Safety Science 149 (2022) 105686

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Pilot errors: Communication comes last

Alexander Thorpe [a], Dominique Estival [b,*], Brett Molesworth [c], Ami Eidels [a]

[a] University of Newcastle, University Dr, Callaghan, NSW 2308, Australia
[b] MARCS, Western Sydney University, Penrith, NSW 2751, Australia
[c] School of Aviation, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Keywords:
Prioritisation
Dual-task effect
Task demands
Error
Workload
Safety

ABSTRACT

This study builds on previous research, which established that in flight simulator experiments the communication performance of pilots was impaired under certain applied conditions. The flight simulator data recording relating to the actions of the pilots were examined to determine the impact of the factors affecting pilots' communication (increased workload, increased demand on memory and, for some groups, increased ATC speech rate) on their flying performance. Using heading error as the dependent variable, no significant effects were found even for flights where pilots committed the most communication errors. Pilots are taught to prioritise tasks in order of operational safety importance, as per the adage "aviate, navigate, communicate". Thus, these results are encouraging as they show that the order of operational importance is adhered to, and that flying performance is maintained even when communication is affected.

1. Introduction

In high-stakes environments, safety relies on processes that are employed to standardise behaviour and provide predictable outcomes. These processes include training to specific competencies and skill levels. In addition, during training, certain behaviours are reinforced until they are internalised and become routine. Such training is crucial in directing attention towards the task at hand and prioritising tasks. In a number of professions, however, it is common to for individuals to perform multiple tasks at once. The simultaneous performance of two tasks commonly leads to performance deficits in these tasks (Watanabe and Funahashi, 2014). In safety critical professions such as aviation where simultaneous dual-task performance is common, any performance deficit can lead to an adverse safety outcome. The aim of this research extends Molesworth and Estival's (2015), Estival and Molesworth (2016, 2020) research which uncovered the effect of known external factors on pilot miscommunication, to determine whether pilots are able to effectively prioritise their attention and prevent the known factors that impact communication from affecting flying performance.

The dual-task effect is thought to provide evidence of the capacity limitations of working memory (Watanabe and Funahashi, 2014). According to Kahneman (1973), working memory contains a finite number of resources. The available resources limit the amount of information that can be processed at any given time. Once this is exceeded, deficits in performance are likely. Completing multiple tasks simultaneously increases the likelihood of exceeding the resource limits of working memory. Baddeley's (2003) model of working memory predicts differing effects based on the nature of the information being presented. In this model, visuospatial information is maintained and manipulated in the visuospatial sketchpad, while auditory information is maintained and manipulated in the phonological loop. Each of these components of the system is theorised to draw on separate resources, resulting in greater dual-task interference when both tasks are of the same *modality*, such as two auditory stimuli, compared to one auditory and one visuospatial stimulus. Alternatively, multitasking may cause decreased performance due to the increased demands on the user's attention – drawing attention away from the process of refreshing the visuospatial sketchpad and phonological loop can cause that information to decay, resulting in information loss and lower performance (Rhodes and Cowan, 2018). The interaction between Baddeley's components of working memory and the participant's directed attention, or inattention, could therefore result in dual-task interference.

Wickens' (2008) Multiple Resource Theory provides an alternative account for dual-task interference. As in Baddeley's three-component model, Multiple Resource Theory theorises separate processing domains for visual and auditory stimuli, which draw on separate resources. Some dual-task inference is predicted under both models regardless of stimulus modality, as executive functions such as planning and

* Corresponding author at: MARCS Institute, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia.
E-mail address: d.estival@westernsydney.edu.au (D. Estival).

https://doi.org/10.1016/j.ssci.2022.105686
Received 25 August 2021; Received in revised form 30 November 2021; Accepted 12 January 2022
Available online 24 January 2022

executing responses are theorised to occur in a central component with a single pool of resources (Helton and Russell, 2011, 2013). Multimodal multitasking will therefore require more resources than completing a single task. This does not necessarily result in a noticeable decrement in performance provided the resource requirement by each individual process is modest enough – a limited-capacity process can perform efficiently as overall capacity requirements are not exceeded (Kantowitz, 1985).

