APPLYING SPATIAL AUDIO TO HUMAN INTERFACES:
25 YEARS OF NASA EXPERIENCE

DURAND R. BEGAULT[1], ELIZABETH M. WENZEL[1],

MARTINE GODFROY[1,2], JOEL D. MILLER[1,3] AND MARK R. ANDERSON[1,3]

[1] *Human Systems Integration Division, NASA Ames Research Center Moffett Field, California, USA*
Durand.R.Begault@nasa.gov
Elizabeth.M.Wenzel@nasa.gov
Martine.Godfroy-1@nasa.gov
[2] *San José State University Foundation, San José, California, USA*
[3] *Dell Services-Perot Systems, Plano, Texas USA*

From the perspective of human factors engineering, the inclusion of spatial audio within a human-machine interface is advantageous from several perspectives. Demonstrated benefits include the ability to monitor multiple streams of speech and non-speech warning tones using a 'cocktail party' advantage, and for aurally-guided visual search. Other potential benefits include the spatial coordination and interaction of multimodal events, and evaluation of new communication technologies and alerting systems using virtual simulation. Many of these technologies were developed at NASA Ames Research Center, beginning in 1985. This paper reviews examples and describes the advantages of spatial sound in NASA-related technologies, including space operations, aeronautics, and search and rescue. The work has involved hardware and software development as well as basic and applied research.

INTRODUCTION

In 1983, work began at NASA Ames Research Center in developing what eventually became known as virtual environment or "virtual reality" systems. The pioneering research of Michael McGreevy and Stephen Ellis with interactive perspective displays led to McGreevy's relatively inexpensive VIVED system (Virtual Visual Environment Display) [1-3]. The seminal VIEW (Virtual Interface Environment Workstation) project developed by McGreevy, Scott Fisher, and Warren Robinette set the interactivity hardware elements used as the basis for such systems that are now well-recognized: (1) a stereo visual display of a computer-generated environment; (2) means for tracking user head and/or body movement to update the graphics, so as to provide a user the illusion of interactively moving through the environment; (3) interactivity with virtual objects via devices such as the "DataGlove" developed by Thomas Zimmerman and Jaron Lanier, where the hand position could be separately tracked and the degree of flex in the fingers separately calculated.

It was a natural extension to add virtual audio, or "3-D sound" to such virtual environments. The continuity of spatial position of visual objects in the virtual world accomplished by head tracking could be complemented by processing sound in a similar manner. The initial motivation was that sound would add to the sensation of "immersion" in the environment, since objects within the graphic world could potentially be audible in a spatially-coordinated way. It was also anticipated by Elizabeth Wenzel (co-author) that sound was also valuable for tracking objects out of the field of view, and for providing feedback for actions and events such as contact. Voice recognition and speech synthesis were also added to the interface to provide additional interactivity (Figure 1).
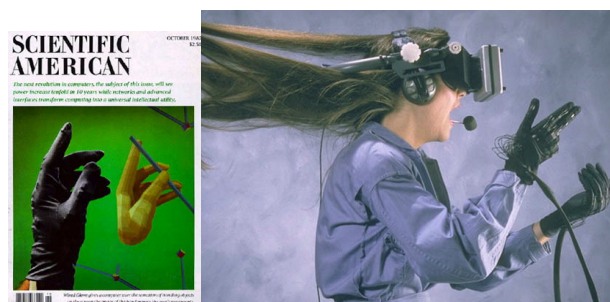

Figure 1: Virtual Reality, circa 1987: (left) Scientific American article on "future computing" featuring DataGlove; (right) NASA publicity shot for VIEW system showing stereo visual display, headphones for 3D audio, voice recognition interface, and Data Gloves for operating virtual menus.

Between 1986-1988, Wenzel collaborated with Scott Foster (Crystal River Engineering Inc.) to develop a prototype real-time virtual audio system known as the *Convolvotron,* so-named because of its convolution signal processing and as a salute to the futuristic device names found in publications such as *Popular Science* [4]. This was originally a 2-card set that fit into the bus of either an IBM PC-XT or PC-AT computer. One card contained a modified TMS320C25 system board and the other a "convolution engine" utilizing IDT7203 and IDT71M135 FIFOS and four IMSA100 DSPs for "double buffering." An electro-magnetic head tracking device (Polhemus *Isotrak*) was used as input to determine 6 degree-of-freedom displacement information in a virtual acoustic space (Figure 2).

Figure 2. Left: *Convolvotron* card set. Right: Polhemus *Isotrak* head tracking hardware, including transmitter and receiver

The novelty of the *Convolvotron* was that it represented one of the first large-scale digital convolution engines that could be updated quickly enough for interactive audio. It could dynamically spatialize up to four input sources at a 50 kHz sample rate using head-related transfer function (HRTF) FIR filtering. The HRTFs mapped 24 azimuth positions (every 15 degrees) on six elevation rings spaced by 18 degrees. Interpolated values were calculated and updated regularly based on positional data from either the head tracking hardware or via software interface. The number of coefficients used to represent left and right HRTFs reduced from 512 coefficients to 128 coefficients proportionally as the number of inputs increased. Hardware limitations motivated initial research into HRTF "simplification", whereby fewer coefficients could be used to represent the HRTF or provide sufficient localization cues [5].

