The effects of party line communication on flight task performance

Helen Hodgetts [1], Eric Farmer[2], Martin Joose[3], Fabrice Parmentier[4], Dirk Schaefer[5], Piet Hoogeboom[3], Mick van Gool[5], & Dylan Jones[1]
[1]School of Psychology, Cardiff University, UK
[2]Centre for Human Sciences, Qinetiq, UK
[3]NLR, Amsterdam, Netherlands
[4]Department of Psychology, University of Plymouth, UK
[5]Eurocontrol Experimental Centre, Bretigny, France

Abstract

Cognitive streaming is an approach to human information processing that regards short-term memory as a series of cognitive processes rather than stores. The approach is used as the theoretical basis for a series of experiments, both laboratory and simulator based, that assessed the disruptive effects of the party line on flight task performance. Initial laboratory work using a computer-based visual monitoring task, a communication task and a conflict detection task demonstrated that meaningful background speech was more disruptive to performance than meaningless reversed background speech or quiet. Moreover, the negative effect of the party line was further substantiated in a more realistic flight simulator study involving eight pilots: The party line condition resulted in a greater deviation from the touchdown point on the runway, and was associated with self reports of increased distraction and workload. Furthermore, an increase in flight checklist completion time was observed when background radio/telephony (R/T) was present, and also slightly more air traffic control (ATC) calls were missed or queried in this condition. The current theme of work extends laboratory findings on the 'irrelevant sound effect' to the aviation domain, and suggests that background sound in the party line not only adds to pilot workload but may also impair flight task mental activities.

Introduction

Background sound disrupts performance on a number of laboratory-based short-term memory tasks such as serial recall (e.g., Jones et al., 1992; 1995), text comprehension (Oswald et al., 2000) and proof reading (Jones et al., 1990). The goal of the current work is to extend these findings to the aviation domain by investigating whether the presence of background speech in the party line may be disruptive to short-term memory elements of flight task performance. Modular approaches to cognition such as multiple resource theory (Wickens, 1992) would suggest that background sound should only be disruptive to those tasks that draw upon the same capacity-limited resource (i.e., a concurrent auditory-verbal task, but

In D. de Waard, K.A. Brookhuis, R. van Egmond, and Th. Boersema (Eds.) (2005), *Human Factors in Design, Safety, and Management* (pp. 1 - 12). Maastricht, the Netherlands: Shaker Publishing.

not a visuo-spatial task). However, cognitive streaming theory offers a different account, proposing instead that interference arises not because of a conflict due to similar *content* but due to similar *processes* (e.g., seriation, or keeping track of order).

The party line refers to the open radio channel through which all aircraft in a given airspace communicate with air traffic control (ATC), a system that allows pilots to hear both their own clearances as well as those of the other aircraft. Although anecdotal evidence suggests that this may be useful for situation awareness (Pritchett & Hansman, 1993), the need to monitor background speech for relevant information would perhaps increase pilot workload in an already demanding multitasking environment. Moreover, laboratory studies indicate that the mere presence of extraneous background sound –even when unattended– impairs performance on a range of cognitive tasks.

The most commonly used task for examining the effect of irrelevant sound in the laboratory is the serial recall task: Participants are required to recall a list of visually or auditorily presented items in serial order, and to ignore any irrelevant sound presented during the trial. Background sound incurs a cost to performance irrespective of its intensity, and regardless of whether the irrelevant stream comprises speech or non-speech (e.g., a series of changing tones; Jones & Macken, 1993). A further finding is that disruption crosses representational domains: Irrelevant sound has been found to disrupt memory for both verbal information (e.g., letters/ digits) and spatial locations (Jones et al., 1995). Traditional approaches to human information processing are unable to explain this pattern of findings. For example, an explanation in terms of the phonological similarity between to-be-remembered items and the to-be-ignored sounds (e.g., Salamé & Baddeley, 1982) does not account for the disruption caused by tones, and multiple resource theory (Wickens, 1992) would have difficulty in accommodating the finding of cross-modal interference.

