*** 1-22.
Summary:	In this paper, the implementation of binaural sound to speech and auditory sound cues is addressed from both applications and technical standpoints. Techniques overviewed include processing by means of filtering with (HRTFs). Application to advanced cockpit-human interface systems is discussed, although the techniques are extendible to any human-machine interface. Research issues pertaining to three-dimensional sound displays under investigation at the Aerospace Human Factors Research Division at NASA-Ames Research Center are described.

Citation: [14]	Doll, T. J., Gerth, J. M., Engelman, W. R., & Folds, D. J. (1986). Development of simulated directional audio for cockpit applications (Tech. Rep. No. AAMRL-TR-86-014). *Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.*
Summary:	This project included three major activities: (1) an extensive review and synthesis of the research literature on auditory localization, (2) the design,

	fabrication, and evaluation of an apparatus for demonstrating simulated auditory localization, and (3) experimental research to determine characteristics of the audio signal, in the time and frequency domains, which enhance localization performance with simulated cues. Previous research is reviewed which describes the cues involved in the perception of sound-source direction, both horizontally and vertically, when the head is stationary. Also reviewed is research on auditory distance perception, the roles of head movement and vision in auditory localization, the perception of auditory motion and volume, and the effects of noise on auditory localization. A feedback control model is presented, which integrates evidence derived from four different theoretical positions concerning the effects of head movement and vision on auditory localization. Possible applications of simulated auditory localization technology in aircraft cockpits are outlined, and the potential benefits of such applications are discussed.

Citation: [15]	D. R. Begault, M. R. Anderson, & B. U. McClain (2003). Spatially-modulated auditory alerts. *Paper presented at the Proceedings of the 2003 International Conference on Auditory Display.*
Summary:	This paper describes a technique for improving the detection of auditory alerts in a noisy environment. The technique involves spatially modulating (encoding) an auditory alert so as to ease its detection in such environments. The 70.7% absolute detection threshold for the spatially jittered alert was on average 7.8 dB lower compared an alert that is not spatially jittered, with noise and signal both presented over headphones using virtual simulation techniques. With the addition of supra-aural headphones to partially attenuate loudspeaker background noise, the threshold for the spatially-jittered alert was 13.4 dB lower than a nonjittered alert.

Citation: [16]	Begault & Pittman (1996). Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. *International Journal of Aviation Psychology, 6(1),* 79-93.
Summary:	The advantage of a head-up auditory display for situational awareness was evaluated in an experiment designed to measure and compare the acquisition time for capturing visual targets under two conditions: - standard headdown Traffic Alert and Collision Avoidance System (TCAS) display - and three-dimensional (3-D) audio TCAS presentation. Results showed a significant difference in target acquisition time between the two conditions, favoring the 3-D audio Traffic Alert and Collision Avoidance System condition by 500 ms.

Citation: [17]	Begault, D. R.,Wenzel, E. M., & Lathrop,W. B. (1997). Augmented TCAS advisories using a 3–D audio guidance system. *Proceedings of the Ninth International Symposium on Aviation Psychology*, 353–357.
Summary:	Scenario software generated 49 targets corresponding to aircraft along a SFO-LAX route; each crew ran this route twice, once with and once without the addition of a 3-D audio cue to process the TCAS advisory. The audio cue consisted of a verbal alert for the TCAS advisory which was presented through a stereo headset so that the perceived lateral spatial location of the sound corresponded to the location of the conflicting traffic. Across all ten crew, the results showed a significant difference in target acquisition time between the two conditions, favoring the 3-D audio TCAS condition by 207 msec; there was no significant difference in the number of targets that were acquired. Questionnaire data showed a favorable attitude from most crews towards augmenting the current TCAS interface with a 3-D audio system.

