SPATIALLY-MODULATED AUDITORY ALERTS

Durand R. Begault[a], *Mark R. Anderson*[b] *and Bryan U. McClain*[c]

[a] Human Factors Research & Technology Division
[b] QSS Group, Inc. [c] San José State University
[a,b,c] Spatial Auditory Display Laboratory
MS 262-2, NASA Ames Research Center
Moffett Field, California 94035-1000 USA
Durand.R.Begault@nasa.gov

ABSTRACT

The technique of spatial modulating ("jittering") an existing or newly created auditory alert along a virtual trajectory to improve its detection in a noisy environment is described. Threshold data from 14 participants were gathered as a function of spatial modulation rate (0, 1.6 and 3.3 Hz) of an avionics "wind sheer" alert (two successive 300 ms square waves) against a steady state background noise (Boeing 737-300 flight deck ambient sound). 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 non-jittered alert. Non-head tracked virtual simulation of the background noise over headphones yielded jittered alert thresholds within 1.5 dB of loudspeaker simulation.

1. INTRODUCTION

Design methodologies for insuring human ability to detect the presence of a non-speech auditory alert (alarm) that is part of a warning system have for the most part concerned with an analytical approach to the amplitude spectrum of the alert and the background noise. International Standard 7731 covers the formation of auditory alerts for danger signals and states that certain frequency components be >= 13 dB above the masked threshold within 1/3 octave bands from 300-3000 Hz [1]. It is well understood from the auditory literature that, by making spectral components of an alert substantially louder than the measured background noise level, one can insure for the audibility or "detection" of such a signal [2, 3]. Such detection is referred to as "release from masking" in that spectral components of the signal are sufficiently greater in amplitude such that they may be heard. However, the technique of unmasking an auditory alert masking by means of spatial manipulation of the signal is unexplored. In addition there has not been a method taught where existing signals can be enhanced.

In high-stress environment such as an airline flight deck, an approach to insuring alert detection via emphasis on the amplitude of spectral components as opposed to spatial manipulation methods is primarily due the relatively poor quality of communication equipment and the use of monaural loudspeaker or single-earpiece headset playback systems. Eventually, binaural headphone or loudspeaker systems may become incorporated into flight deck, 911 consoles, machine operator transportation, media, and communications devices that will require the human to process a great deal of information. Also, the background noise of these types of environments will be continually reduced, allowing for presentation of informative or alerting types of signals without the need for levels that have the potential to create a "startle" effect [4]. The design of an auditory alert using spectral amplitude as a criterion is potentially problematic because other desirable signals as opposed to noise may be masked. There is a compelling motivation to provide critical alert information in a manner that allows the alert to be both audible and recognizable, but in as "quiet spoken" a manner as is possible.

In light of the above, a method has been described in a patent application that provides for the synthesis and two-channel playback of an auditory alert that is relatively more detectable against a background of noise, relative to one-channel "monaural" or "diotic" playback of the same auditory alert [5]. There are three components to this method, of which the "spatial modulation" technique is described in the current paper.

The technique of spatial modulation ("jitter") of an auditory alert along the auditory azimuth involves taking a one-channel alert signal and processing it into a two-channel signal for headphone playback, moving from a central location at 0 degrees azimuth to a position 45-90 degrees azimuth to the right, and then to the mirror image position at least 45-90 degrees azimuth to the left, and then back to 0 degrees: see Figure 1. In the presence of steady state background noise, which is relatively unvarying in its spatial properties, it was hypothesized that a spatially jittered alert is more detectable than one which is not spatially jittered.

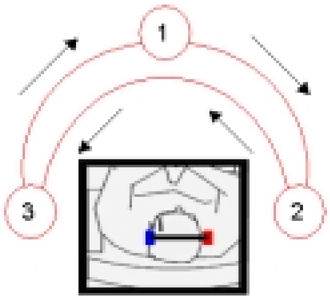

Figure 1. *A "jittered" spatial modulation.*

From a design standpoint, the alert should not bring attention to any particular spatial location during its excursion. As exemplified by everyday experience with fast-moving insects and vehicles, attention may be stimulated by the activation of the auditory system's response to dynamic changes in interaural relationships, but within certain ranges of source velocity it is difficult to actually localize a moving source.