An alternative approach to the irrelevant sound effect is ‘cognitive streaming’ (e.g., Jones, 1999), by which short-term memory is regarded as a series of cognitive processes rather than stores. Incoming information, both primary task items and unattended/ irrelevant material, is processed in similar ways and is represented in streams containing information about the order of events. For example, background sound changing in pitch (either speech or tones) comprises a series of events whose order is automatically registered in streams, just as order information is processed consciously and deliberately for to-be-remembered items in a serial recall task. Interference arises as a result of conflict between two streams of order cues; those yielded from the irrelevant sound clash with those representing the to-be-remembered items.

In the laboratory, activities that draw heavily upon short-term memory for order are particularly vulnerable to disruption by irrelevant sound. In the field, this should encompass activities that (a) involve dealing with novel information, (b) require short-term response to unpredictable events, and (c) call upon reproduction of sequences (not just spoken sequences but other sequential actions also). Cognitive streaming theory would therefore predict that performance on many types of flight task would be impaired by irrelevant sound, and not just those aspects that involve the processing of verbal information.

Experiment 1

Experiment 1 aimed to identify the range of tasks that may be prone to disruption by irrelevant sound by testing a number of tasks performed in combination, and under three different sound conditions: quiet, speech, and meaningless reversed speech. For this purpose a new task battery was developed, namely the Aviation Multi-Tasking Environment (AMTE; Figure 1), comprising four main tasks designed to represent some of those that may be undertaken on the flight deck: audio-monitoring/data entry, tracking, conflict detection and visual monitoring. Based on cognitive streaming theory, it was expected that the data entry task would be particularly disrupted by irrelevant sound since it involves the processing of order information. Conversely, background speech was expected to have little or no effect on the tracking task that involved simple psychomotor control. The conflict detection and visual monitoring tasks were more exploratory because it is unclear to what extent unattended sound will interfere with visual or auditory vigilance tasks.

Method

Participants

Twelve undergraduate students at Cardiff University were firstly trained in each task of the AMTE in isolation, and then with all in combination.

Apparatus and materials

The task was presented on a *Windows 98* PC using the *AMTE* software written in *Visual Basic 6.0*, and using sound files recorded in *Sound Forge 4.5* (Sonic Foundry Inc.). Task-relevant sounds relating to the auditory-based elements of AMTE were recorded in a female voice for presentation through the left ear of headphones, whilst task-irrelevant sound was narrative speech recorded in a male voice and presented through the right ear of headphones. The same narrative was digitally edited using the 'reverse' function of Sound Forge for use as irrelevant reversed speech. The presentation of relevant and irrelevant messages in different ears and in dissimilar voices was implemented in order to minimise the possibility of masking, whereby one message is not heard due to purely perceptual interference by a concurrent sound.

The audio-monitoring/ data entry task required participants to monitor the relevant-speech stream for their call sign and for instructions to set one of four radio ports to a six-digit frequency. The tracking task involved four main flight instruments and participants were to maintain a target heading and altitude for the duration of the trial using a joystick. For the conflict detection task participants were required to monitor party line style messages in the task-relevant stream in order to detect potential conflicts with their own aircraft due to changes in heading or altitude. Finally, the visual monitoring task required participants to check that the level of the gauges remained between 45-55%, and to take appropriate action should they deviate by pressing the corresponding key F1-F4 on the keyboard.

