Detection and Identification of Simultaneous Communications in a Simulated Flying Task

Ken McAnally, Russell Martin, Jodie Doman, Geoff Eberle & Simon Parker
Air Operations Division, Defence Science and Technology Organisation,
PO Box 4331, Melbourne 3001
Australia

ken.mcanally@dsto.defence.gov.au

SUMMARY

Operators of military flight vehicles are often required to attend to more than one source of communications signal. Previous research has shown that the intelligibility of a speech message in a background of speech distractors is improved if the signals are presented using a 3-dimensional (3-d) audio display rather than the traditional diotic configuration. However, whether infrequent target messages (e.g., callsigns) are more reliably detected in a continuous monitoring task with high temporal uncertainty when using a 3-d audio display has not been examined. This study examined participants' ability to detect a target callsign and identify a colour/number combination associated with it while engaged in a 20-minute, simulated formation-flying task. Participants were required to monitor 5 communications channels in each of which messages were presented at random intervals. (On average, 2.4 channels were simultaneously active.) Thirty targets were presented over the 20-minute period. There were three audio display conditions: diotic, all channels in front, and channels separated in azimuth (3-d). Detection of target callsigns was significantly higher in the 3-d condition compared to the other conditions. Detections and false alarms were combined to calculate sensitivity and criterion measures using signal detection theory. Sensitivity was significantly higher in the 3-d condition compared with the other conditions, but there were no differences in criterion. Also, consistent with previous results, correct identification of the target number/colour combination was significantly higher in the 3-d condition compared with the other conditions.

1.0 INTRODUCTION

Previous research has shown that spatial separation of competing speech signals increases intelligibility compared with that associated with co-located talkers (see [1] for a review). A popular paradigm for conducting research on intelligibility under conditions of simultaneous communication is the coordinate response measure (CRM)[2]. A speech corpus comprising spoken messages of the form "Ready (callsign) go to (colour) (number) now" has been released to facilitate research using this paradigm [3]. In the most commonly adopted form of the paradigm, listeners are instructed to attend to the message addressed to a given callsign (e.g., Baron) and respond to the colour/number combination contained in that message, while ignoring distracting messages. This task involves a high degree of temporal certainty because the target and distractor messages are begun simultaneously in order to maximise informational masking. However, in most operational scenarios the timing of target messages is highly uncertain and operators have to monitor communications channels over long periods.

This study adopted a paradigm in which the timing of messages was random and the frequency of target messages was low, such that at any time there was high uncertainty about the presence of a target message. Participants were engaged in a simulated formation-flying task to ensure that their background workload was similar to that potentially present in operational environments.

McAnally, K.; Martin, R.; Doman, J.; Eberle, G.; Parker, S. (2005) Detection and Identification of Simultaneous Communications in a Simulated Flying Task. In *New Directions for Improving Audio Effectiveness* (pp. 31-1 – 31-6). Meeting Proceedings RTO-MP-HFM-123, Paper 31. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.aps.

2.0 METHODS

2.1 Participants

Six male volunteers aged from 26 to 46 years (mean age 34.5 years) recruited from the Defence Science and Technology Organisation participated in the study. Four had some previous flight experience, and all had experience with PC-based flight simulators. All had normal hearing and normal or corrected-to-normal vision.

2.2 Experimental Design

Each of three conditions of audio display (diotic, in-front, and 3-d) was evaluated during two 20-minute flight simulations for each participant. In the diotic condition, all stimuli were presented without any spatial processing. In the in-front condition, all stimuli were filtered by the individual's head-related transfer functions (HRTFs) for the location directly in front (i.e., 0 degrees of azimuth and elevation). In the 3-d condition, each of five channels of audio was spatialised to a different location (-90,0 -40,0 0,0 40,0 or 90,0 degrees of azimuth, elevation) using individualised HRTFs.