Evidence has also been gathered recently for an auditory motion detector, independent of a mechanism for localization, that is potentially analogous to a visual motion detection mechanism in the cererbal cortex [6-9]. For the technique described here, the spatial modulation frequency (which corresponds to the frequency of amplitude or time delay modulation) should be in the region of 2-4 Hz, and no greater than 10 Hz. For interaural time delay manipulation, this range is where the auditory system perceives movement, but is relatively bad at following the location, a phenomenon known as "binaural sluggishness" [10-12].

There are many methods by which to implement spatial jitter. It can be done by means of linear or exponential amplitude panning, or by continuously varying a time delay to each ear in the range 0 – ca. 0.8 ms. Finally, binaural variations in time and amplitude as a function of HRTF convolution can be implemented via a 3-D sound interface that allows movement of a virtual source to a listener. The latter technique was used in the current study.

The advantage of spatially modulating an existing alert is that it is still inherently recognizable as the alert itself, despite the spatial modification. Although spectral components are modified in amplitude by dynamic Head-Related Transfer Function (HRTF) filtering, the temporal-spectral gestalt of the alert remains recognizable to a pilot or other machine operator.

2. EXPERIMENT METHODOLOGY, SUBJECTS

Absolute thresholds for auditory alerts in the presence of background noise were measured under six different experimental conditions (block types) that were selected for a series of planned comparisons in the analysis. The acoustic environment was modeled on data and recordings made in a 737-300 flight deck from the jump seat position [13]. The experiment was designed to allow the following alert thresholds comparisons to be made: (1) monaural loudspeaker (representative of the "existing" flight deck condition) versus a spatially jittered, headphone delivered alert; (2) non-spatially jittered versus jittered alerts, delivered over headphones; (3) comparison of 1.66 versus 3.33 Hz jitter rate; and (4) comparison of alert thresholds using virtual (headphone) versus loudspeaker simulation of background noise. Fourteen participants (subjects) (ages 18-29, 8 male, 6 female) participated; all were screened for normal hearing (< 30 dB HL from 125 Hz – 8 kHz).

The alert used was a two-tone alarm that precedes a verbal "wind shear" warning on a Boeing 737-300 (two successive 300 ms square waves, 0.25 and 1 kHz, total duration 600 ms). The background noise used was a 2.5 recording of the ambient noise of the Boeing 737-300 flight deck under cruise conditions. Alert and noise levels were calibrated for both headphone and loudspeaker conditions using a head and torso simulator (Bruel and Kjaer 4100D) and a real-time frequency analyzer (Hewlett-Packard H5670A). The background noise level (2 second Leq) was set at 79 dB, A-weighted. The alert level was calibrated to 79 dB using peak-hold averaging over a 125 ms interval. The r.m.s. level of the two square waves were identical.

The alert and the background noise were presented either through loudspeakers or headphones (Senheisser HD 480s), and the alert was modulated at different rates, depending on the experimental condition. Table I summarizes these conditions, which made up the experimental block types. Participants were run under each block type four times, for a total of 24 experimental blocks per participant.

Block type	Noise Source	Alert Source	Spatial Modulation (Hz)
1	Loudspeaker	Loudspeaker	0
2	Loudspeaker	Headphone	1.66
3	Loudspeaker	Headphone	3.33
4	Headphone	Headphone	0
5	Headphone	Headphone	1.66
6	Headphone	Headphone	3.33

Table I. *Experimental conditions.*

Figure 2 shows the physical aspects of the experimental configuration. The experiment was conducted in double-walled soundproof booth having a background noise level of 15 dB (A-weighted). Participants were seated about 4 feet from loudspeaker 2 (JBL "Proformer", 5 inch driver) which provided the loudspeaker alert for the first condition (block type 1); participants were instructed to remove headphones for this condition. During all trials, subjects were instructed to fixate gaze at a point indicated on loudspeaker 2. Loudspeakers 1 and 3 (Genelec 1029A) and dodecahedron loudspeaker 4 (Bruel and Kjaer 4186) provided the aircraft noise for conditions 1-3, and were arranged to make the sound field diffuse as possible (the A-weighted level changed less than +/- 1 dB with normal head movement). The loudspeakers were not used for conditions 4-6 (both the alert and the noise were played over headphones). The aircraft noise for conditions 4-6 was produced from a binaural recording of the noise under condition 1-3 made at the position of the participant.