One coping mechanism to overcome the deleterious effects of dual-task interference is to shed or actively ignore information (Lachman et al., 2015). The effectiveness of this strategy is contingent on the importance of the information being prioritised. In aviation, this prioritisation technique is actively taught to pilots. Specifically, pilots are taught to prioritise their attention, as per the adage "aviate, navigate, communicate" (e.g., FAA, 2018; CASA, 2019). Attention prioritisation to manage the dual-task effect is not new or unique to aviation. Prioritisation techniques have been employed and tested in the laboratory with both young and old adults, and patients with and without a diagnosed medical condition to lower the safety risk when walking. Such techniques focus on instructions pertaining to the priority activity, in this case walking pattern – gait technique. Attention prioritisation often leads to a reduction in gait speed in the dual-task condition (walking and cognitive task) compared to control (no attention prioritisation; Verghese et al., 2007; Yogev-Seligmann et al., 2010; Yogev-Seligmann et al., 2012). In aviation, similar results have been found in reducing pilot altitude deviations during an Instrument Flight Rules (IFR) flight in the face of competing, less serious prospective memory tasks. Pilots who were taught a task management procedure called APE (Assess, Prioritize, Execute) showed greater improvements pre-post intervention on both the altitude deviation and memory tasks (Bishara and Funk, 2002).

There remains a question over whether the prioritisation behaviour described above would occur without training. Natural task prioritisation is apparent in some tasks, but not others. In the example of gait technique given above, directed prioritisation was required to adjust gait speed. This is also the case with swimming (Stets et al., 2019), whereby swimming speed is negatively affected by increased cognitive load. However, there is some evidence that climbing speed is unaffected by increased cognitive load (Darling and Helton, 2014; Blakely et al., 2021). This lack of interference was found despite climbing having a deleterious effect on cognitive tasks, suggesting the interference only ran one way. One explanation for this effect is the level of risk involved in the activity. Whereas walking and swimming are relatively low-risk activities with little penalty for slow(er) performance, climbing is higher-risk, with poor performance carrying potentially catastrophic consequences (Blakely et al., 2021). There is nevertheless some evidence that climbing can be affected by increased cognitive load, though this may be an indirect effect – if cognitive demands incur physical fatigue, this fatigue could be the cause of slower climbing (Woodham et al., 2016). These findings have mixed predictions for aviation. Flying is certainly a high-stakes behaviour, so natural prioritisation of flying tasks may take place, though the low physical demand and relative detachment of the pilot from the physical process of flight may imply that piloting will not be naturally prioritised. In any case, there is evidence that optimal prioritisation can take place even under high cognitive workload (Wickens et al., 2003), so training can prepare pilots to prioritise piloting behaviour over distracting stimuli.

While the attention prioritisation research highlights the effectiveness of targeted and simple instructions to effectively prioritise tasks, how such a technique applies in aviation where pilots need to balance multiple tasks such as flying an aircraft, navigating though three-dimensional space, monitoring and responding to communications, performing calculations, maintaining situation awareness, is still open to research. A number of important questions applicable to attention prioritisation in aviation remain unanswered, and possibly the most prominent relates to the effectiveness of such a technique on the known factors that affect pilot performance.

Molesworth and Estival (2015), Estival and Molesworth (2016, 2020) investigated the communication performance of pilots in flight simulator experiments providing conditions known or expected to affect the communication performance of pilots: increased ATC speech rate (Taylor et al., 1994; Burki-Cohen, 1995; Morrow and Rodvold, 1998; EUROCONTROL, 2006b; Said, 2011; EUROCONTROL, 2014), increased information density (Cardosi, 1993; Barshi and Farris, 2013), increased pilot workload (Morrow et al., 1993) and increased radio frequency congestion (Morrow and Rodvold, 1993; Orlady and Orlady, 1999). They found that increased ATC speech rate, increased information density in ATC instructions, and especially increased pilot workload (completing a secondary task such as revising fuel calculation), all adversely affected pilots' communication performance as measured by the number and nature of the communication errors committed by the pilots while flying.