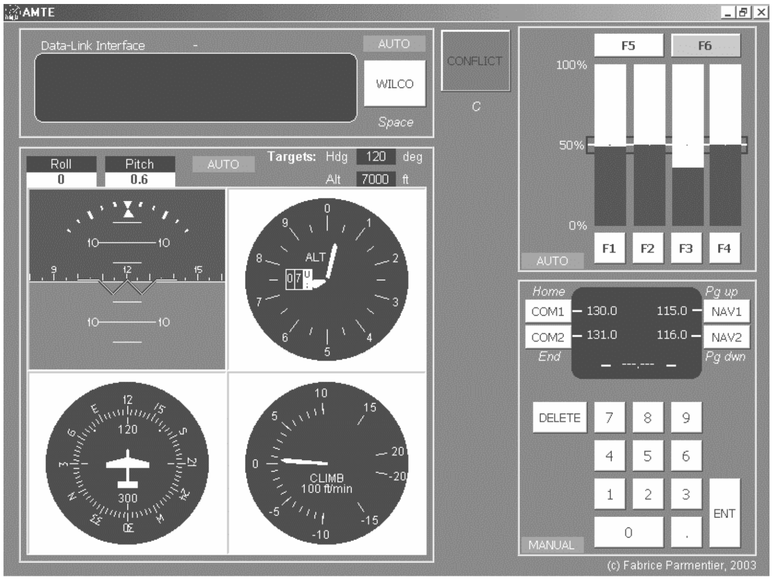

Figure 1. The Aviation Multi-tasking Environment.

Design

The four elements of the AMTE task battery were performed in combination. Each participant completed 12 trials, four in each of the three sound conditions: quiet, reversed speech and forward speech, the order of which was counterbalanced. Each trial lasted a total of 6 minutes and 40 seconds, and events were randomised within each trial.

Procedure

Participants first completed a session of training on the AMTE task. During the experimental trials, participants were told that in addition to the ‘party line’ messages in the left ear of their headphones, they might also hear irrelevant speech in their right ear. They were instructed to ignore this sound and to perform as well as possible on each of the elements of the AMTE task battery.

Results and discussion

Data entry task

Serial position data for the required six-digit radio frequencies are shown in Figure 2. A repeated measures analysis of variance (ANOVA) demonstrated a significant main effect of serial position ($F(5, 55) = 39.49$, $MSE = .004$, $p < .001$). The main effect of sound condition however, failed to reach significance ($F(2, 22) = 1.69$, $MSE = .077$, $p = .21$) although this could perhaps be attributed to a lack of power. Further analyses revealed quite a large effect size (Cohen's $d = .78$) but low power (power to detect medium effect (Cohen's d of 0.5) = .42), indicating that this non-significant result is likely due to the small sample size ($n = 12$).

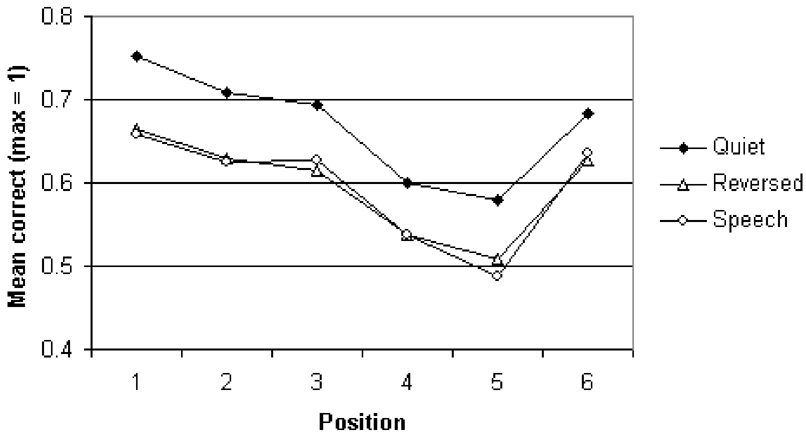

Figure 2. Serial recall performance in each of the three irrelevant sound conditions.

Figure 2 shows a clear trend for poorer performance in the two speech conditions relative to quiet. The finding that background sound –even meaningless reversed speech– disrupts performance on a task requiring seriation is in line with cognitive streaming theory. The meaning or phonological similarity of the irrelevant speech to the to-be-remembered items is not an important feature; rather, the fact that both streams yield a series of order cues is the key to disruption.

Tracking task

There was no difference between sound conditions on measures of deviation from heading or altitude. This finding is also in accordance with cognitive streaming theory because a simple psychomotor control task imposes a low cognitive load, and would not be expected to be affected by the processing of the order of the sounds in the irrelevant stream.

