INTELLIGIBILITY OF STEREO AND 3D-AUDIO CALL SIGNS FOR FIRE AND RESCUE COMMAND OPERATORS

Otto Carlander, Mattias Kindström and Lars Eriksson

FOI, Swedish Defence Research Agency,
Div. of Command and Control Systems,
Dept. of Man-System Interaction,
Linköping, Sweden.
otto.carlander@foi.se

ABSTRACT

A command operator of fire and rescue units may need to pay attention to several radio calls in the coordination of simultaneous emergency missions. An experiment investigated command operators' ability to discern stereo and 3D-audio call signs presented in background noise of added voice sources. Each of 10 command operators listened to one to four call signs combined with two to four background voices, with the primary task to discriminate the voice of each call sign. A secondary visual and manual response task induced an overall high mental workload. 3D-audio presentation resulted in a slightly increased number of correctly identified call signs. Four background voices reduced accuracy compared to two, and both three and four simultaneously presented call signs resulted in lower accuracy compared to sets of one and two, respectively. The results are discussed in relation to the potential for improving the 3D-audio presentation aiming for increased intelligibility and operator effectiveness.

1. INTRODUCTION

Good performance in demanding tasks often requires intense attention, especially when the task-crucial information comes in practically simultaneous quick bursts. Command operators at Swedish fire and rescue departments handle several simultaneous short-period voice streams to coordinate materiel and personnel for ongoing rescue missions. That is, a command operator may abruptly process up to four sets of brief auditory stimuli from four separate radio channels. Managing the important auditory information during intense periods generates relatively high mental workload conditions.

Today, the command operators have headphones and speakers as options for listening to their radio channels. Stereo panning is used for adjusting the separation of sound sources in the headphones. For the speakers, each radio channel corresponds to one (and only one) chosen speaker, with speakers distributed horizontally over about 1 m with fixed separation. Both headphones and speakers result in limited intelligibility of sound sources, partly because of their limited spatial separation.

The utilization of 3D-audio technology introduces the possibility to position sounds in a virtual space at numerous positions relative to the listener. Previous research has shown that angular separation increases intelligibility [1][2][3]. Using a 3D auditory display increases the capability of listening to parallel channels, and the risk of misinterpreting *who* says *what* can decrease [4]. Thus, 3D-audio technology can improve radio communication and lay a foundation for more effective command operators that perhaps also need less mental effort.

2. EXPERIMENT

The current auditory tools for radio communication at Swedish fire and rescue departments can be improved. It could be crucial that a command operator is more alert for, and can better manage, simultaneous or almost simultaneous calls on several radio channels. An experiment was therefore carried out to investigate whether 3D-audio could increase intelligibility and performance compared to stereo presentation. The main difference compared to previous studies is that the experiment included an overall high workload setting, roughly resembling intense work periods for command operators, in combination with using a novel 3D auditory display based on a commercials-off-the-shelf (COTS) soundcard.

2.1. Method

Apparatus. A PC with monitor and a soundcard (Hercules Gamesurround MUSE Pocket) was used. All auditory stimuli were recorded on the PC with a Shure M58 microphone and a microphone preamp, and the speech signals were high-pass filtered at 100 Hz and low-pass filtered at 8 kHz. Speech signals were presented in AKG k240 studio headphones with a frequency range of 15 – 25000 Hz.

Design and stimuli. The experiment had a two × four × three factorial within subjects design. It included two auditory display technologies used for presenting one to four call signs in background noise. The two auditory displays generated 3D-audio and stereo sound, representing one novel and one more traditional way of presenting radio communication. Each of the four levels of call signs consisted of presenting a single call sign, two, three or four simultaneous call signs. The three levels of background voices comprised two, three or four simultaneously presented background voices reading different texts with at least one background voice per ear and a SNR of 0 dB. Each call sign had a duration of 2.5 s, and call signs were slightly adjusted out of phase but completed within 3 s. The interval between presentations of each set of call signs was randomized between 4 and 16 s. Each condition was repeated three times, resulting in 72 presentations for each participant.