The order of conditions was balanced across participants using all possible orders. Conditions were also balanced within participants by presenting the second block of conditions in reversed order. Differences in performance across conditions were assessed using repeated-measures analyses of variance (ANOVAs) employing Greenhouse-Geisser corrections (where $df > 1$) and a criterion of significance of .05.

2.3 Flight task

Each participant completed six simulation runs, each around 20 minutes in duration. Participants were required to follow a lead aircraft (an F-111) that performed gentle (30-degree bank) turns and gentle climbs and dives. The measure of performance for the flight task was the proportion of time the participant was able to keep the lead aircraft within a 7.5-degree circular reticule centred on the ownship's x-axis. Workload was adjusted for each participant by changing the following distance (between 200 and 400 m) so that the lead aircraft was maintained in the reticule for up to about 95 % of the time of the simulation. One of six scripts specifying the track of the lead aircraft was randomly assigned to each simulation run.

The out-of-the-window display was generated and rendered at 60 Hz (Silicon Graphics, Onyx), and projected onto a 4-channel (i.e., left side, front, right side, above) cube display. Participants sat on a seat within the cube and flew using both the throttle and the stick. Head-down displays of flight instruments and a situation awareness display were presented.

2.4 Stimuli

Speech stimuli were sentences taken from a speech corpus [3]. Sentences were of the form "Ready Baron go to red three now", where the callsign (in this case Baron), the colour and the number varied across items.

Participants were required to monitor five channels of communication. The delay between messages in each channel was randomly selected from the range encompassing 1.3 to 4 seconds. This resulted in a talker density that varied from 0 to 5 simultaneous talkers with an average of 2.4. The target callsign to be monitored was "Baron". There were 30 target messages in each 20-minute flight simulation, with a minimum interval between target messages of 20 seconds. Thus, the frequency of target calls was low (only 2.5 % of all messages were targets) and temporal uncertainty was high. The talker, colour and number for the target calls were chosen randomly from those available in the corpus. All distractor messages were addressed to callsigns other than "Baron". The talker, callsign (other than "Baron"), colour and number for distractor messages were chosen at random.

To simulate the effects of transmission though a radio channel, stimuli were low-pass filtered at 4 kHz and clipped following amplification by 20 dB.

2.5 3-Dimensional Audio

Individualised HRTFs were recorded following the procedure previously described [4]. Briefly, miniature microphones were placed in the participant's ear canals and test signals were played from a loudspeaker that could be positioned in space around the participant. The recorded signals were then processed to derive the filtering effect of the head, torso and external ears (i.e., the HRTFs). The HRTFs were corrected for the transfer functions of the loudspeaker, microphones and headphones that were used later to present the audio stimuli. Each participant was able to accurately localise "virtual" stimuli in azimuth. The average azimuth error (adjusted for convergence at poles) was 8.5 degrees in the free field and 8.8 degrees using the 3-d audio display. This difference was not significant ($t[5] = 0.90$, $p = .41$).

2.6 Data Collection

A matrix of coloured buttons was placed close to the throttle for making responses to target communications signals. Response time was measured in ms from the initiation of a target call. The number of detected target calls, the number of false alarms, the proportion of correct colour/number response combinations and the reaction time to correct responses were calculated. The time window following a target message during which a response was accepted was 5 seconds. The rate of correct detections and false alarms were combined to calculate sensitivity (d') and response criterion using signal detection theory [5].

3.0 RESULTS

The proportion of detected target calls is shown for each display condition in figure 1a. The proportion of detected target calls differed significantly across display conditions ($F[1.8,9.0] = 6.9$, $p = .017$). Planned comparisons revealed that the proportion of detected calls was significantly higher in the 3-d condition compared with either the diotic or the in-front condition (diotic: $t[5] = 3.55$, $p = .016$; in-front: $t[5] = 3.50$, $p = .017$) and that the proportion of detected calls for the diotic and in-front conditions did not differ significantly ($t[5] = 0.12$, $p = .91$).

The average number of false alarms made in each 20-minute simulation is shown for each display condition in figure 1b. The number of false alarms was very low and did not differ significantly across display conditions ($F[1.4,6.8] = 0.49$, $p = .56$).