Conflict detection task

Repeated measures ANOVAs were conducted on hit rate, false alarms and reaction times. There was a trend for improved conflict detection in the quiet condition compared to the two speech conditions (Figure 3). This did not quite reach statistical significance ($F(2, 22) = 2.63$, $MSE = .09$, $p = .09$), but again this could perhaps be due to the small sample size: Further analyses indicated a large effect size (Cohen's d = 0.98) but low power (power to detect medium effect = 0.58). A Fisher's LSD post hoc analysis demonstrated that hit rate in the quiet condition was marginally better than in either of the two speech conditions (both p < .06).

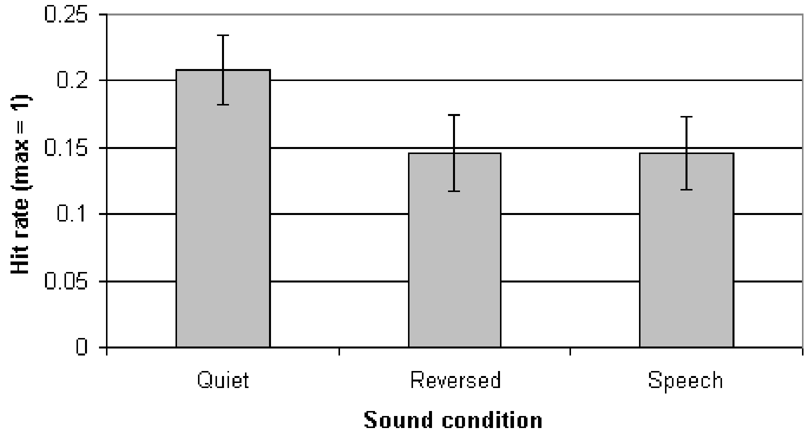

Figure 3. Mean hits in the conflict detection task. Error bars indicate standard error.

A marginally significant effect of sound condition was found in an analysis of false alarm data, ($F(2, 22) = 3.15$, MSE = 1.23, $p < .06$). A Fisher's LSD post hoc test indicated that more incorrect responses were made in the presence of forward speech than reversed speech ($p < .025$) or quiet ($p < .08$, Figure 4).

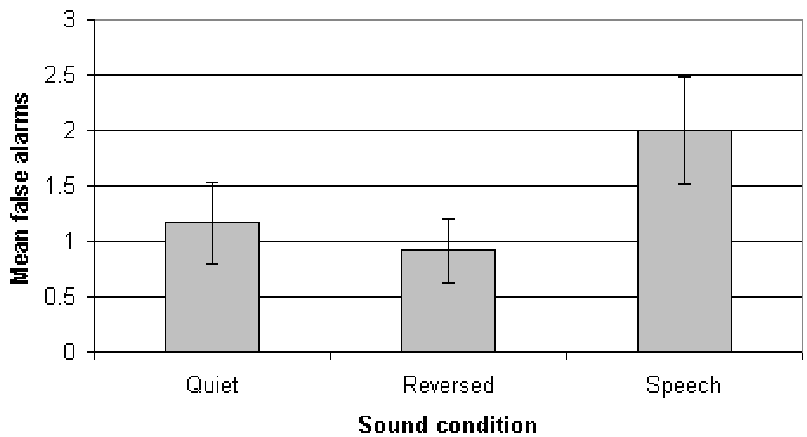

Figure 4. Mean false alarms in the conflict detection task. Error bars indicate standard error.

Reaction times were also recorded. These demonstrated a clear trend for slower conflict detection in the reversed speech condition than in quiet, and slower reaction times still in the forward speech condition. These findings show that unattended background sound can affect both accuracy and reaction times in a task of auditory vigilance.

Visual monitoring task

Both hit rate and false alarm data were collected, but neither demonstrated a significant effect of sound condition. Reaction times were also recorded (Figure 5) and demonstrated a significant effect of irrelevant sound, ($F(2, 22) = 6.57$, $MSE = 265396$, $p < .01$), with reaction times in the speech condition significantly longer than those in the quiet ($p < .001$) or reversed speech ($p < .04$) conditions.