Citation: [18]	Oving, A. P., & Bronkhorst, A. W. (1999). Application of a three-dimensional auditory display for TCAS warnings. In *Proceedings of the Tenth International Symposium on Aviation Psychology*, 26–31.
Summary:	In this study, 3D sound and verbal location information were used in the TCAS warning to convey spatial information about an intruding aircraft. The verbal information involved such cues as up-left and down-right to the warning. The participants were required to visually search the external world scene in order to indicate the specific orientation of the intruding aircraft. The results of the experiment indicated that the response time was reduced with both 3D sound and verbal location information (12% and 11% respectively). The results also indicated additional benefits in 3D sound that were not available with verbal cues. First, the standard deviation of the response time was reduced considerably, indicating a more consistent search performance for 3D sound. Second, the localization of targets close to the line-of-sight was improved. The best performance was observed when 3D sound and verbal location cues were combined (22% reduction).

Citation: [19]	Veltman, Oving & Bronkhorst (2004). Effectiveness of 3-D audio for warnings in the cockpit. *International Journal of Aviation Psychology, 14(3),* 257-276.
Summary:	This paper presents the results of an experiment conducted in order to investigate the application of 3-dimensional audio for presenting auditory warnings in the cockpit in two simulator experiments. Pilots had to respond

	to warnings of the Traffic Alert and Collision Avoidance System and to system malfunction warnings. Visual warning displays were always present. The first experiment showed 12% faster response times when 3–D audio was used compared to mono sound. This was only observed when flight path error information was presented on a visual head-down display. No effect was observed when an auditory error display was used because this enabled pilots to pay more attention to the visual warning displays. In the second experiment, only a visual error display was employed, and the effects of 3–D audio and verbal directional information (e.g., "up") were tested. Each type of cue resulted in a 12% reduction of response times. The combination of 3–D audio and verbal cues resulted in the best performance, with response time reductions of approximately 23%.

Citation: [20]	Moroney, B.W., Nelson, W.T., Hettinger, L.J., Warm, J.S., Dember, W.N., Stoffregen, T.A., & Haas, M.W. (1999). An evaluation of unisensory and multisensory adaptive flight path navigation displays: An initial investigation. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, 71-75.
Summary:	In this paper, unisensory and multisensory adaptive interfaces for precision aircraft navigation were tested under varying concurrent task demands. Participants, composed of 12 USAF pilots, were required to perform a simulated terrain following, terrain-avoidance navigation task, including evasive maneuvers, while also performing: (1) no additional task; (2) a visual search task; or (3) an auditory monitoring task. Real-time performance efficiency, as measured by lateral deviation from the flight course, was used to activate the adaptive navigation displays consisting of a visual azimuth steering line on the head-up display, a spatial auditory beacon, or a combination of the two displays. A completely factorial, within-subjects design was used to assess the effects of secondary task loading and adaptive interface configurations on flight performance. The results indicated that the effectiveness of multisensory, adaptive navigation displays depends not only upon the supplementary task confronting the pilots, but also upon the type of flight task performed and the strategies they adopted to acquire and use the information offered.

Citation: [21]	Nelson W. T., Bolia R.S., Ericson M. A., and McKinley R.L. (1999). Spatial audio displays for speech communications: A comparison of free field and virtual acoustic environments. *Proceedings of the Human Factors and Ergonomics society 43rd annual meeting*, 1202-1205.
Summary:	The ability of listeners to detect, identify, and monitor multiple simultaneous speech signals was measured in free field and virtual acoustic environments. Factorial combinations of four variables, including audio

	condition, spatial condition, the number of speech signals, and the sex of the talker were employed using a within-subjects design. Participants were required to detect the presentation of a critical speech signal among a background of non-signal speech events. Results indicated that spatial separation increased the percentage of correctly identified critical speech signals as the number of competing messages increased.

Citation: [22]	Veltman, Oving & Bronkhorst (2004). 3-D audio in the fighter cockpit improves task performance. *International Journal of Aviation Psychology, 14(3),* 239-256.
Summary:	In this study, a flight simulator experiment was used to study the effect of 3–D audio on the visual scan behavior of pilots. The effect of 3–D audio on mental workload were also studied, by keeping record of several physiological measurements that are related to mental workload as well as subjective effort ratings. Also in this paper, the effects on task performance and workload when two auditory sources are presented concurrently were studied by using two different auditory sources to present information for two different tasks. In some of the experimental conditions, these tasks had to be performed concurrently making both auditory sources relevant for task execution at the same time. The performance on several tasks was improved when 3-D audio was present, whereas no negative performance effects were found. Physiological measures were not affected indicating that mental effort was the same in all conditions. Pilots were also able to process the information from 2-independent 3D auditory displays that were present at the same time.