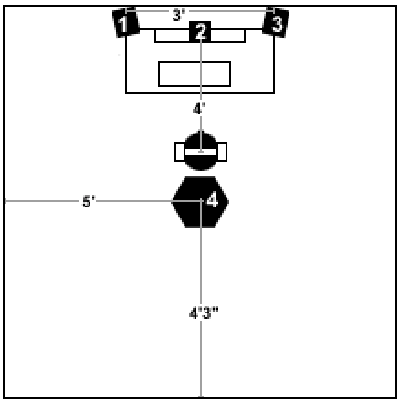

Figure 2. *Experiment configuration of loudspeakers (numbered) and listener within the soundproof booth.*

Azimuth angles (referenced to 0° at a point directly in front of the listener) were simulated via real-time head-related transfer function (HRTF)-filtering via the SLAB real-time, software-based 3-D audio processor [14, 15]. Client software communicated with the SLAB server via a tcp/ip loopback connection. Data was gathered from the subject via a two-button switchbox interfaced to the mouse port. For the headphone-based alerts, the SLAB processor's trajectory interface was programmed to provide modulation of 1.66 Hz (0.6 degrees/ms) or 3.33 Hz (1.2 degrees/ms) to left and right 90 degrees of the listener, as indicated in Figure 1. Non-individualized HRTFs were used. Calibration of signals and change of configuration between blocks was accomplished via

MIDI program changes to a digital mixer (Tascam DM24) from the host computer.

Each trial within an experimental block consisted of two iterations of the 2.5 s noise recording, with a 1 second interval of silence in between; the alert was randomly assigned to either the first or second iteration. Participants indicated their response via a push-button interface as to which stimulus interval contained an auditory alert. Using a two-alternative forced-choice paradigm, thresholds were obtained at the 70.7% level within a tolerance of 1 dB with a "one up-two down" adaptive staircase algorithm that adjusted the level of the alert relative to a fixed noise level [16]. The alarm level (peak) was initially presented at –4 dB relative to the noise level (Leq). The staircase began with an 8 dB step size, and reduced in level by 50% until the 1 dB step size was reached. An experimental block consisted of as many trials a necessary to reach 4 reversals at the minimum step size. Thresholds were defined for each subject and for each block as the mean value of the four final staircase reversals at the minimum level of 1 dB; a typical block took about 5 minutes to complete.

3. RESULTS

3.1. Threshold comparison: Non-jittered loudspeaker alert versus a spatially jittered, headphone delivered alert.

This comparison was made to determine how the "existing" condition for auditory alerts on the flight deck, where the alert is provided through a monaural loudspeaker, compares to a spatially jittered, headphone delivered alert. This corresponds to conditions 1 versus condition 2 and 3 in Table 1. Figure 3 shows the data for these conditions (mean and standard deviation for all participants), with the threshold plotted relative to the r.m.s level of the background noise. The advantage of mechanical noise attenuation provided by the headphone, combined with the jittered alert, was on average 13.4 dB compared to the loudspeaker alert.

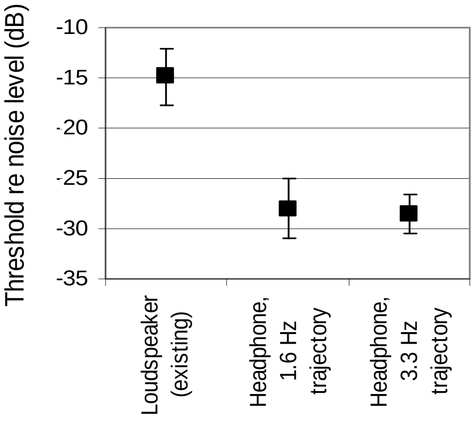

Figure 3. *Loudspeaker alert versus headphone alert.*

The passive attenuation of the noise provided by the headphones is shown in Figure 4. Below the 2500 Hz octave band, the measured attenuation provided is less than 1 dB, possibly due to the open venting of the circumaural shell of the headphone. The square wave alert (particularly that with the 1 kHz fundamental) will have significant harmonic energy in the frequency region where the headphone provides maximal attenuation.