How these factors affect the primary task of flying is the focus of this study. Specifically, this paper reports on the analysis of the flight simulator data which recorded the actions of the pilots during the simulated flights reported in Molesworth and Estival's (2015) research. The flight simulator recorded data pertaining to Altitude, Indicated Airspeed, Vertical speed, Pitch and Roll angle, and Heading. Using Heading error as the dependent variable, analyses were conducted to determine whether the factors which had been shown to affect pilots' communication also affected their flying performance.

With the background that all pilots are taught to prioritise tasks in order of operational safety importance, as per the adage "aviate, navigate, communicate", and the effect of the known factors on pilot communication performance as identified by Molesworth and Estival (2015), the current study sought to answer the following two research questions.

Research questions

Are pilots able to effectively prioritise the tasks of aviating, navigating, and communicating?

Does the type of task demand (increased ATC speech rate, increased ATC information density, or increased pilot workload) affect pilots' ability to manage the flying task?

2. Method

2.1. Participants

The participants in the study were all pilots recruited from local flight training schools, performing in a medium fidelity flight simulator. A total of 17 pilots (one female), eight of whom were native English speakers (NES) and nine non-native English speakers (NNES)[1] volunteered for the research. The average age of the participants was 30.82 (SD = 13.97) years. All participants completed the set of eight flights, in an average of two hours. Before the experiments, the pilots filled an informed consent form and a demographic questionnaire. Pilot experience ranged from 42 to 3,500 h, with seven pilots holding a Private Pilot Licence (PPL) or lower qualification, and 10 pilots a Commercial Pilot Licence (CPL) or higher.

2.2. Design

The experimental design comprised a 2 × 4 repeated measures design. The first within-group factor, Task Load contained two levels (baseline vs. manipulation), while the second within-group factor, Flight Scenario contained four levels (Speech Rate, Information Density, Pilot

[1] The native language of the non-native English speakers (English as a second language - ESL) included: Cantonese (4), Chinese (1), Malayalam (1), Italian (1), Danish (1), Russian (1). On average, the NNES pilots reported to have spoken English for 17.11 (SD = 11.96; range 2–35) years.

Workload and Radio Congestion). The data analysis however, comprised a 2 × 3 design since the Radio Congestion scenario did not feature because the pilots did not have to respond to the manipulated operation condition. In the increased Radio Congestion flight, none of the additional transmissions were directed to the pilot in command (thus, no heading instructions).

The baseline flight in each 'Flight Scenario' reflected operational conditions (i.e., Task Load) that were deemed to be normal. This flight was called Flight A. The manipulation flight, referred to as Flight B, varied from the baseline flight based on the operational condition being manipulated in accordance with the flight scenario. In the Speech Rate scenario, Flight B contained ATC radio transmissions at a speech rate double that recommended by ICAO and used in Flight A. In the Information Density scenario, Flight A contained fewer than three items per ATC transmission, while Flight B contained four or more items per ATC transmission. In the Pilot Workload scenario, Flight B involved pilots having to calculate the fuel required for a diversion, in addition to communicating and flying the aircraft, thus increasing cognitive workload.

The dependent variable was heading error. Heading error was the absolute difference between observed heading and prescribed heading (as instructed by ATC) for each section of flight.

Sections of the flight (referred to as time series) were identified based on radio instructions – a section would begin when the ATC instruction to change heading was given, and end either after the next heading instruction, two further non-heading radio instructions, or the end of the flight, whichever came first. This allowed the collection of multiple observations per run. These sections were further divided into *call* and *response* segments – the time during which the instructions were given was considered the *call* segment, while the time after the instructions ended and the pilot was required to respond verbally and behaviourally was considered the *response* segment.

The time series was divided further as depicted in Fig. 1, which depicts the flight path from an example trial, illustrating where each of the three segments fell within a trial. The first response segment, hereafter 'R_1'', was taken from the beginning of the response, and the second response segment, hereafter 'R_2'', was taken from the end of the response. Each of these segments was of equal length to the *call* segment, hereafter 'C', to ensure equal observations in each segment. This factor is hereafter referred to as *call segment*, with three levels ($C/R_1/R_2$). This division is depicted in Fig. 2, which shows the translation from the flight path in space to a timeline of the example time series section. The purpose of this division was to allow pilots sufficient time to reach their expected heading, allowing for fair comparison between the C segment and the response. Responses during the R_1 segment represented the point in flight when pilots were most likely to be affected by ATC instruction calls and the need to verbally respond to those instructions, whereas responses during the R_2 segment represented the point when they would be least likely to be affected by ATC calls. It was, therefore, expected that heading error would be highest in the R_1 segment. The critical comparison was between C and R_2 segments, as the latter segment was the point in the time series when pilots had been given sufficient time to make a safe turn to the expected heading. If the presence of ATC calls affected flying performance, it would therefore be apparent in the comparison of the C and R_2 segments.