Citation: [23]	Bronkhorst, A.W. (1995). Localization of real and virtual sound sources. *Journal of the Acoustical Society of America, 98,* 2542–2553.
Summary:	Localization of real and virtual sound sources was studied using two tasks. In the first task, subjects had to turn their head while the sound was continuously on and press a button when they thought they faced the source. In the second task, the source only produced a short sound and the subjects had to indicate, by pressing one of eight buttons, in which quadrant the source was located, and whether it was located above or below the horizontal plane. Virtual sound sources were created using HRTFs, measured with probe microphones placed in the ear canals of the subjects. Sound stimuli were harmonic signals with a fundamental frequency of 250 Hz and an upper frequency ranging from 4 to 15 kHz. Results, obtained from eight subjects, show that localization performance for real and virtual sources was similar in both tasks, provided that the stimuli did not contain frequencies above 7 kHz. When frequencies up to 15 kHz were included, performance for virtual sources was, in general,

	poorer than for real sources. Differences between results for real and virtual sources were relatively small in the first task, provided that individualized HRTFs were used to create the virtual sources, but quite large (a factor of 2) in the second task. The differences were probably caused by a distortion of high-frequency spectral cues in the HRTFs, introduced by the probe microphone measurement in the ear canal.

Citation: [24]	Haas, E. C. (1998). Can 3–D auditory warnings enhance helicopter cockpit safety? *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 1117–1121.
Summary:	A study was conducted to determine how quickly helicopter pilots could respond to helicopter malfunction warning signals in a simulated cockpit environment when four different signal functions (fire in left engine, fire in right engine, chips in transmission, shaft-driven compressor failure) were presented in three different presentation modes (visual only, visual plus 3-D auditory speech signals, visual plus 3-D auditory icons). The dependent variable was pilot response time to the warning signal, from the time of signal onset to the time that the pilot manipulated the collective control in the correct manner. Subjects were 12 U.S. Army pilots between the ages of 18 and 35 who possessed hearing and visual acuity within thresholds acceptable to the U.S. Army. Results indicated that signal presentation was the only significant effect. Signal function x signal function interaction and the signal presentation x signal function interaction were not significant. Post hoc test results indicated that pilot response time to the visual signals supplemented with 3-D audio speech or auditory icon signals was significantly shorter than that to visual signals only.

Citation: [25]	Martin, R. L., McAnally, K. I.,& Senova, M. A. (2001). Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society*, *49,* 14–22.
Summary:	An evaluation was conducted by comparing, for three participants, virtual and free-field localization performance across a wide range of sound-source locations. For each participant, virtual localization was found to be as good as free-field localization, as measured by both front-back confusion rate and average localization error. The results provide a demonstration of the feasibility of achieving free-field equivalent localization of virtual audio.

Citation: [26]	Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1998). Monitoring the simultaneous presentation of multiple spatialized speech signals in the free field. In D.W. Martin (Ed.), *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America*, 2341–2342.
Summary:	In this paper, an experiment is described where the effect of spatial auditory information on a listener's ability to detect, identify, and monitor multiple simultaneous speech signals was evaluated in the free field. Five spatialization conditions (front right quadrant (RQ), front hemifield (FH), right hemifield (RH), full 360° (F), and a non-spatialized control (C)) were combined factorially with eight talker conditions (1,2,3,4,5,6,7, and 8 talkers) and the sex of the critical speech signal (male and female) to provide 80 experimental conditions. This was a within-subjects design. Participants were required to detect the presentation of a critical speech signal among a background of non-signal speech events. Results indicated that the spatialization of simultaneous speech signals (1) increased the percentage of correctly identified critical signals and (2) lowered ratings of perceived mental workload as compared to a non-spatialized control condition.