Simultaneously with development of the *Convolvotron,* Wenzel formed the Spatial Auditory Display Laboratory at NASA Ames Research Center for the purpose of evaluating localization of "virtual acoustic stimuli" [6]. Between about 1986 and 1992, a collaborative research agreement was established with Frederic Wightman of the University of Wisconsin-Madison to provide head-related transfer functions and conduct collaborative psychoacoustic research related to the project. The laboratory also partially funded Dr. Wightman's two

seminal works on measurement and localization of HRTF-based stimuli [7,8].

This work and subsequent activity led to several important publications including an evaluation of the localization ability when "listening through another person's ears" [9]. Possible degradation of localization performance with "non-individualized HRTFs" has always been considered an important consideration for human interface design.

The emphasis on psychoacoustic research was in part motivated by the desire to have performance-based metrics that could be incorporated into virtual environment interfaces. Stephen Ellis and Bernard Adelstein pursued quantitative measurements of human performance as a function of system fidelity [10, 11]. The Spatial Auditory Display Laboratory was part of the Aerospace Human Factors Division, at the time one of the largest human factors research facilities in the world with emphases in both the space and aeronautic missions of NASA. The Spatial Auditory Display Laboratory still operates but has expanded its role to include research into multimodal perception, communications engineering, HRTF measurement, virtual acoustic software development, speech intelligibility testing and rapid prototyping of advanced acoustic interfaces. It is currently part of the Human Systems Integration Division, which "…advances human-centered design and operations of complex aerospace systems through analysis, experimentation, and modeling of human performance and human-automation interaction to make dramatic improvements in safety, efficiency, and mission success" [12].

1 IMPROVED INTELLIGIBLITY FOR MULTIPLE SPEECH STREAMS

As early as 1989, the concept of using HRTF processing specifically for speech communications was addressed at NASA [13]. As well-understood from the 'cocktail party effect' described by Cherry, the ability to conduct a single conversation within a context of multiple streams of speech is greatly enhanced with binaural compared to single-ear (monotic) listening [14]. Most communications workstations (e.g., aeronautics, emergency) are designed with a single earpiece or two channel monaural signal (diotic) for listening to multiple communication channels. At NASA, launch personnel at Kennedy Space Center are required to listen up to seven different communication channels at once over a single earpiece. The "exposed" ear is exposed to background noise, face-to-face conversation and/or telephone communication. Effectively, the binaural advantage used in everyday listening is lost.

The concept of spatially separating multiple speech communications was not new in itself. Amongst early examples, the second issue of the Journal of the Audio Engineering Society (1953) featured an article by John
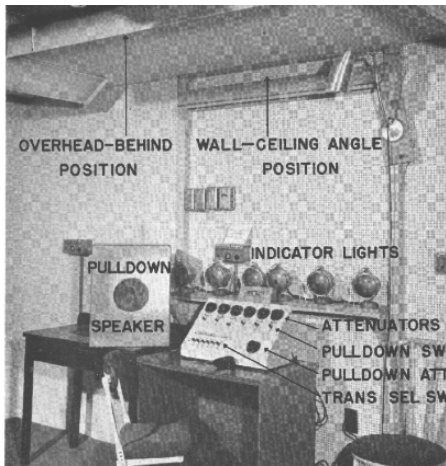
Figure 3: Synthetic control tower with multiple loudspeakers co-located with the indicator lights. In some conditions, the loudspeakers were mounted overhead.
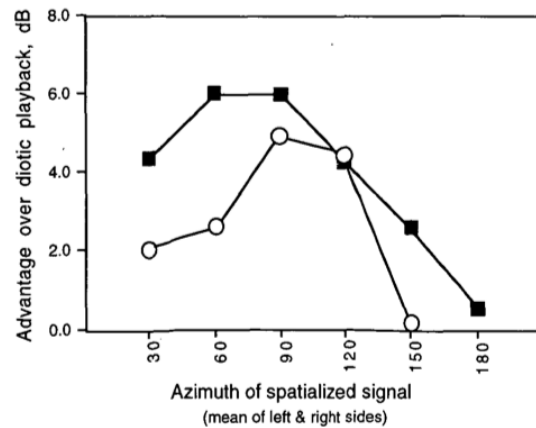

Figure 4: Intelligibility advantage of the binaural speech display for full-bandwidth (filled squares) and telephone bandwidth (open circles) speech, relative to diotic playback. From [17].

Webster and Paul Thompson of the Human Factors Division of the US Navy Electronics Laboratory describing a "synthetic control tower" where six communication streams were fed either to separate loudspeakers or to one loudspeaker (Figure 3) [15]. By means of a switch, a communication channel monitored from loudspeakers mounted overhead could be sent to a "pulldown" speaker.