2.3. Materials

The laboratory equipment comprised: X-Plane 6.21 featuring a Cessna 172 aircraft (with call sign "ABC''), a Personal Aviation Training Devices (PCATD) with one 21-inch flat screen monitor, Elite rudder pedals, and two additional computers: one to play the audio stimuli through an aviation headset and one to record the pilot's verbal responses with the Audacity software. In order to replicate the applied environment as much as possible, reproduced aircraft noise of a Cessna 172 during cruise at 65 dBA was played throughout each flight.

The test documentation comprised: an information sheet, a consent form, and a demographics questionnaire asking participants to provide their age, sex, native language, number of years speaking English and flying experience.

2.4. Procedure

The four flight pairings (eight flights in total) were presented in a counterbalanced order as per a 4 × 4 Latin square design (as opposed to a balanced Latin square design, because of the undesirable adjacency which a balanced Latin Square would have given to the two flights in each pair). A total of 17 pilots completed the task, for a total number of 136 flights. The data recorded for each flight was of two types, audio for the pilots' verbal responses and flight simulator data for the pilot

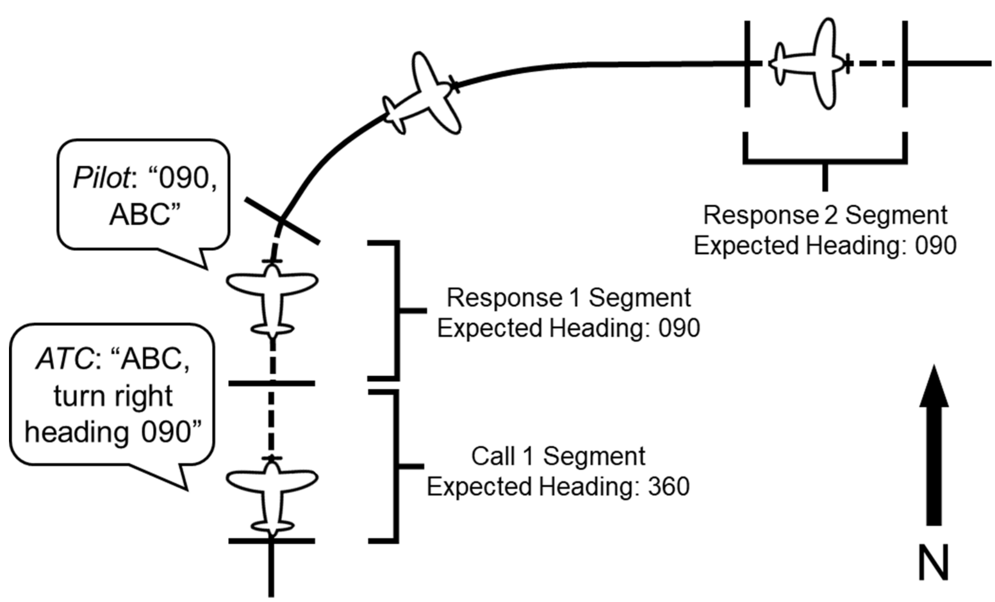

Fig. 1. Flight path of an example time series section. In this example, pilots begin at heading 360 and receive instructions to change heading to 090 degrees ("ABC, turn right heading 090") during the call segment, with the two response segments occurring at the beginning and end of the remainder of the time series. The duration of each of the three segments (Call 1, Response 1, Response 2) is equal.