Sensitivity (d') and response criterion were calculated from the hit and false alarm rates using signal detection theory. Sensitivity is presented for each display condition in figure 1c. Statistical analysis revealed a significant effect of display condition ($F[1.1,5.7] = 9.23$, $p = .023$). Planned comparisons revealed that sensitivities in the diotic and in-front conditions did not differ significantly ($t[5] = 0.82$, $p = .45$) and that sensitivity in the 3-d condition was significantly higher than that in the diotic ($t[5] = 2.77$, $p = .039$) or in-front condition ($t[5] = 3.54$, $p = .017$).

The response criterion is shown for each display condition in figure 1d. Criterion did not differ significantly across display conditions ($F[1.3,6.6] = 0.30$, $p = .66$).

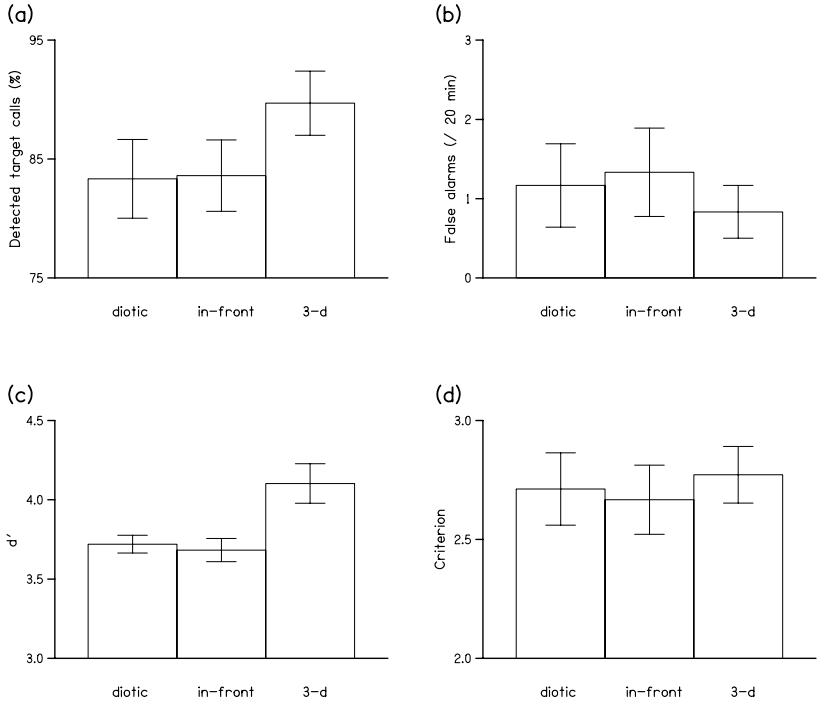

Figure 1. (a) Proportion of detected target messages, (b) number of false alarms, (c) sensitivity, and (d) response criterion for each audio display condition.

From the responses made to target calls, the proportion that correctly identified the colour/number combination was calculated. Participants made a similar proportion of correct responses in the diotic and in-front conditions, but more in the 3-d condition (figure 2a). Statistical analysis revealed a significant effect of display condition ($F[1.7,8.4] = 16.66$, $p = .001$). Planned comparisons revealed that the proportion of correct responses did not differ significantly between the diotic and in-front conditions ($t[5] = 0.91$, $p = .41$), but was significantly higher in the 3-d condition compared with either the diotic ($t[5] = 5.04$, $p = 0.004$) or the in-front condition ($t[5] = 5.82$, $p = .002$).