Citation: [27]	Perrot, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1995). Aurally aided detection and identification of visual targets. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting,* 104–108.
Summary:	The experiments described in this report provide baseline performance measures of aurally directed detection and search for visual targets in an observer's immediate space. While the simple target detection task was restricted to the frontal hemi-field, visual search performance (discrimination of which of two light arrays was present on a given trial) was evaluated for both the frontal and rear hemi-fields. In both tasks, the capacity to process information from the visual channel was improved substantially (1—50% reduction in latency) when spatial information from the auditory modality was provided concurrently. While performance gains were greatest for events in the rear hemi-field and in the peripheral regions of the frontal hemi-field, significant effects were also evident for events within the subject's center visual field.

Citation: [28]	Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America, 94,*111–123.
Summary:	In this study, inexperienced listeners judged the apparent direction

	(azimuth and elevation) of wideband noisebursts presented in the free-field or over headphones; headphone stimuli were synthesized using HRTFs. When confusions (up-down and front-back) were resolved (the incorrect responses were coded as if they were correct), because it was unknown how to treat these types of errors fairly, localization of virtual sources was quite accurate and comparable to the free-field sources for 12 of the 16 subjects. Of the remaining subjects, 2 showed poor elevation accuracy in both stimulus conditions, and 2 showed degraded elevation accuracy with virtual sources. Many of the listeners also showed high rates of front-back and up-down confusions that increased significantly for virtual sources compared to the free-field stimuli.

Citation: [29]	Wightman F.L, Kistler D.J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America, 85(2),* , 858-867
Summary:	This paper conducted an experiment where participants reported the apparent spatial positions of wideband noise bursts that were presented either by loudspeakers in free field or by headphones. The headphone stimuli were localized by eight subjects in virtually the same positions as the corresponding free-field stimuli. For each subject in each of the six regions of auditory space there was a very close correspondence between judgments of the apparent positions of real sources in free field, and judgments of the apparent positions of digitally synthesized virtual sources presented under headphones. The data also indicated that there are some relatively subtle aspects of free-field listening, which are not captured in headphone simulation. For example, there is a substantial increase in the number of front-back confusions in the headphone condition. It is also important to note that the differences among the subjects and the differences between the free-field data and the headphone data appear almost entirely in the elevation components of the responses.

Meta Analysis:

In this meta-analysis, we will revisit the general themes of the papers and discuss their implications for spatial audio support in supervisory control of unmanned vehicles.

- **Spatial Audio and Target Acquisition ([2], [3], [4], [5], [7], [8], [9], [10], [11], [17], [18], [27])**

  A large number of the papers in the literature review focused on using spatial audio to aid target acquisition. These studies showed that spatial audio, when used in conjunction with the already present visual information, can reduce search times and reaction times for target acquisition in the visual field. The implication of this is that UV operators are sometimes required to perform target acquisition by monitoring the video being transmitted from the on-board payload. Although

the field of view (width of the computer screen or screens) presented to UV operators is generally much narrower than that presented to pilots in a cockpit, the benefits of spatial audio cues could still be realizable. This is due to the fact that even searches for targets that were located near the center of the visual field improved with the addition of spatial audio.

However, augmenting the primary visual information source with an additional visual display (such as the radar visual display) was equally as effective for conveying 3D information as augmenting the primary source with a spatial auditory display. However when augmenting the primary information source such as a camera display simultaneously with displays for both the visual and auditory channels, there was further reduction in search time. However, because UV visual interfaces have limited real estate, it might be advantageous to augment the primary visual source with a spatial audio display over a visual display wherever supplementing both at the same time is not feasible.

The research also showed us that supplementing a visual display with verbal cues achieved similar benefits to those achieved by adding spatial audio, but had some limitations Therefore in environments were headsets for UV operators are not possible, verbal cues can be used as auditory aids to the visual source of information, but this performance would generally not be as high as if headsets were used.

Whenever questionnaires were conducted in the studies, they led to statistically significant results indicating that participants liked 3D audio support. For ground operators of UVs, it is important that they approve of any new technology given to them, and therefore a similar study would need to be run with UV controllers to see their acceptance of audio support.