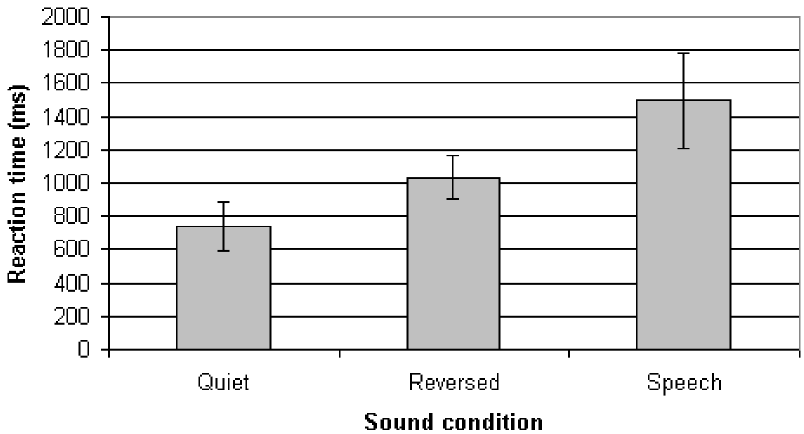

Figure 5. Mean reaction times in the visual monitoring task. Error bars indicate standard error

This result is interesting as it demonstrates that auditory-verbal irrelevant material can disrupt performance on a visual task, a finding that is difficult to accommodate within traditional modular theories of cognition (e.g., Wickens, 1992). Further research would be required to determine the exact processes involved in a visually based vigilance task and why they are susceptible to the effects of processing unattended sound.

Experiment 2

Experiment 2 examined the extent to which the effects of irrelevant sound observed in the laboratory could be demonstrated in a more realistic flying environment. Performance data, observational data, and subjective opinions were collected from four crews (eight pilots) tested under two sound conditions in a high fidelity fixed-base flight simulator. It was expected that the presence of the party line would be associated with self reports of increased frustration, distraction, pressure and workload. Moreover, it was expected that the potentially negative effects of background R/T would be evident in the objective measures of flight task performance.

Method

Participants

Eight Dutch pilots (paid volunteers) took part in the experiment. Captains had an average of around 4650 hours of flight experience and First Officers had approximately 160 hours.

Simulator

Crews were tested in the Generic Research Aircraft Cockpit Environment (GRACE) configured as a Boeing 747 with B747-400 enhanced EFIS displays and simulated Flight Management System. Pilots were instructed to disconnect the autopilot but used the flight director to help them to maintain heading and altitude.

Design

Each crew performed a total of six landings, three with the Captain as pilot flying (PF) and three with the First Officer as PF. In three of the flights the crew received 'minimal-required' R/T, and in the three remaining flights received additional 'party line' R/T. Conditions were counterbalanced between trials. The simulator collected performance data with respect to pitch angle deviation, deviation of flight director pitch angle commands, roll angle deviations, deviation of flight director roll angle commands, runway proximity (horizontal and vertical ground path approaches), and touchdown position on the runway.

Results and discussion

Questionnaire data

Post flight questionnaires (in English) were administered to both crew members immediately after each landing. An adapted version of the NASA TLX was used whereby pilots indicated on a scale of 0 – 100 how they rated each of eight items relating to the preceding flight task: mental and perceptual activity, perceived time pressure, success in accomplishing tasks, mental workload, frustration, distraction by R/T communications, perceived performance impairment by R/T, and perceived flight safety. Relative to the condition with minimal required R/T, the party line was judged to increase workload ($F(1, 7) = 62.94$, $MSE = 13.45$, $p < .01$), frustration ($F(1, 7) = 14.08$, $MSE = 104.03$, $p < .01$) and distraction ($F(1, 7) = 9.66$, $MSE = 483.14$, $p < .02$). A significant effect of party line was obtained on all items except 'success in accomplishing tasks'. Although pilots felt that task success was comparable across flights, the ratings of the other measures suggest that the crew may have had to work harder in the presence of party line to achieve this level of success.