The design of an HRTF-based speech communication processor was initially recognized as being simpler than development of a virtual environment convolution engine like the *Convolvotron*. Each speech communication stream only needed to be processed to a single unique virtual location, eliminating the need for dynamically updating or interpolating HRTF coefficients. The first prototype used four Digidesign Audio Media sound cards independent of a host PC, using a removable EPROM with HRTF data that was interfaced to the onboard Motorola 56000 DSP chip. Development of hardware prototypes with mixing capabilities and other features followed, with the first prototypes for NASA KSC launch personnel and for NASA JSC astronaut training facilities. A US patent was issued in 1995 [16].

Studies were conducted and published in the Journal of the AES on the advantage of using the display for full-bandwidth and telephone-bandwidth (4 kHz maximum frequency) [17]. Besides the immediately apparent increase in intelligibility, an additional benefit was observed: individual volume controls for each channel needed to be manipulated less often compared to monotic or diotic playback. This advantage is due to the well-known ability of humans to successfully 'stream' audio from a given source [18]. The concept has been successfully implemented into teleconferencing applications as well as radio communication systems for search and rescue [19, 20].

2 AURALLY-GUIDED VISUAL SEARCH

The current implementation of the Traffic Alert and Collision Avoidance System (TCAS II) uses both auditory and visual map displays of information to supply flight crews with real-time information about proximate aircraft. However, the visual display is the only component delegated to convey spatial information about surrounding aircraft, while the auditory component is used as a redundant warning or, in the most critical scenarios, for issuing instructions for evasive action.

Begault and colleagues evaluated the effectiveness of a 3-D head-up auditory TCAS display in several studies by measuring target acquisition time [21, 22]. The experiments used professional pilots in a fully-equipped, motion-based flight simulator. The direction of the auditory alert "Traffic-Traffic" was linked to a visual target location's azimuth (an incoming aircraft at 2 miles), but not its elevation. However, instead of literally mapping the location to the pilot's head position, as might be done in a virtual environment simulation, only a limited set of discrete positions were used for cueing direction (Figure 5). Despite this simplification, the results showed a significant reduction in visual acquisition time when using spatialized sound to guide head direction (0.5 to 2.2 seconds, depending on the experimental conditions).

An important observation from these studies was that the perceived spatial position could be referenced to the pilot's exocentric reference to their aircraft, independent of head position (with "0 degrees" corresponding to the nose of the plane, rather than to their own nose). This allowed two pilots in different relative positions to take advantage of the same spatial cues. Overall, the most important advantage of spatial audio is that it allows

pilots to keep their visual gaze "out the window" looking for traffic without needing to move the head downward to the visual display and then back up. Situational awareness is improved, and the visual perceptual modality is freed to concentrate on other tasks, as necessary.
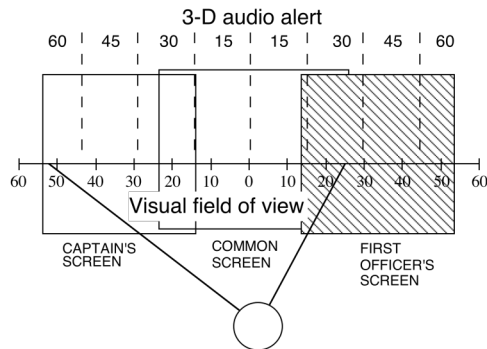

3-D audio alert

Figure 5. The horizontal field-of-view in the simulator from the perspective of the left seat (captain's position). The numbers within the dashed lines show the mapping between visual target azimuths and the specific azimuth position of the 3-D sound cue that was used for the TCAS alert.

3 MULTI-MODAL TARGET ACQUSITION

Determining the optimal means of integrating visual and auditory perceptual information is key for safe and effective use in advanced human interfaces. Such an "ecological" approach to the presentation of information reflects the fact that users are actively localizing within an environment using multiple sensory cues [23].

Martine Godfroy (co-author) joined the Spatial Auditory Displays Laboratory in 2005. She extended her work in the area of the multimodal perception of congruent but also non-congruent visual (V) and auditory (A) stimuli from work previously conducted at IMASSA (French Army Aeronautic Cognitive Sciences Laboratory, now renamed IRBA) [24].

Evidence for localization enhancement using for multimodal targets (possibly including haptic feedback) compared to unimodal targets is of great interest in the context of the development of advanced multimodal interfaces. In recent studies, synchronous bimodal auditory-visual (AV) spatial localization precision has been shown to exceed that of vision, which is superior to audition for unimodal localization. In 2004 Godfroy et al. provided evidence of bimodal AV integration in two dimensions in the frontal field (one degree of visual angle spot of light and pink noise sound burst, 100 msec duration, in free-field, for 35 positions tested within a 80 degree azimuth by 60 degree elevation) [25]. These results confirmed near optimal integration, with localization precision for the bimodal AV targets being

better than that of the more precise visual modality, while accuracy for the bimodal stimulus tending to be a compromise between the values of individual modalities in favor of vision.