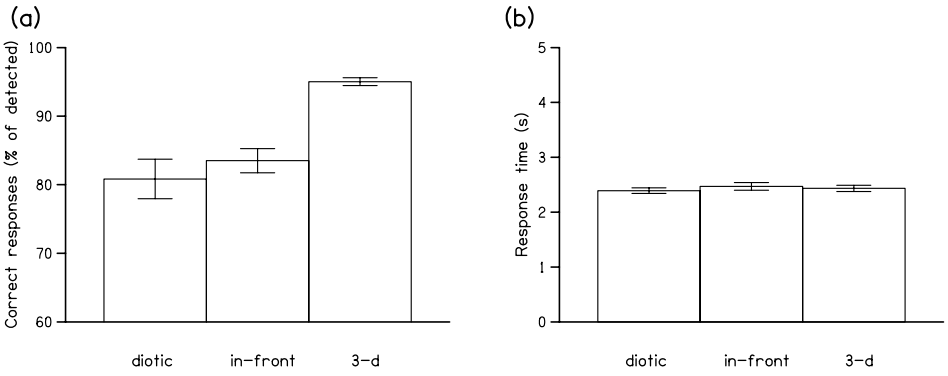

Figure 2. (a) Proportion of detected messages for which a correct response was made, and (b) reaction time for correct responses in each audio display condition.

Median response times were calculated for correct colour/number responses for each participant. Average median response times were close to 2.4 seconds (figure 2b) and did not differ significantly across display conditions ($F[1.7,8.8] = 3.78$, $p = .07$).

4.0 DISCUSSION

This study used a paradigm of high temporal uncertainty in the presence of a sustained background perceptual-motor task and found that the ability of listeners to detect a target callsign is improved when speech messages are distributed in azimuth. The study also found that the accuracy of responses was improved for spatially distributed messages, compared to messages presented via a traditional diotic display.

The detection of target messages has been studied previously using the CRM [6] and has been found to be significantly higher in spatialised audio conditions than in diotic conditions. However, in that study there was little temporal uncertainty because each trial was independent and contained messages that began simultaneously. In contrast, the present study was conducted under conditions of high temporal uncertainty.

An advantage for spatial presentation of simultaneous speech messages with respect to response accuracy has also been demonstrated in other previous studies employing the CRM under conditions of high temporal certainty. For an accurate response to be made in the CRM task, the callsign has to be identified, and features of the target voice (e.g., its location, timbre, or prosody) have to be used to link the target callsign and coordinates (i.e., colour and number). Previous work has shown that the predominant error involves responding to coordinates associated with a distracting talker [7], so the main difficulty encountered when listening to co-located talkers is linking callsigns and coordinates correctly rather than understanding the messages.

One previous study [8] has investigated the detection and identification of multiple communications in a simulated flying task. In that study, pilot's verbal responses to communications messages were more correct when communications were presented via a 3-d audio rather than a diotic display. However, performance in callsign detection and message identification were combined into a single response score so it is not clear how each of these tasks contributed to the result.

The task in the present study approximates those existing in operational environments more closely than do the tasks in many previous studies. Our results, therefore, provide a further demonstration of the potential for 3-d audio communications displays in operational environments.

[1] Bronkhorst, A.W. (2000) The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acustica-Acta Acustica 86, 117-128.

[2] Moore, T. (1981) Voice communication jamming research. In AGARD Conference Proceedings 331: Aural Communication in Aviation (pp. 2:1-2:6). AGARD: Neuilly-sur-Seine, France.

[3] Bolia, R.S., Nelson, W.T, Ericson, M.A. & Simpson, B.D. (2000) A speech corpus for multitalker communications research. Journal of the Acoustical Society of America 107, 1065-1066.

[4] 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.

[5] Green, D.M. & Swets, J.A. (1966) Signal detection theory and psychophysics. Wiley, New York.

[6] Nelson, W.T., Bolia, R.S., Ericson, M.A and McKinley, R.L. (1998) Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment. Proceedings of the 1998 IMAGE conference.

[7] Brungart, D.S. (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. Journal of the Acoustical Society of America 109, 1101-1109.

[8] Haas, E.C, Gainer, C., Wightman, D., Couch, M. & Shilling, R. (1997) Enhancing system safety with 3-d audio displays. Proceedings of the Human Factors and Ergonomics Society 41st annual meeting.