The primary task was to identify one (single) to four (set) simultaneous call signs among two to four background voices. A correctly identified single call sign was defined as the identification of a single speaker. A correct identification of a complete set of call signs was defined by the identification of all speakers in a set. The call signs consisted of a spoken command call "102 over, 102 over" and, when identified, the

subject used the computer mouse to indicate who or whom that called. This was done in a response form shown on the computer screen seen in Figure 1 below. A secondary task was used for inducing a heightened mental workload. The secondary task consisted of a visual and manual-response task shown in Figure 1.

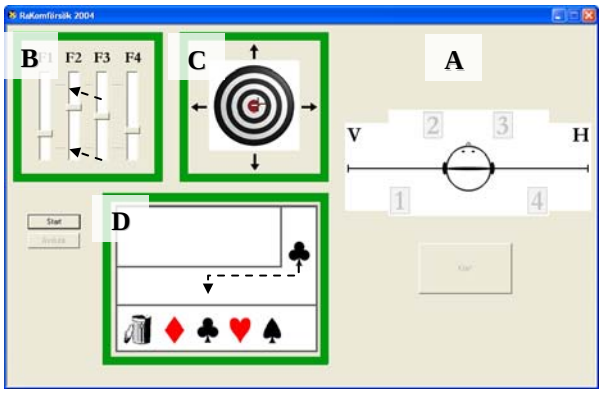

Figure 1. *(A) The call sign response form, primary task. (B) Levelers were supposed to be kept within the small markings. (C) Cross aim was supposed to be kept over bull's-eye. (D) Symbols were supposed to be placed at their match. (B), (C) and (D) composed the secondary task that was controlled by the keyboard.*

Each participant performed two training sessions, one for stereo and one for 3D-audio, both with low workload. The workload was adjusted by parameter settings of the secondary task such as speed of symbols, levelers and cross aim movements. The experiment with a higher workload was then performed with the presentation order of stereo and 3D-audio counterbalanced over participants. The experimental conditions were randomized within each session.

Procedure. After a brief introduction, the participant read written instructions followed by verbal instructions by the experimenter. The training session with an overall low workload condition was then completed, consisting of two blocks, stereo and 3D-audio presentation respectively. Next, the experiment proper with the higher workload condition began with 36 auditory stimuli for both stereo and 3D-audio presentation, respectively, totaling 72 presentations. Each session lasted about 1 h for each participant.

Participants. 10 male command operators from the staff at a Fire and Rescue Department in Stockholm participated. They were all naïve about using 3D-auditory displays, but familiar with stereo displays.

2.2. Results and conclusions

Repeated ANOVA measures were applied to each of the means of correctly identified single call signs and correctly identified complete sets of call signs. Each analysis included 24 means (2 × 4 × 3 = 24) for each participant, with each mean calculated from three trials of each call sign or set of call signs in each condition. All ANOVA *p*-values are hereafter given with the Greenhouse-Geisser correction values.

Identified single call signs. The ANOVA of correctly identified single call signs showed significant main effects of technology, $F(1, 9) = 6.51$, $p < .05$, and background voices, $F(2, 18) = 5.84$, $p < .025$, with no other significant effects. The stereo technology with mean proportion correct (M) = 0.57 and standard error of (SE) = 0.04 generated less accuracy compared to 3D-audio with M = 0.63 and SE = 0.03. Figure 2 illustrates the main effect of presentation technology. The main effect of background voices is shown in Figure 3. A Tukey HSD test revealed a higher accuracy with two background voices, M = 0.64, SE = 0.03, compared to four, M = 0.55, SE = 0.04 ($p <$ .01).

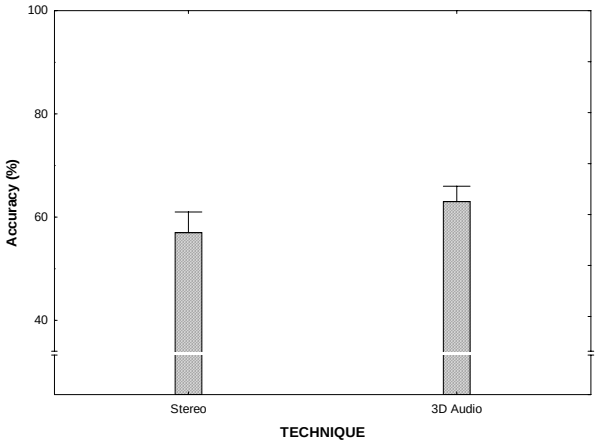

Figure 2. *Mean accuracy of identified call signs with stereo and 3D-audio. Mean + SE.*