An important question remained whether these results could be replicated using virtual sound sources. Studies were conducted in which visual stimuli in a spherical 30 by 30 degree frontal field were combined with headphone-based virtual acoustic sound sources (generic and individualized HRTFs) (Figure 6) [26, 27]. As with real sound sources, localization precision for the bimodal AV targets showed to be better than that of the more precise modality (vision), even in elevation where auditory localization has the greatest error.

These results have an important outcome in the context of the development of the advanced interfaces for aeronautic and space applications, where the major requirement is safe and efficient performance when developing displays that access multiple sensory modalities. One modality can replace another, as for example in the context of visual overload, where information can be presented via the auditory modality. Two modalities, presented simultaneously in space and time, provide a much more precise and robust estimate of the localization of a target, with the benefit increasing with target's eccentricity relative to the observer's central fixation. Other advantages of visual-auditory information compared to unimodal presentation include faster localization time and faster orientation toward a target [28].

Potential applications of these results are evident: when vision is not available, or degraded, the auditory channel provides a very ecological substitute, both for the content and the localization of the message. In a saturated visual environment, or to provide fast detection, localization, and/or identification of a target, multimodal presentation is the ideal solution.
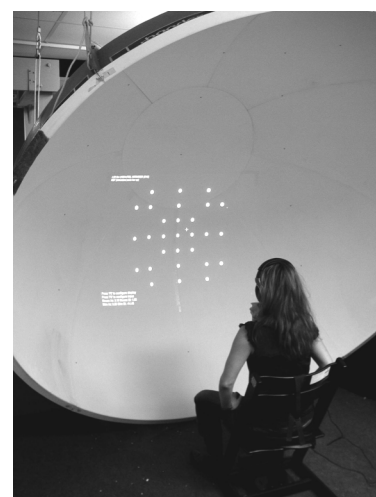

Figure 6. Localization study involving visual and auditory stimuli

4 SOUND LAB: A REAL-TIME, SOFTWARE-BASED SYSTEM FOR SPATIAL SOUND SYNTHESIS

In 1998, Wenzel, Miller (co-authors) and Abel began development of the software equivalent of the *Convolvotron*. *slab3d* (formerly known as *SLAB,* for Sound Lab) is a software-based, real-time virtual acoustic environment rendering system developed as a tool for the study of spatial hearing and for general purpose virtual acoustic rendering [29, 30]. The rendering engine was developed and continues to be maintained by Miller; it is used by research facilities such as the US Air Force Research Laboratory. Recent and future developments include sound mixing, audio-visual virtual environment prototyping, DIS (distributed interactive simulation) radio and VoIP (voice over internet protocol) implementation. *slab3d* is designed to take advantage of the low-cost personal computer platform while providing a flexible, maintainable, and extensible architecture to enable the quick development of experiments. The software provides an API (Application Programming Interface) for specifying the acoustic scene as well as an extensible architecture for exploring multiple rendering strategies. The *SLAB Render API* supports a number of parameters including sound source specification (waveform and signal generation), source gain, source location, source trajectory, listener position, listener HRTF (Head-Related Transfer Function) database, surface location, surface material type, render plug-in specification, scripting, and low-level signal processing parameters.
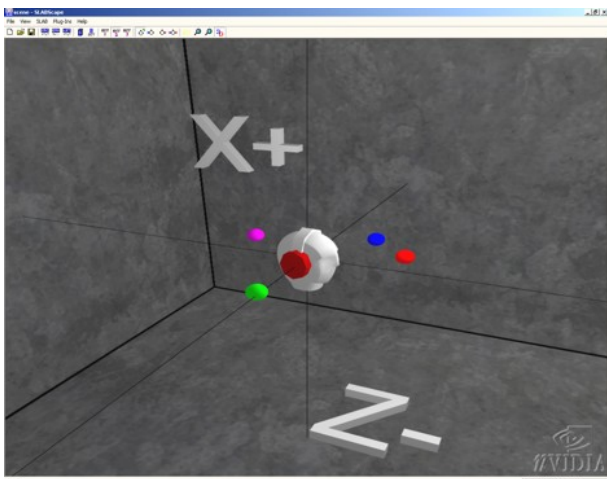

Figure 7. The *SLABScape* graphical user interface allows the user to experiment with the *SLAB Render API* and access and manipulate the acoustic scene parameters.

slab3d is available for free for non-commercial use to provide a research solution for a number of virtual acoustic environment applications that include aerospace display research, virtual reality for training, enhanced communications, and improved situational awareness [31]. For these applications and others, *slab3d* provides a low-cost system for dynamic synthesis of virtual audio over headphones without the need of special purpose signal processing hardware.