Performance data

Data periods were analysed in which stable flight conditions occurred, but the presence of the party line had no significant effect in terms of pitch angle or roll angle deviation. When considering the landing period however, there was a main effect of background R/T on touchdown: The party line condition was associated with a higher mean standard deviation of the longitudinal position on the runway if compared to the no party line condition. This finding is surprising since almost no background R/T was present during the final approach, although perhaps the touchdown accuracy reflects a cumulative effect of background sound throughout the trial. Arguably, the landing period is the very point at which moment-to-moment correction and responsiveness to the aircraft environment is at its most pressing; one may therefore expect the effects of irrelevant sound to manifest at points when workload is greatest.

Observational data

Observations of ATC calls indicated that slightly more calls were missed, queried or incorrectly read back in the party line than the no party line condition (Table 1). However, the differences were small and there were insufficient samples to perform statistical analyses.

Table 1. Instances of ATC calls missed, queried or read back incorrectly, for Party line (P) and No Party line (NP) conditions

	Crew 1	Crew 2	Crew 3	Crew 4	Total
Calls missed (NP)	0	0	0	0	0
Calls missed (P)	0	1	0	0	1
Calls queried (NP)	1	1	0	1	3
Calls queried (P)	1	1	0	2	4
Incorrect read back (NP)	0	1	1	1	3
Incorrect read back (P)	1	3	0	1	5

Observations of checklist errors showed no apparent differences between the party line conditions: Omissions and repetitions of checklist items were perhaps more dependent upon *interruption* – the suspension and resumption of the checklist – rather than mere *distraction.* That is, a call to the specific aircraft requiring action and read back (in either R/T condition) was more disruptive than simply the presence of background R/T in the party line condition (Table 2). This is consistent with findings that interruptions are disruptive to flight deck performance (Latorella, 1999).

Table 2. Instances of checklist items omitted, repeated or resumed correctly at the next item for those occasions when an ATC call to the own aircraft caused checklist performance to be interrupted (R/T conditions combined)

	Crew 1	Crew 2	Crew 3	Crew 4	Total
Resumed at next item	2	0	1	0	3
Item omitted	1	1	0	1	3
Item repeated	2	0	0	0	2

The time taken to complete each checklist (after excluding time spent dealing with actual interruptions) was recorded and subjected to a 2 (party line condition) x 3 (checklist) repeated measures ANOVA. Although there was no main effect of party line or of checklist, there was a significant interaction, ($F(2, 4) = 11.92$, $MSE = 0.01$, $p < .02$). Completion of the approach and landing checklists was unaffected by background R/T, but completion of the descent checklist took significantly longer in the party line than the no party line condition.

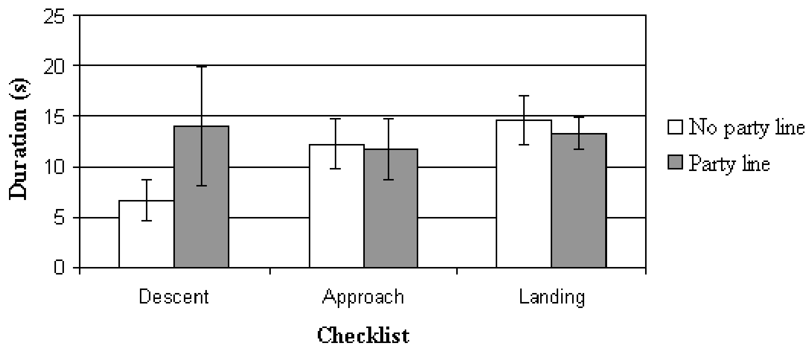

Figure 6. Checklist duration according to party line condition.

This may be because the items on the descent checklist (terrain clearance and approach preparation) involved an element of discussion between crew members, whereas the other two lists involved just quick checks and so were less susceptible to disruption from the background sound.