- **3D Audio Displays to Enhance Alerts and Warning Detection ([15], [16], [18], [19], [24])**

  Unlike fighter cockpits where targets are constantly moving back and forth in the field of view, targets in UV interfaces are, in most cases, moving at a slower pace especially in surveillance type missions, where targets are monitored from a distance. Examining the papers concentrating on spatial audio to support alert and warning detection revealed that alerts that are embedded with spatial audio information generated faster response times, therefore supporting alert and warning detection. Again in all these studies, the primary source of information, the visual warning display, was always present.

  These results encourage the use of spatial audio to alert the UV controller to look to a particular location on the interface in order to read information about the warning. This is definitely important since it is very easy for controllers to miss warnings that are on screens other than the ones they are currently working on. Change blindness is a known problem in human-computer interaction that relies primarily on the visual field, thus using spatial audio to improve health and status monitoring as well as specific mission/task contextual alerts. Also by further encoding the alerts to as to indicate the level of urgency, one could allow the operators to prioritize alerts.

Other experiments, however, show that the application of verbal location information to the warning in a monoaural display can result in comparable performance improvements as for 3D audio, and hence can serve as an alternative to spatial audio. This is important for the supervisory control of unmanned vehicles because it gives us an alternative to using 3D headphones when they are not a feasible option as mentioned earlier.

Sanders & McCormick (1993) presents design recommendations for the selection or design of warning and alarm signals. Those that pertain to headset design are listed below:

- Use frequencies between 200 and 5000 Hz, and preferably between 500 and 3000 Hz, because the ear is most sensitive to this middle range.
- Use a modulated signal (1 to 8 beeps per second, or warbling sounds varying from 1 to 3 times per second) since it is different enough from normal sounds to demand attention.
- Use signals with frequencies different from those that dominate any background noise, to minimize masking.
- If different warning signals are used to represent different conditions requiring different responses, each should be discriminable from the others, and moderate-intensity signals should be used.

These considerations should be taken into account when designing a spatial audio interface to support alert and warning detection.

- **Workload ([5], [22])**

Results show that pilots can perform control and mission tasks more efficiently when they are supported by 3D audio. In UV control, operators are often required to multitask, and the support of 3D audio could allow for more efficient completion of those tasks. The spatial audio system must not require more attention, or else the workload might increase. Two different results can be expected when a spatial audio system is provided to the controllers: (a) the overall workload will decrease or (b) the controller will redistribute his or her workload by paying more attention to other tasks. For complex human supervisory control tasks such as UV operation, it is expected that with any effective aural aids, UV operators would refer less to the visual information due to the aural presentation. Because the operators would be able to redistribute their workload (or add additional tasks), the overall workload would remain the same but the overall level of performance would increase.

- **Real vs. Virtual Sound Sources ([1], [2], [23], [25], [28], [29])**

We can conclude that localization accuracy using virtual 3D audio is almost the same as the accuracy achievable when using real sound sources. Certain differences between the two types of sources were highlighted such as the fact that when the same stimuli were presented through virtual sources as opposed to real sources, the vertical localization error increased.

Due to the remoteness of the control station from the vehicle and mission environment, real audio cues are difficult to provide. Virtual sounds created with individualized HRTFs (HTRFs that were obtained from the controller him/her

self) appear to be the most effective. Such HRTFs can be localized almost as accurately as real sources provided that head movements can be made and that the sound is sufficiently long.

Also, if virtual sources of sound were used in UV control stations, frequencies above 7 kHz should be avoided as they are not simulated correctly at such frequencies. Also, when orienting their head toward a sound source, listeners give priority to the left/right cues first, and consider up/ down cues at a later stage. It is therefore important to try and separate information laterally as opposed to vertically wherever possible on the different screens used by the operator.

- **Speech Communications ([10], [12], [13], [21], [26])**

  This section is important because if 3D headsets are used for warnings, then speech must also be conveyed from co-located or remotely located team members. Otherwise, the 3D headsets would marginalize two-way communication with the operator(s) wearing them.