5 AUDITORY BEACONS FOR EXTRA-VEHICULAR ACTIVITIES

During extra-vehicular activity (EVA) sorties (short duration missions), astronauts must maintain situational awareness of a number of spatially distributed "targets" including other team members (both human and robotic), rover vehicles and other critical resources, and the lander/outpost or other safe havens. These targets are often outside the astronaut's immediate field of view (or are not visible from current location). Further, visual resources may be needed for other task demands.
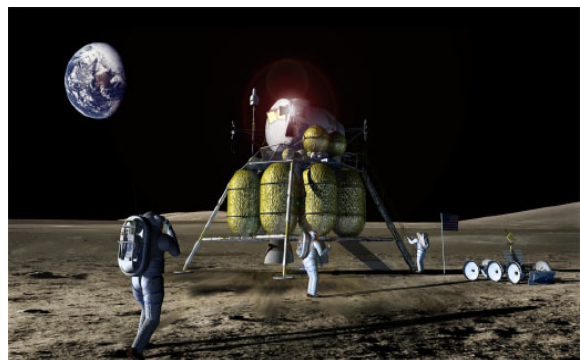

Figure 8. Artist's concept of a lunar sortie mission.

In 2008, initial development efforts in auditory displays for situational awareness resulted in a demonstration of an auditory "orientation beacon" display at NASA Ames. (A similar concept has been previously demonstrated for providing navigation assistance to blind persons; e.g., [32]). This auditory beacon display prototype created non-intrusive guidance sounds that enhance situational awareness without imposing undue distraction or workload. The auditory display prototype demonstrated an advanced communication concept involving the use of directional auditory cues that can be potentially used by crewmembers on a planetary surface to safely find their way to a habitat, rover, or other crewmembers. This "beacon" display could supplement visually displayed information or be a critical backup if visual systems fail. A demo of the beacon virtual acoustic environment simulates an augmented reality auditory display for an astronaut conducting EVA sorties on the moon (downloadable as part of the *slab3d* software; see [31]). Three auditory beacons assist the astronaut in locating a rover, the lander, and another astronaut ("partner"). Voice commands are used to interact with the display. To experience the complete demo, a head tracker is required to locate the position of the listener.

Current work focuses on development of a software test bed for experimental evaluation of the efficacy of a revised beacon display prototype, a Mars EVA audio-visual simulation of a spatial audio augmented-reality display. A joystick will be used in place of a head-tracker to position a listener as they steer a Mars Buggy. Future work will also include the development of caution, warning, and emergency cueing for off-nominal situations (e.g., injured astronaut, loss of signal/communications).

6 SPATIAL ALARMS

Design methodologies for insuring human ability to detect the presence of an alarm have for the most part been based on the amplitude spectrum of the alert and the signal-noise ratio of the alarm. International Standard ISO 7731 covers the formation of auditory alerts for danger signals and indicates that the level must be at least 10 dB above the background noise within at least one octave band between 300-3000 Hz [33]. It is well understood from the auditory literature that, by making spectral components of an alert substantially higher than the measured background noise level, one can insure the audibility or "detection" of such a signal. However, the technique of reducing the masking of an auditory alert by means of spatial manipulation of the signal is a novel concept.

Begault (co-author) described a technique using spatial modulation, whereby an alarm signal is moved laterally in space at a rate of 2-10 Hz in the manner of an insect moving about the head [34]. A study using a simulated flight deck environment indicated that spatial modulation of an alarm sound allows it to be about 7 dB more detectible against a stationary background noise, compared to when it is not moving [35].
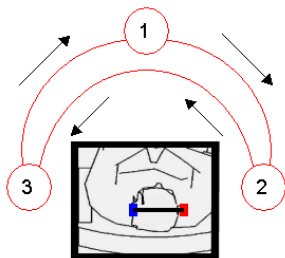

Figure 9. Spatial modulation of an alert (1.6 or 3.3. Hz period), from position 1 to position 2 to position 3, and back to position 1.

7 RAPID ACQUISITION OF HRTFS

The Spatial Auditory Display Laboratory personnel collectively have had considerable experience with a number of HRTF measurement systems. The system used in collaborative work with the University of Wisconsin in the late 1980s was very time-consuming

and required an anechoic chamber [7]. Later, an innovative solution called *Snapshot* was developed and implemented at the Spatial Auditory Display Laboratory, principally through the work of Jonathan Abel (then at Crystal River Engineering). This system used a single loudspeaker and was capable of measuring a rough grid of 30 degree azimuth intervals at six elevations in a reflective environment. The measurement technique used microphones positioned in the blocked meatus with Golay codes as the probe stimulus. In contrast to previous HRTF measurement systems, with *Snapshot* the listener moved on a rotating stool relative to a fixed loudspeaker position in order to measure different azimuth positions; the loudspeaker was then repositioned for different elevation measurements. The system was limited to a single diffuse-field equalization based on a specific set of headphones.