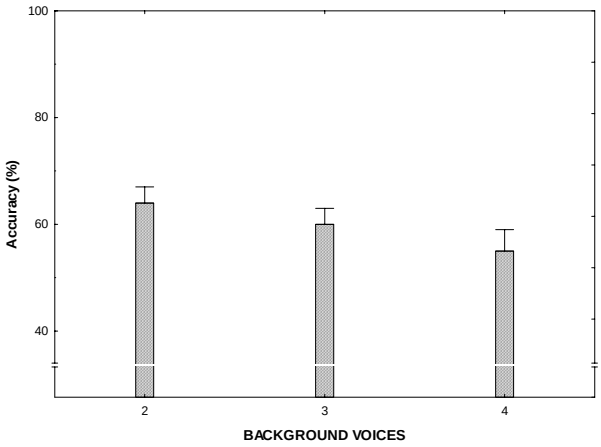

Figure 3. *Mean accuracy of identified call signs with 2-4 background voices. Mean + SE.*

Identified complete sets of call signs. The ANOVA of correctly identified sets of call signs revealed significant main effects of technology, $F(1, 9) = 5.89$, $p < .05$, background voices, $F(2, 18) = 5.94$, $p < .05$, and set size of call signs, $F(3, 27) = 74.98$, $p < .0001$, with no other significant effects. Stereo, M = 0.27, SE = 0.04, generated less accuracy compared to 3D-audio, M = 0.32, SE = 0.04, see Figure 4.

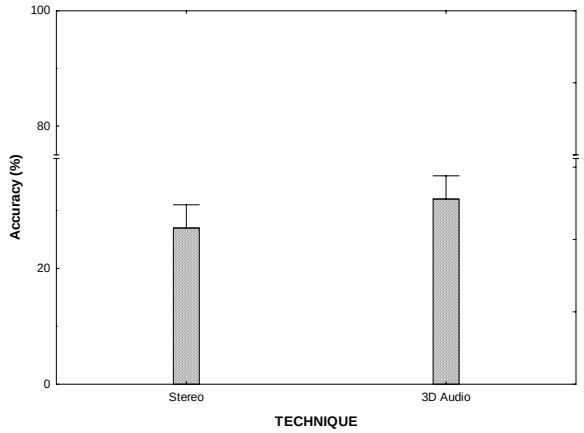

Figure 4. *Mean accuracy of identified complete sets of call signs with stereo and 3D-audio. Mean + SE.*

A Tukey HSD test revealed that two background voices, M = 0.35, SE = 0.04, resulted in higher accuracy than four, M = 0.24, SE = 0.04 ($p < .0001$). Set sizes of both one and two call signs, M = 0.65, SE = 0.05, and M = 0.38, SE = 0.06, showed higher accuracy than sets of three and four, respectively, M = 0.09, SE = 0.04, M = 0.06, SE = 0.03 ($p < .05$ for all comparisons). The main effects of background voices and set sizes are illustrated in Figures 5 and 6, respectively.

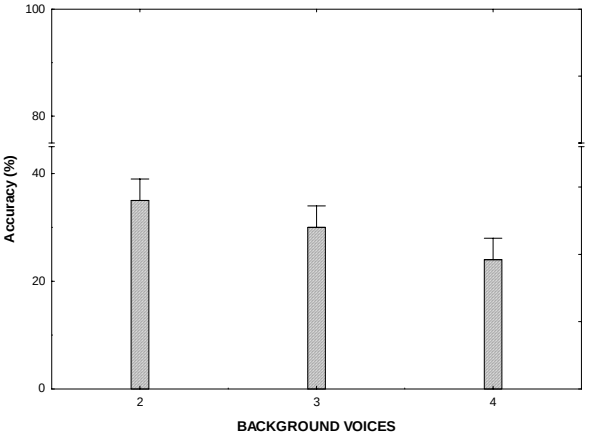

Figure 5. *Mean accuracy of identified complete sets of call signs with 2 to 4 background voices. Mean + SE.*