  We can interpret from the papers on speech communications that for both (a) radio transmissions and (b) speech, binaural sound manipulations can improve the intelligibility of speech sources against noise and assist in the segregation of multiple sound sources. Spatial separation of speech signals in the horizontal plane enhances one's ability to identify critical speech signals when they occur in competing message environments as would be the case in ground control stations. In the case of UV control stations, for speech originating in the control station (i.e. from other co-located controllers), sound spatialization can also be used to organize locations in perceptual auditory space and to convey urgency or establish redundancy. Headphones could be used within a binaural system that includes an intercom from the other controllers spatialized to the proper side. This addresses the reservations UV operators have in wearing headphones that would otherwise isolate them from other crew members.

  Also due to binaural summation of loudness, a binaurally equipped UV operator would need less amplitude of the signal at each ear, giving an additional advantage in suppressing hearing fatigue.

  The results of these studies can also be extended to the case where 3D headsets are not used and instead the controllers communicate directly with each other. In that case, the results encourage that we look more closely at the placement of the controllers if we wish them to communicate with each other efficiently.

- **Adaptive Interfaces ([6], [20])**

  The research on adaptive interfaces revealed that when displaying redundant information on both the visual and aural channel, an operator could suppress one of the modalities during certain mission phases. For UV control stations, an adaptive interface could be designed such that the aural display could be turned off when the operator is not under stressful phases of the mission and does not need redundant information. During critical mission phases, the auditory interface could be turned on so as to relieve the operator from having to look at the corresponding visual display and hence redistribute workload to those other tasks that also need attention. This dynamic capability of being able to control the

presence or not of aural information could insure that aural information is only presented when it is necessary and suppressed when it does not add any additional value.

Potential Areas for Future Research:

As a result of this literature review, we feel that the following areas have been well-researched and thus there would be little added to the scientific body of knowledge:

- Spatial audio for threat localization
- Spatial audio for speech communications

However, this literature review also uncovered many areas that have not either been addressed, or addressed in-depth which include:

- Verbal cueing vs. spatial audio for enhancing alerts and warnings
- Simultaneous presentation of auditory warning and alerts
- Priority queueing of audio warnings and alerts
- How to spatially associate audio warnings and alerts to specific displays in a multi-screen UV interface
- How to present different types of auditory cues, e.g., visual threat localization versus warnings and alerts, through the same headset so that the UV controller can differentiate between the different alert types
- Identifying the mission phases in UV supervisory control where auditory information could be suppressed so as to take advantage of auditory adaptive interfaces
- The use of ambient noise headphones and associated filtering techniques for which no significant literature has been found.

Additional References:

Bolia, R.S. (2003a). Effects of Spatial Intercoms and Active Noise Reduction Headsets on Speech Intelligibility in an AWACS Environment. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, 100-103.

Bolia, R.S. (2003b). Spatial Intercoms for Air Battle Managers: Does Visually Cueing Talker Location Improve Speech Intelligibility? *Proceedings of the 12th International Symposium on Aviation Psychology*, 136-139.

Haas, E. C. (1998). Can 3–D auditory warnings enhance helicopter cockpit safety? In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 1117–1121.

Heffner, R. S. (1997). Comparative Study of Sound Localization and its Anatomical Correlates in Mammals. *Acta Otolaryngologica Supplement, 532*, 46-53.

Heffner, R. S., & Heffner, H. E. (1992). Visual Factors in sound localization in mammals. *Journal of Comparative Neurology, 317*, 219-232.