In 1998, the laboratory began work with William Chapin and Agnieszka Roginska (AuSIM Incorporated) to make major improvements to the *Snapshot* system, while keeping its best features. The system specifications for the improved system, called *Headzap,* can be summarized as follows:
 • Increased density of the measurement grid from 72 to 432 measurements on a sphere about the listener, using 12 loudspeakers on 3 poles (Figure 10).
 •Faster means of subject positioning while maintaining reliability and repeatability of measurement.
 • Improved fidelity in the time domain and optimized microphone signal-to-noise ratio;
 • Improved method for equalization to arbitrary headphone response;
 • Tools to easily adapt measured Head-Related Impulse Responses (HRIRs) to the PC-based rendering system used in our laboratory (*slab3d*).

Each of these desired specifications were addressed in the overall system design for both software and hardware, described in detail in [36]. One novel concept of the system was a visual feedback system for correct positioning of the head (*Aimer*), developed by Mark Anderson (co-author). A representation of the head position is supplied on a panel containing thirty red LEDs surrounding a central green LED, using real-time data from a head tracker (Figure 10, bottom). When the head is in the correct position, the green LED illuminates, the subject holds still, and the measurement is taken. This interactive 'closed-loop' system proved to be far faster than verbal communication between the experimenter and the subject. Ultimately, the advantage was that a complete HRTF set comprised of 432 measurements could be measured in less than half an hour, and subsequently rendered for use in spatial hearing experiments.
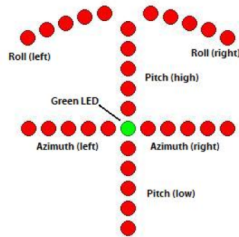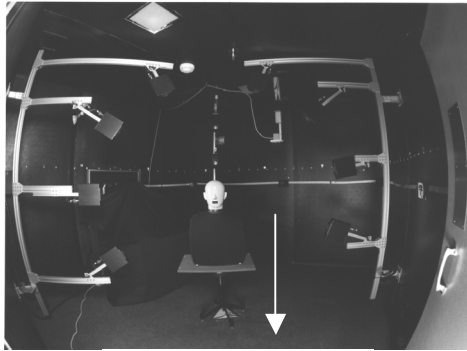
Figure 10. Top: *Headzap* as installed in a soundproof booth. Bottom: *Aimer* system for visual feedback (arrow shows location of LED panel).

8 CREW SOUND: VIRTUAL ACOUSTIC DISPLAYS FOR FLIGHT DECKS

From the standpoint of communications engineering, the design of an integrated approach for every type of acoustic information that arrives at the ears of a pilot is referred to as an *auditory display*–including radio communications, synthetic speech, caution and warning, and confirmatory audio feedback, relative to undesired "noise." Due to the anticipated reduction of air-ground communications in the Next Generation Air Transport System (NextGen), there will be an increased need for pilots to be aware of the current status or any significant changes within automated systems that control not only "ownship" (the pilot's own aircraft) but other aircraft in the vicinity. Increased interaction on the flight deck with automated systems for activities such flight planning, spacing, and "weather avoidance" will be necessary, and so it is incumbent on these systems to subtly yet effectively (from the standpoint of safety) make their users aware of the status of these multiple layers of automation.

The role of synthetic speech in an auditory display will likely be more prominent compared to its use for caution-warning systems. Synthetic speech driven by digital information can be used in place of actual radio communication voices for "party line" information from other aircraft that are currently monitored on the "same frequency" as own ship. Compared to purely non-speech alerts, synthetic speech can also be used to give supplemental informational content for problem solving

or situational awareness. As information rates increase, the need for multimodal presentation both visually and alternatively via an auditory display (or even via a haptic display) increases. Compared to listening, text-based systems have an obvious 'bottle neck' effect on the effective rate of information that can be transmitted and then acknowledged [37].

Given the potential advantages for future flight decks of varying prosody and speaking rate in spoken data link messages [38], a recent study by the authors sought to determine whether these manipulations impact the subjective comprehension effort and overall quality of the communications [39]. Ratings of overall quality and comprehension effort were obtained as a function of voice type, synthesized speech rate, and sentence prosody. Rank-order data analyses showed that both overall quality and comprehension effort were affected by speech rate: under the "fast rate" condition (vs. "default rate"), overall quality decreased and comprehension effort increased. However, the introduction of "prosodic emphasis" (pitch and level changes for specific phrases) in fast rate sentences produced a relative improvement in both comprehension and quality ratings. For both speaking rates, the introduction of "prosodic emphasis" resulted in higher quality ratings and lower comprehension effort ratings. The data suggest that faster speaking rates, which may improve message throughput in a display, may be viable when combined with prosodic emphasis.

The creation of synthetic speech displays was integrated into a real-time signal processing engine for an auditory display that handled all aspects of the audio simulation, including engine sounds, radio communications, spatialized and non-spatialized audio alerts, and synthetic audio feedback (e.g., switch sounds when interacting with a touch panel). This engine, comprised of off-the-shelf hardware and custom software, is collectively referred to as *CrewSound*. The initial hardware implementation of the *CrewSound* system included a fully configurable 24-channel audio interface (MOTU 24 I/O core system PCIe), multiple loudspeakers, supra-aural stereo aviation headsets with active noise cancellation and a customized push-to-talk capability (Sennheiser HMEC46-BV-K), and other peripherals. Software implementation included a custom graphical user interface enabling the specification, signal routing, and generation of up to 24 channels of synthesized speech messages and/or non-speech alerts by the user.