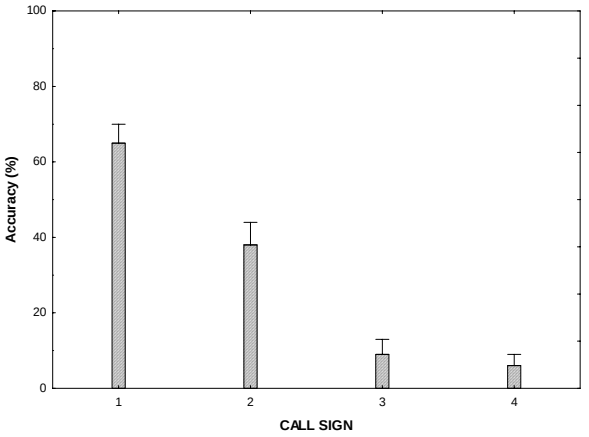

Figure 6. *Mean accuracy of identified complete sets of call signs. Mean + SE.*

In sum, stereo presentation generated less proportion of correctly identified separate call signs. Identification of separate call signs also showed a higher accuracy with two background voices present compared to four. Stereo presentation also revealed less accuracy for identified complete sets of call signs. Two background voices resulted in higher accuracy than four, and set sizes of both one and two call signs showed higher accuracy than sets of three and four, respectively.

3. DISCUSSION

The results imply that command operator ability to discern call signs is improved by 3D-audio, although with only a small improvement compared to stereo. The conclusion is that 3D-audio offers a slightly better intelligibility of call signs in the high workload condition used, and most probably because of the increased spatial separation of call signs.

The experimental setup for the high workload condition was intended to resemble the workload potentially occurring for an operator. However, by primarily introducing many background voices and a SNR of 0 dB, the primary task becomes difficult. While this was partly expected [2][4][6][8] the many background voices in combination with a demanding secondary task probably made conditions too difficult regardless of auditory display.

Previous experiments with COTS technology for 3D-audio have shown good performance compared to more advanced systems. In comparison to a professional research platform regarding azimuth "localization" of one sound source, the COTS technology performance was in fact superior [5]. Thus, this 3D-audio technology was chosen for this experiment. However, it had not previously been tested for simultaneous sound sources. Some of the participants claimed that the perception of the 3D-audio presentation was not spacious and that sound sources blended. A professional research platform for 3D-audio might overcome this by its better support of simultaneous presentation of sound sources.

Previous studies show that 3D-audio can be more effective for presenting simultaneous sound sources [1][2][4][6][7]. Improved intelligibility of radio communication could be of vital importance for fire and rescue command operators, reducing the risk of misinterpretations and missed call signs.

Further investigations include using a 3D-audio platform that better handles simultaneous sound sources, reducing the number of background voices to one in each ear, and an easier secondary task.

4. REFERENCES

[1] D.R. Begault, "Virtual acoustic display for teleconferencing: Intelligibility advantage for '"telephone-grade'" audio," *Journal of Audio Engineering Society*, vol. 47, no. 10, pp. 824-828, 1999.

[2] W.T. Nelson, R.S. Bolia, M.A. Ericson, and R.L. McKinley, "Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment," in *Proceedings of the IMAGE Conference*, Chandler, USA, 1998, pp. 159-166.

[3] T.J. Doll, T.E. Hanna, and J.S. Russotti, "Masking in three-dimensional auditory displays," *Human Factors*, vol. 34, no. 3, pp. 255-265, 1992.

[4] E.C. Haas, "Utilizing 3-D auditory display to enhance safety in systems with multiple radio communications," in

S. Kumar (Ed.), *Advances in Occupational Ergonomics and Safety*. IOS Press, Amsterdam, The Netherlands, 1998.

[5] O. Carlander, L. Eriksson, and M. Kindström, ”Perceived accuracy of horizontally distributed sounds of two 3D-audio display technologies.,” Manuscript in preparation.

[6] R. Drullman, A. W. Bronkhorts, “Multichannel speech intelligibility and talker recognition using, monaural, binaural and three-dimensional auditory presentation,” *Journal of Acoustical Society of America*, vol. 107, no. 4, pp. 2224-2235, April 2000.

[7] J. Baldis, “Effects of Spatial Audio on Memory, Comprehension, and Preference During Desktop Conferences,” in *Proceedings of CHI 2001*, Seattle, USA: March-April, 2001, pp. 166-173.

[8] C. Wickens and J. Holland, *Engineering psychology and human performance.* Prentice-Hall Inc, Upper Saddle River, USA, 2000.