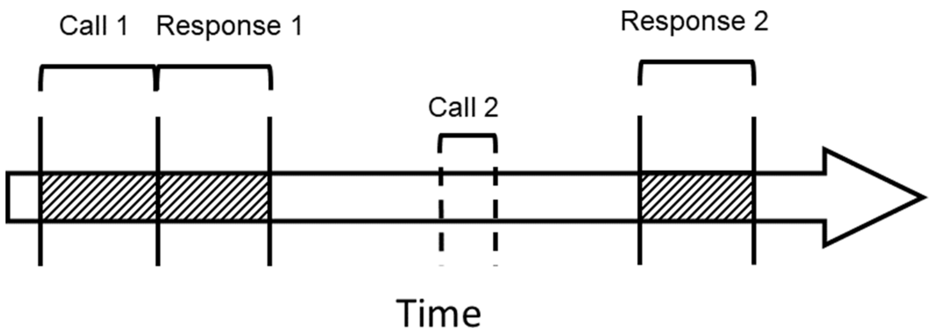

Fig. 2. Timeline of an example time series section where Call 1 provided heading instructions and where Call 2 provided no heading instruction. Shaded areas represent the data used for the current analysis. Responses 1 and 2 represent the beginning and end of the response time series, respectively.

actions. The audio data and the flight simulator data were aligned for all the flights.

2.5. Data analysis

The dependent variable, heading error, was analysed using a 3 × 2 × 3 repeated measures ANOVA, with within-subjects factors of Call Segment (C/R_1/R_2), Task Load (Low/High), and Flight Scenario (1/2/3). As explained in Section 2.2, only the first three Flight Scenarios were used in the current analysis. Prior to all analyses, violations of the ANOVA test assumptions were checked, and alpha was set at 0.05. Alpha remained unadjusted in accordance with Rothman (1990) and Armstrong (2014), when subsequent post hoc tests were necessary in the pursuit of the interpretation of a single test, as opposed to across tests. According to Rothman and Armstrong, such a method protects against type ii error, as well as controls for type i error since no data was repeatedly used across tests. Effect sizes are presented in cases of significant differences as Cohen's *d* for *t*-tests and eta-squared for ANOVAs. Effect sizes were interpreted based on Cohen's (1988) conventions.

Bayesian mixed ANOVAs were used in cases of non-significant results. These tests have the advantage of being able to evaluate the evidence in favour of a null hypothesis, unlike conventional null hypothesis testing. In these cases, Bayes Factors representing the strength of an effect are presented. There are two types of Bayes Factors. Bayes Factors labelled BF_{10} represent the strength of evidence in favour of an alternative hypothesis compared to a null hypothesis, whereas those labelled $BF_{\text{Inclusion}}$ represent the strength of including a specific factor in an explanatory model of the observed data. In both cases, Bayes Factors were interpreted according to the conventions in Jeffreys (1961) and presented in Table 1.

Prior to the mixed repeated analysis, it was important to check whether differences in flying performance were evident between the two language background groups (NES vs. NNES) that featured in Molesworth and Estival (2015), Estival and Molesworth (2016, 2020). Performance differences, if any, would be expected in the manipulation flight (Flight B), therefore three separate *t* tests (one for each flight scenario) were performed. The results of the three *t* tests, failed to reveal a significant difference based on heading error (largest *t*, $t(13) = 1.19$, $p = .25$ – Flight 1). As a result, the two groups were collapsed.

Table 1
Interpretation of Bayes Factors, based on the conventions in Jeffreys (1961).

Bayes Factor BF_{10}	Interpretation
>100	Extreme evidence for H_1
30–100	Very strong evidence for H_1
10–30	Strong evidence for H_1
3–10	Moderate evidence for H_1
1–3	Anecdotal evidence for H_1
1	No evidence
1/3–1	Anecdotal evidence for H_0
1/10–1/3	Moderate evidence for H_0
1/30–1/10	Strong evidence for H_0
1/100–1/30	Very strong evidence for H_0
<1/100	Extreme evidence for H_0

3. Results

Prior to analysing the results in relation to the two research questions, a manipulation check was conducted to determine whether differences existed between call segments. As demonstrated in the Design section above, differences in call segment between R1 and R2 were expected, since in R1 pilots would be initiating their turn based on the ATC instructions, and at R2, they should be on the new required heading or near approaching this new heading. It is also possible that differences exist between C and R1, based on pilots' early initiation of the new heading. As can be seen in Table 2, the results revealed a medium-sized main effect for Call Segment, $F(2, 20) = 44.48$, $p < .001$, $\eta 2 = 0.06$.