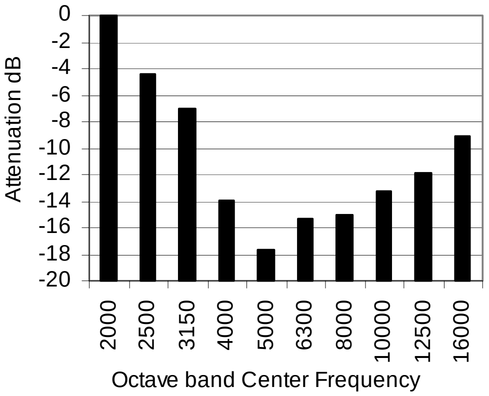

Figure 4. *Attenuation of noise by Sennheiser HD-480.*

3.2. Non-spatially jittered versus jittered alerts, delivered over headphones

Paired comparisons were made of experimental conditions 4-6 (ref. Table I), which used headphone presentation of the alarm and binaural simulation of the loudspeaker-delivered aircraft noise. These analyses were made to test the effect of alert jittering and to determine the effect of trajectory velocity rate. The independent variable was trajectory velocity (0.0 Hz, or no jitter; 1.6 Hz trajectory; and 3.3 Hz trajectory).

Figure 5 indicates the mean and standard deviation values of the results across the fourteen subjects tested. The mean thresholds were –13.3 dB for no jittering; -21.2 dB for the 1.6 Hz trajectory, and –21.0 for the 3.3 Hz trajectory. Analyses of variance (ANOVAs) were run for each pair wise comparison. Condition 4 was significantly different than condition 5 [F(1,13) = 203, p < .000], and significantly different than condition 6 [F(1,13) = 274, p < .000], but there was no significant difference between condition 5 and 6, that is, between the two moving trajectory velocities tested. The overall reduction in threshold caused by spatial jittering at either rate is about 7.8 dB.

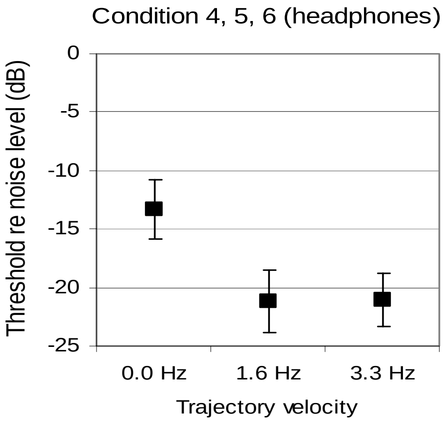

Figure 5. *Effect of trajectory velocity on alarm threshold.*

3.3. Comparison of alert thresholds under equivalent loudspeaker and virtual (headphone) conditions

A comparison was made of thresholds obtained using loudspeakers versus headphones as the source of an unjittered alert and noise simulating the 737-400 flight deck, for conditions 1 and 4 described in Table I. This allows an estimate of the quality of the 3-D audio virtual simulation of loudspeakers for threshold estimation. Condition 1 used the loudspeakers shown in Figure 2 for the background noise (loudspeakers numbered 1,3 and 4) and for the alert (loudspeaker number 2). Condition 4 used a virtual simulation of loudspeaker 2 for the alert (the alert was played through a 0 degree azimuth, 0 degree elevation HRTF) and a calibrated, binaural recording of noise played through loudspeakers 1,3 and 4 in Figure 2.

The mean thresholds were very close: –14.8 dB for condition 1 (real loudspeakers) versus -13.3 dB for condition 4 (virtual loudspeakers); see Figure 6. An ANOVA indicated that the difference was significant [$F(1,13) = 6.0$, $p = .029$]. This 1.5 dB lower threshold under the real loudspeaker condition may have been due to slight head movements made by the participants, despite instructions to keep the head steady and eyes fixated at loudspeaker 2. Slight head movements could have effectively introduced a “jittering” advantage to the alert, in that the source is moved relative to the listener.