Sanders, M.S., & McCormick, E.J. (1993). Human Factors in Engineering and Design. *Seventh edition, McGraw-Hill.*

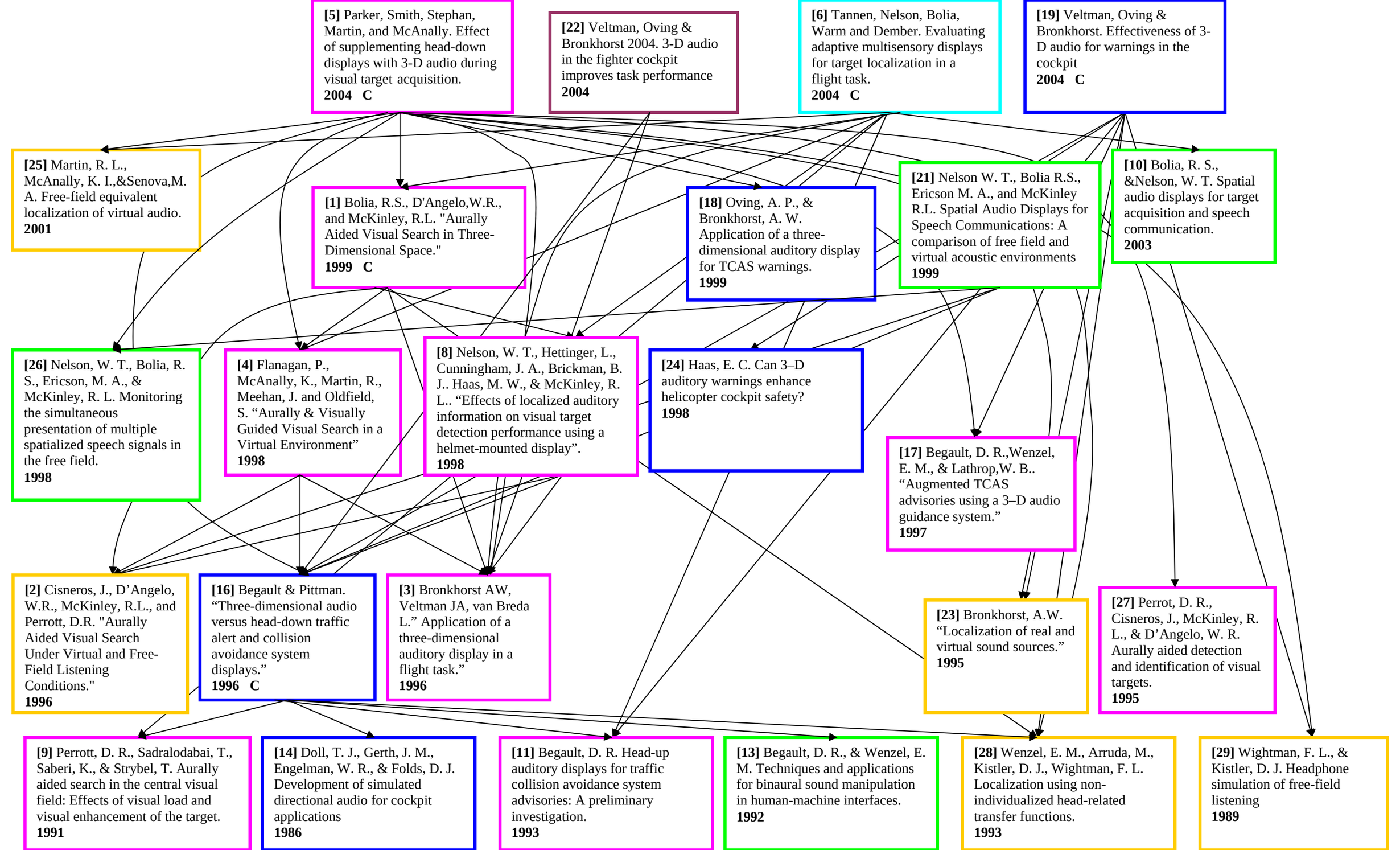
Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening I: stimulus synthesis. *Journal of the Acoustical Society of America, 85,* 858-867.