The post hoc analysis for Call Segment revealed that R_1 was statistically different to both the C segment, $t(10) = 9.10$, $p < .001$, $d = 2.75$, and the R_2 segment, $t(10) = 6.69$, $p < .001$, $d = 2.02$. These represent extremely large effect sizes. Mean heading error in the C segment was also significantly different to error in the R_2 segment, $t(10) = 2.41$, $p = .026$, $d = 0.73$. This represents a medium effect size, but it is much smaller than the effects seen in comparisons of the R_1 segment to the other segments. This pattern of results suggests pilots were following ATC instructions and that, given sufficient time, they could reach the appropriate heading.

Having determined that pilots were indeed following the heading instructions given by ATC, it was important to determine whether pilots were able to effectively prioritise tasks whilst flying (Research Question 1). If pilots were unable to prioritise tasks (flying and communicating in Scenarios 1 and 2; flying and performing calculations in Scenario 3), a main effect for Task Load (Flight A vs. Flight B) would be expected. As can be seen in Table 2, and by the separation between lines in Fig. 3, no main effect was evident. Bayesian analysis indicated moderate evidence

Table 2
Test statistics from 3 × 2 × 3 repeated measures ANOVA. Significant results are marked with asterisks.

Factor	Mean Square	Degrees of Freedom	*F*	*p*
Task Load	38.57	1, 20	0.03	0.876
Flight	727.72	2, 20	0.27	0.763
Call Segment	4588.94	2, 20	44.48	<0.001***
Task Load * Flight	2272.55	2, 20	1.35	0.281
Task Load * Call Segment	40.17	2, 20	0.94	0.406
Flight * Call Segment	245.83	4, 40	2.61	.05[a]
Task Load * Flight * Call Segment	29.69	4, 40	0.57	0.689

[a] This result was ambiguous, as it approached significance (p < .05), so was examined further below.

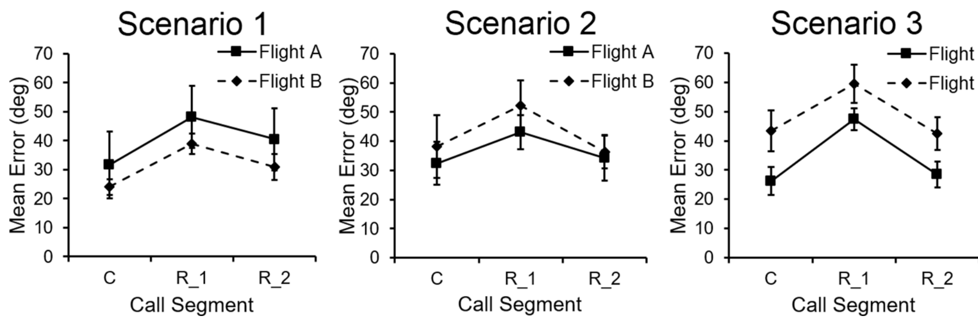

Fig. 3. Mean heading error across levels of call segment and task load. Each panel represents data from the two flights in each scenario.

against including this factor in an explanatory model of the data, $BF_{Inclusion} = 0.15$. This result indicates that pilots were able to prioritise tasks.

The next analysis sought to determine how the task demand (i.e., increased ATC speech rate, increased ATC information density, and increased pilot workload) affected pilots' ability to manage the flying task (Research Question 2). If these different demands affected pilots differently, an interaction between Task Load (Flight A vs Flight B) and Flight Scenario (speech rate, information density, pilot workload) would be expected. As can be seen in Table 2, the results failed to reveal a significant effect. Anecdotal evidence (as defined in Table 1) for an interaction was found, $BF_{Inclusion} = 1.97$, though this evidence lacks the strength to reject the null hypothesis.

The Bayesian analyses above found evidence against a difference in flying performance based on Task Load (Research Question 1) or on Flight Scenario (Research Question 2). Indeed, only models containing Call Segment as a factor explained the data better than the null model. It should be noted that, unlike frequentist ANOVA, these tests provide positive evidence for the null hypotheses. In relation to the research questions, they indicate that pilots can effectively prioritise the task of flying over communicating and performing calculations, and that the type of task demand does not affect this ability.