CrewSound was recently evaluated in a part-task simulation experiment, with professional pilots using an advanced concepts 777 flight simulator at NASA (Figure 11, top) [40]. The system was used to form a virtual acoustic auditory display where distinctive synthesized voices or alerts and communications

emanated from specific virtual audio locations over headsets, depending on function and urgency. This was acoustically overlaid on a loudspeaker-produced display of engine and system sounds. For example, standard caution and warning tones are emitted from a centralized loudspeaker; display announcements for off-nominal conditions such as "check spacing speed" used a female voice from a virtual position from the right; uplink messages used a distinct male voice from the right; etc.

In addition, *audio twitter* information consisting of route changes, altitude changes, and other pertinent information on the path or status of selected aircraft was presented to provide situational awareness in the manner of party line communications. *Audio twitters* were directionally referenced to the location of the subject aircraft, allowing an intelligibility advantage for multiple talkers.

Spatialized audio was also used to provide auditory feedback for conflict detection and routing tools. For example, a conflict alert that would normally be spatially acquired via a visual display (Figure 11, bottom) would emanate spatially from a position relative to ownship, much like the TCAS system described in section 2, above. In addition, auditory feedback cues were supplied to provide feedback that a correct sequence of "arm" and "engage" controls was performed.

Figure 11. Top: NASA's Flight Displays Research Simulator. Bottom: radar display showing conflicts.

9   CONCLUSIONS & ACKNOWLEDGEMENTS

These examples have been presented to give an overview of NASA Ames Research Center's role in developing improved human-machine interfaces using spatial audio. The efforts over the last 25 years have focused on a combination of hardware and software development, psychoacoustic experimentation, human factors applications research, and communications engineering. The Spatial Auditory Display Laboratory acknowledges support from within NASA from the Space Human Factors Engineering Program, and the Intelligent Integrated Flight Deck Project within the Aviation Safety Program. Development of *slab3d* has also been supported by the US Air Force Research Laboratory. The laboratory also acknowledges the support of their supervisors and collaborators both within and outside of the Human Systems Integration Division at NASA, and particularly the dedication and assistance we have received from our laboratory staff; including Richard Shrum, Brian Lathrop, Laura Tran, Alex Lee, Mitchell Clapp, and Bryan McClain.

REFERENCES

[1]   K. Pimentel, K. Teixeira *Virtual Reality. Through the New Looking Glass*, 2nd ed., McGraw Hill (1995).

[2]   Personal communication, Stephen Ellis, April 2010.

[3]   S. S. Fisher, E. M. Wenzel, C. Coler, & M. W. McGreevy, "Virtual interface environment workstations" *Proceedings of the Human Factors Society,* vol. 32, pp. 91-95 (1988).

[4]   E. M. Wenzel, F. L. Wightman, F. L., & S. H. Foster, "A virtual display system for conveying three-dimensional acoustic information" *Proceedings of the Human Factors Society,* vol. 32, pp. 86-90 (1988).

[5]   D. R. Begault, "Challenges to the successful implementation of 3-D sound" *Journal of the Audio Engineering Society*, vol. 39, no. 11, pp. 864-870 (1991).

[6]   E. M. Wenzel, "Localization in virtual acoustic displays" *Presence,* 1, 80-107 (1992).

[7]   F. L. Wightman, & D. J. Kistler, "Headphone simulation of free–field listening. I: Stimulus synthesis" *Journal of the Acoustical Society of America* vol. 85, pp. 858–867 (1989).

[8] F. L. Wightman, & D. J. Kistler, "Headphone simulation of free–field listening. II: Psychophysical validation" *Journal of the Acoustical Society of America* vol. 85, pp. 868–878 (1989).

[9] E. M. Wenzel, M. Arruda, D. J. Kistler, & F. L. Wightman, "Localization using non-individualized head-related transfer functions" *Journal of the Acoustical Society of America*, vol. 94, pp. 111-123 (1993).

[10] S. R. Ellis, S. R. Smith, A. J. Grunwald, & M. W. McGreevy, "Direction judgement error in computer generated displays and actual scenes" *Pictorial communication in virtual and real environments. Taylor and Francis*. London (1991).

[11] B. D. Adelstein, E. R. Johnston, & S. R. Ellis, "Spatial sensor lag in virtual environment systems" *Proceedings, SPIE Conference on Telemanipulator Technology, Bellingham WA*, vol. 1833, pp. 187-198 (1992).

[12] http://human-factors.arc.nasa.gov/

[13] D. R. Begault & E. M. Wenzel, "Technical aspects of a demonstration tape for three-dimensional auditory displays" *NASA Technical Memorandum* No. 102286 (1990).

[14] E. C. Cherry, "Some experiments on the recognition of speech with one and two ears" *Journal of the Acoustical Society of America*, vol. 25, pp. 975–979 (1953).