General Discussion

Three of the four tasks employed in AMTE showed some degree of disruption when performed concurrently and in the presence of extraneous sound. The data entry task showed typical effects of irrelevant sound on serial recall: Performance was depressed in conditions in which irrelevant material was presented, echoing the findings of stringent laboratory-based tasks (e.g., Colle & Welsh, 1976; Salamé & Baddeley, 1982). Serial recall in the reversed speech and forward speech conditions did not differ, suggesting that the physical properties of the irrelevant sounds are more critical to the degree of disruption than any meaning they may contain. In both the conflict detection task and the visual monitoring task, reactions to events slowed in the presence of irrelevant sound. This may be attributed to the nature of multi-tasking environments in which participants must divide their attention such that each task is performed to a modest level. It is likely that this process requires some form of serial order in shifting from one task to another. The conflict detection and visual monitoring tasks each required the monitoring of sources of information, one auditory and one visual. Although these tasks may not typically be disturbed by irrelevant sound when presented in isolation, the process of shifting attention from one to the other, as well as to the tracking and communication tasks, clearly utilises processes that are liable to disruption by irrelevant sound.

The findings of the simulator study were mixed. The systematic subjective reports are unequivocal in showing that the aircrew reported that the effects of party line were negative. Ratings relating to workload and distraction were higher in the presence of irrelevant R/T messages. Although the cognitive streaming model does not make firm predictions about subjective response to sound, these results are consistent with the notion that an irrelevant stream of information is difficult to ignore. In terms of the performance data, an effect of party line was obtained only in the final stage of flight. Perhaps the effects of workload and distraction were partly cumulative, so that the predicted effects were found only at the point when the demands of the task were greatest. Although differences in flight task performance measures were not observed for the main period of flight, it is interesting that a simple measure of checklist completion time showed the expected decrement in the presence of the party line. The effects of background sound in the simulator study may have been constrained by the small sample size, however, given that a few interesting findings did emerge, this is an avenue of research that warrants further study. Future research could speak more directly to the question of whether the replacement of the party line with digital data link technology may not only reduce workload, but may also circumvent the harmful effects that the party line may otherwise have on flight task mental activities.

Acknowledgement

These experiments were part of the Cognitive Streaming project, funded by EUROCONTROL in the context of the 'CARE Innovative Action'. More information on the Cognitive Streaming project can be found at www.eurocontrol.int/care/innovative/projects2002/cs.

References

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour, 15*, 17-32.

Hoogeboom, P., Hanson, E., Joosse, M., Hodgetts, H., Jones, D., Salmoni, A., Farmer, E., & Straussberger, S. (2004). *Cognitive streaming project, report on work package 4: 'Real World' effects.* Care Innovative Action, Eurocontrol CARE-IA-CS-NLR-WP4-D4-02-C. Available: www.eurocontrol.int/care/innovative/projects2002/cs/

Jones, D.M. (1999). The cognitive psychology of auditory distraction: The 1997 BPS Broadbent Lecture. *British Journal of Psychology, 90*, 167-187.

Jones, D.M., Farrand, P.A., Stuart, G.P., & Morris, N. (1995). The functional equivalence of verbal and spatial information in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1008-1018.

Jones, D.M., & Macken, W.J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 369-381.

Jones, D.M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44A*, 645-669.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proof-reading by irrelevant speech: effects of attention, arousal or memory? *Applied Cognitive Psychology, 4,* 89-108.

Latorella K.A. (1999) *Investigating Interruptions: Implications for Flightdeck Performance,* NASA/TM-1999-209707. Washington: National Aviation and Space Administration (also published in 1996 as Doctoral Dissertation, State University of New York at Buffalo).

Oswald, C. J. P., Tremblay, S., & Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory, 8,* 345-350.

Pritchett, A., & Hansman, R.J. (1993). Preliminary analysis of pilot rankings of ‘party line’ information importance. In R.S. Jensen and D. Neumeister (Eds.), *Seventh International Symposium on Aviation Psychology*. LOCATION: PUBLISHER

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology, 41A,* 107-122.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.