4. Discussion

The main aim of this study was to determine whether pilots can effectively prioritise their attention and thus prevent factors known to impact communication from affecting their flying performance. The results revealed that indeed pilots were able to effectively prioritise their attention, and importantly that the factors which have been shown to affect communication such as increased ATC speech rate, information density, and workload, failed to interfere with their prioritising the task of flying. Given the findings of Molesworth and Estival (2015) and Estival and Molesworth (2020) that these sources of load interfered with verbal responses for the same cohort of participants on the same tasks, this lack of interference suggests that the pilots focused on piloting rather than verbal tasks.

The primary finding of this analysis was the expected difference between the R_1 segment and the other segments within the time series, with a larger deviation from the required flight trajectory on R_1 relative to other segments. The simplest explanation for this difference is that the R_1 segment represents a time before pilots could have responded to the instructions given in the call. The more relevant comparison was between the C and R_2 segments, which would have detected a difference due to the demands of the verbal task. The lack of significant difference here, along with the lack of differences across Task Load conditions or Flight Scenarios, have two possible explanations. It might be that the conditions were truly not different. However, the difference in verbal response accuracy across Task Load conditions in Scenarios 2 and 3 (Molesworth and Estival, 2015; Estival and Molesworth, 2020) conflicts with this explanation. Another explanation might be that the behavioural measures are not sufficiently sensitive to detect true differences across the conditions. Our primary finding of a difference between the R_1 segment and the other call segments, however, suggests it is indeed sensitive to different patterns of behaviour. Another explanation still might be that the current analysis simply lacked sufficient power – the highest number of pilots in a single condition was fifteen, while the lowest was twelve. This precluded the between-subjects comparison between NES and NNES pilots. With more observations from a larger cohort of pilots, this factor could have been included in the analysis, which may have affected our findings if native language interacted with the other factors.

4.1. Theoretical and applied implications

Our findings can be understood in terms of Baddeley's (2003) model of working memory. The flight scenario with the highest level of dual-task interference, Flight 3, was also the only scenario with two visuospatial tasks. The other scenarios involved auditory/verbal task (i.e., listen to ATC transmission and respond) and one visuospatial task (i.e., piloting), though the lack of statistical significance between the levels of task demand is not predicted by this model. The non-significant level of interference may be explained by Multiple Resource Theory, which predicts that multimodal multitasking incurs less dual-task interference than single-mode multitasking. As all scenarios featured auditory secondary tasks alongside the visual aviation task, they may have caused less interference than a visual secondary task. When taken together with Molesworth and Estival (2015)'s original analysis of this data, the current study's findings suggest individuals can effectively prioritise competing tasks – performance on the verbal tasks was affected by increased task demand, but flight performance was not similarly affected. These findings mirror those in other research laboratories examining simple instructions articulating which task an individual should prioritise to achieve a desired level of performance (Bishara and Funk, 2002; Verghese et al., 2007; Yogev-Seligmann et al., 2010; Yogev-Seligmann et al., 2012). The results also illustrate that task prioritisation can overcome the debilitating effect of high cognitive load, which is known to affect other aspects of performance (e.g., information density and increased workload, which affect communication performance while flying (Molesworth and Estival, 2015)).

From an applied perspective, these findings reflect positively on the current emphasis of task prioritisation during flight training. The findings do, however, raise questions about the causes of some crashes that result from failures in task prioritisations, such as Eastern Air Lines Flight 401 and Pakistan International Airlines Flight 8303. For both flights, the accident investigation body found pilot distraction during the flight to be the leading contributing factor (NTSB, 1973; AAIB, 2020) One of the differences between the flights undertaken in the current research and these two fatal flights was the number of pilots onboard, and it is plausible that the larger number of crew members led to other known psychological phenomena such as diffusion of responsibility or failure in communication or leadership (e.g., failure to effectively delegate).