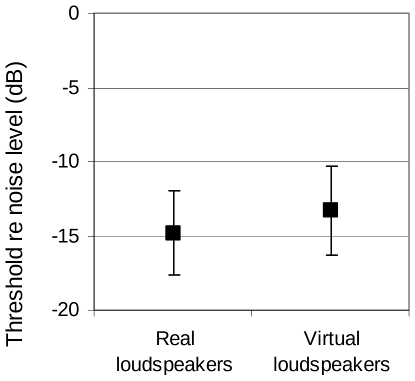

Figure 6. *Thresholds under real and virtual conditions (experimental conditions 1 and 4).*

4. CONCLUSIONS

These data support a new approach to the design of an auditory alert for flight decks or other high-stress human interfaces where stereo headphones can be used. It is not known as of yet what the advantage would be for spatial modulation using multiple loudspeakers. Through the use of headphone presentation and spatial modulation, it is possible to improve the detection of an alarm by up to 13.4 dB, partially due to the passive noise reduction afforded by the headphones. The effect of spatial modulation of an alert, at velocities of 1.6 or 3.3 Hz relative to an unmodulated alert, is to improve detection by 7.8 dB. Virtual simulation of loudspeaker alerts can bias thresholds slightly upwards, compared to evaluations made using virtual 3-D audio simulations, possibly due to slight head movements made in front of the real loudspeaker. Further testing of threshold equivalency between virtual and real loudspeakers using a chin rest or other means for preventing head movement would allow for a more controlled comparison between these conditions.

5. REFERENCES

[1] ISO 7731. "Danger signals for work places-Auditory danger signals." 1995, *International Organization for Standardization*, Geneva.

[2] Moore, B.C.J., ed. *Hearing. 2nd ed.* "Handbook of Perception and Cognition*"*. 1995, Academic Press: San Diego.

[3] Zwicker, E., and Fastl H., *Psychoacoustics.* "Facts and Models." 1990, Berlin: Springer-Verlag.

[4] Patterson, R., "Guidelines for auditory warning systems on civil aircraft." 1982, *Civil Aviation Authority*, London.

[5] Begault, D.R., "Making audible alarms more noticeable in noisy environments*." NASA Tech Briefs*, 2002. 26(9).

[6] Rashbass, C., "The relationship between saccadic and smooth tracking eye movements." *Journal of Physiology* (London), 1961. 159: p. 326-338.

[7] Krukowski, A.E., Begault, D.R., Wenzel, E.M., and Stone, L.S., "Human oculomotor responses to virtual auditory motion." *Society for Neuroscience 31st Annual Meeting*. 2001. San Diego, CA.

[8] Krukowski, A.E., Begault, D.R., Wenzel, E.M., and Stone, L.S., "Human smooth pursuit eye movement responses to visual, auditory, and imagined target motion (abstract)." *The Society for the Neural Control of Movement Twelfth Annual Meeting.* 2002. Naples, Florida.

[9] Krukowski, A.E., Begault, D.R., Wenzel, E.M., and Stone, L.S., "Pursuing the invisible: Human oculomotor response to auditory motion." *Nature* (submitted), 2003.

[10] Grantham, D.W., "Spatial Hearing and related phenomenon," *Hearing*, B.C.J. Moore, Editor. 1995, Academic Press: San Diego.

[11] Grantham, D.W., and Wightman, F.L., "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation." *Journal of the Acoustical Society of America*, 1979. 65: p. 1509-1517.

[12] Grantham, D.W., and Wightman, F.L., "Detectability of varying interaural temporal differences." *Journal of the Acoustical Society of America*, 1978. 63: p. 511-523.

[13] Begault, D.R., "Virtual acoustics, aeronautics and communications." *Journal of the Audio Engineering Society*, 1998. 46(6): p. 520-530.

[14] Miller, J.D., and Wenzel, E. M., "Recent Developments in SLAB: A Software-Based System for Interactive Spatial Sound Synthesis." *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan. 2002. Kyoto, JP.

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

[16] Levitt, H., "Transformed up-down methods in psychoacoustics." *Journal of the Acoustical Society of America*, 1970. 49(2): p. 467-477.