Figure 1:

Figure 1, shown below, describes the relationships between the most prominent papers that were reviewed. An arrow pointing from one paper to another implies that the first paper references the second. This is also implied by the vertical location of the papers on the page. Papers at the top reference others that are located lower on the page. Although the abundance of arrows makes quickly perceiving the relationships between papers hard, the graph quickly highlights which papers have an extensive reference list, and also signals to the reader how often each paper has been cited. At a second tier, the graph allows the reader to follow the arrows and make connections between papers. In addition, the coloring scheme allows the reader to quickly tell which area of research each paper belongs. The papers are grouped by color. The different groupings are:

Purple: Spatial Audio and Target Acquisition
Blue: 3D Audio Displays to Enhance Alerts and Warning Detection
Burgundy: Workload
Orange: Real vs. Virtual Sound Sources
Green: Speech Communications
Turquoise: Adaptive Interfaces

For each paper in the figure, the corresponding reference list was studied, and references that were related to any of the five categories above were extracted. This list is not complete because not every paper referenced is listed here and there are likely other papers that relate, but were not included due to the focus on human supervisory control in unmanned vehicle applications.


[5] Parker, Smith, Stephan, Martin, and McAnally. Effect of supplementing head-down displays with 3-D audio during visual target acquisition. 2004 C
[22] Veltman, Oving & Bronkhorst 2004. 3-D audio in the fighter cockpit improves task performance 2004
[6] Tannen, Nelson, Bolia, Warm and Dember. Evaluating adaptive multisensory displays for target localization in a flight task. 2004 C
[19] Veltman, Oving & Bronkhorst. Effectiveness of 3-D audio for warnings in the cockpit 2004 C
[25] Martin, R. L., McAnally, K. I.,&Senova,M. A. Free-field equivalent localization of virtual audio. 2001
[1] Bolia, R.S., D'Angelo,W.R., and McKinley, R.L. "Aurally Aided Visual Search in Three-Dimensional Space." 1999 C
[18] Oving, A. P., & Bronkhorst, A. W. Application of a three-dimensional auditory display for TCAS warnings. 1999
[21] Nelson W. T., Bolia R.S., Ericson M. A., and McKinley R.L. Spatial Audio Displays for Speech Communications: A comparison of free field and virtual acoustic environments 1999
[10] Bolia, R. S., &Nelson, W. T. Spatial audio displays for target acquisition and speech communication. 2003
[26] Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. Monitoring the simultaneous presentation of multiple spatialized speech signals in the free field. 1998
[4] Flanagan, P., McAnally, K., Martin, R., Meehan, J. and Oldfield, S. “Aurally & Visually Guided Visual Search in a Virtual Environment” 1998
[8] Nelson, W. T., Hettinger, L., Cunningham, J. A., Brickman, B. J.. Haas, M. W., & McKinley, R. L.. “Effects of localized auditory information on visual target detection performance using a helmet-mounted display”. 1998
[24] Haas, E. C. Can 3–D auditory warnings enhance helicopter cockpit safety? 1998
[17] Begault, D. R.,Wenzel, E. M., & Lathrop,W. B.. “Augmented TCAS advisories using a 3–D audio guidance system.” 1997
[2] Cisneros, J., D’Angelo, W.R., McKinley, R.L., and Perrott, D.R. "Aurally Aided Visual Search Under Virtual and Free-Field Listening Conditions." 1996
[16] Begault & Pittman. “Three-dimensional audio versus head-down traffic alert and collision avoidance system displays.” 1996 C
[3] Bronkhorst AW, Veltman JA, van Breda L.” Application of a three-dimensional auditory display in a flight task.” 1996
[23] Bronkhorst, A.W. “Localization of real and virtual sound sources.” 1995
[27] Perrot, D. R., Cisneros, J., McKinley, R. L., & D’Angelo, W. R. Aurally aided detection and identification of visual targets. 1995
[9] Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Aurally aided search in the central visual field: Effects of visual load and visual enhancement of the target. 1991
[14] Doll, T. J., Gerth, J. M., Engelman, W. R., & Folds, D. J. Development of simulated directional audio for cockpit applications 1986
[11] Begault, D. R. Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. 1993
[13] Begault, D. R., & Wenzel, E. M. Techniques and applications for binaural sound manipulation in human-machine interfaces. 1992
[28] Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L. Localization using non-individualized head-related transfer functions. 1993
[29] Wightman, F. L., & Kistler, D. J. Headphone simulation of free-field listening 1989