[15] J. C. Webster & P. O. Thompson, "Some audio considerations in air control towers" *Journal of the Audio Engineering Society* vol. 2, pp. 171-175 (1953).

[16] D. R. Begault, "Multi-channel spatialization system for audio signals" U.S. patent 5,438,623 (1995).

[17] D. R. Begault, "Virtual acoustic displays teleconferencing: intelligibility for 'telephone-grade' audio" *Journal of the Audio Engineering Society*, vol. 47, pp. 824-828 (1999).

[18] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA: MIT Press (1990).

[19] D. R. Begault, M. R. Anderson, J. D. Miller, &

B. M. McClain, "Audio-Visual Communication Monitoring System for Enhanced Situational Awareness", Poster presentation, *Working Together: R&D Partnerships in Homeland Security Conference*, Boston (2005).

[20] D. R. Begault, M. R. Anderson, J. D. Miller, & B. M. McClain, "Reconfigurable Auditory-visual Display" U.S. Patent 7,378,963 (2008).

[21] D. R. Begault, "A head-up auditory display for TCAS advisories" *Human Factors*, vol. 35, pp. 707-717 (1993).

[22] D. R. Begault & M. T. Pittman, "Three-Dimensional Audio Versus Head-Down Traffic Alert and Collision Avoidance System Displays" *International Journal of Aviation Psychology*, vol. 6, pp. 79-93 (1996).

[23] J. J. Gibson, *The ecological approach to visual perception*. Boston: Houghton Mifflin (1979).

[24] M. Godfroy, C. Roumes, and P. Dauchy, "Spatial variations of visual-auditory fusion areas" *Perception* vol. 32, pp. 1233-1245 (2003).

[25] M. Godfroy & C. Roumes, "Multisensory enhancement of localization with synergistic visual-auditory cues" *Proceedings of the International Conference on Cognitive Sciences, COGSCI 2004*, Chicago (2004).

[26] M. Godfroy & R. B. Welch, "Two-dimensional localization with spatially and temporally congruent visual-auditory cues" *Proceedings of the 46th conference of the Psychonomic Society*, Toronto (2005).

[27] M. Godfroy, R. B. Welch, P. M. Sandor, & C. Roumes, "The Interaction of Vision and Audition in Two-dimensional Space" *Perception* (submitted).

[28] F. Frassinetti, N. Bolognini, & E. Ladavas, "Enhancement of visual perception by crossmodal visuo-auditory interaction" *Experimental Brain Research* vol. 147, pp. 332-343 (2002).

[29] E. M. Wenzel, J. D. Miller, & J. S. Abel, "Sound Lab: A real-time, software-based system for the study of spatial hearing" *Audio Engineering Society 108th Convention*, Paris, Preprint 5140 (2000).

[30] J. D. Miller & E. M. Wenzel, "Recent developments in SLAB: A software-based system for interactive spatial sound synthesis" *ICAD 2002, International Conference on Auditory Display,* Kyoto (2002).

[31] http://slab3d.sonisphere.com/

[32] J. M. Loomis, J. R. Marston, R. G. Golledge, & R. L. Klatzky, "Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance" *Journal of Visual Impairment & Blindness*, vol. 99, pp. 219-232 (2005).

[33] ISO 7731-2003, "Ergonomics. Danger signals for public and work areas. Auditory danger signals" International Organization for Standardization, Geneva (2003).

[34] D. R. Begault, "Auditory alert systems with enhanced detectability", U.S. Patent 7,346,172 (2008).

[35] D. R. Begault, M. R. Anderson, & B. M. McClain, "Spatially-modulated auditory alerts for aviation" *Journal of the Audio Engineering Society* vol. 55, pp. 276-282 (2006).

[36] D. R. Begault, M. Godfroy, J. D. Miller, A. Roginska, M. Anderson, & E. M. Wenzel, (2006). "Design and verification of HEADZAP, a semi-automated HRIR measurement system" *Audio Engineering Society 120th Convention*, Paris, preprint 6655 (2006).

[37] S. A. Lozito, A. McGann, & K. Corker, "Data link air traffic control and flight deck environments: Experiments in flight crew performance" *Proceedings of the 7th International Symposium on Aviation Psychology,* Columbus, OH, pp. 1009-1015 (1993).

[38] D. S. Brungart, B. D. Simpson, & N. Iyer, "Maximizing information transfer in auditory speech displays" *Proceedings of the 19th International Congress on Acoustics. Madrid* (2007).

[39] M. Godfroy, E. M. Wenzel, & D. R. Begault, "Speech Synthesis for Data Link: A Study of Overall Quality and Comprehension Effort" *Proceedings of the 15th International Symposium on Aviation Psychology,* Dayton, OH (2009).

[40] E. M. Wenzel & D. R. Begault, "CrewSound: An Auditory Display System for the Study of Advanced Cockpit Communications", report for NASA Intelligent Integrated Flight Deck project (2009).