4.2. Limitations and future research

While the results of this research are positive, they are not without their limitations. Although Molesworth and Estival (2015) found that increased information density in ATC transmission and increased pilot workload adversely affected the communication performance of all pilots, while increased ATC speech rate adversely affected only non-native English-speaking pilots with low levels of flight experience, the current study found no effect of these three different conditions on flying performance on any of the pilot groups. It is plausible that these factors did not affect flying performance because of their perceived lower importance to the flight itself. A distractor such as a low fuel level or an oil warning light, which would relate to the functionality of the engine might be more distracting than the two factors which affected communication and might affect flying performance. This is an area for future research.

In addition, the current study focused on single pilot operations. With both Eastern Air Lines Flight 401 and Pakistan International Airlines Flight 8303, there were at least two crew members on the flight deck at the time of the crash. Task prioritisation may be affected by leadership or communication skills (Sexton and Helmreich, 2000), but how such skills interact with task prioritisation is another area for future research. More generally, cockpit environments with multiple pilots may feature more or different sources of cognitive workload than those with a single pilot. For example, collaborative tasks that require team members to communicate and share workload impose unique source of workload (Sellers et al., 2014) while the expertise of individual pilots in collaborative environments may not be reflected in effective pilot/co-pilot collaboration (Helton et al., 2014). Future research into multi-pilot operations will need to account for a more complex set of task demands, which may evoke different effects than the current study. Another area for future research is the effectiveness of various instructions techniques to facilitate in task prioritisation.

Finally, the current study relied on performance metrics to detect a dual-task effect. This has the advantage of being objective, whereas commonly used workload measures such as the NASA-TLX (NASA, 2020), which were not employed in the current study, are subjective. However, such measures have the advantage of directness, labelling the item under measure and gauging a response. Performance metrics, by contrast, can be affected by other sources of task demand, such as physical demand or fatigue. Future studies investigating the relationship between cognitive demand and flight performance would be strengthened by validating performance measures against established measures of cognitive workload.

5. Conclusion

Pilots from the outset are taught of the importance of task prioritisation for safe and efficient flight. This is simplified in the "adage, aviate, navigate, communicate". The results from the present research illustrate that when faced with factors that are known to affect communication, pilots are able to effectively prioritise their attention to reduce the impact of these factors on their operation of the aircraft. Using heading error as the dependent variable, no significant effects were found even for flights where pilots committed the most communication errors (Flight 2B with increased information density, and Flight 3B with increased pilot workload). When taken alongside the previous analysis of this cohort's performance on verbal tasks (Molesworth and Estival, 2015; Estival and Molesworth, 2020), these results provide evidence for effective prioritisation of piloting over verbal communication. Thus, these results are encouraging as they show that if the order of operational importance is adhered to, safe flight operations can be maintained even when communication is affected.

CRediT authorship contribution statement

Alexander Thorpe: Methodology, Software, Formal Analysis, Writing. **Dominique Estival:** Conceptualisation, Resources, Funding Acquisition, Writing. **Brett Molesworth:** Conceptualisation, Validation, Writing – Reviewing and Editing. **Ami Eidels:** Supervision, Writing – Reviewing and Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors acknowledge the support of the MARCS Institute, whose grant to Estival made this research possible.

References

AAIB, 2020. Preliminary Report: Crash of PIA AP-BLD 22 May 2020. Aircraft Accident Investigation Board Pakistan, Pakistan.
Armstrong, R.A., 2014. When to use the Bonferroni correction. Ophthalmic Physiol. Opt. 34 (5), 502–508. https://doi.org/10.1111/opo.12131.
Baddeley, A., 2003. Working memory: looking back and looking forward. Nat. Rev. Neurosci. 4 (10), 829–839. https://doi.org/10.1038/nrn1201.
Barshi, I., Farris, C., 2013. Misunderstandings in ATC Communication. Ashgate, Farnham, England.
Bishara, S., Funk, K., 2002. Training pilots to prioritize tasks. Proc. Human Factors Ergonom. Soc. Ann. Meet. 46 (1), 96–100.
Blakely, M.J., Smith, S.L., Russell, P.N., Helton, W.S., 2021. The impact of cognitive load on climbing and climbing on cognitive performance. Appl. Ergon. 94, 103413. https://doi.org/10.1016/j.apergo.2021.103413.
Burki-Cohen, J., 1995. Analysis of tower ground controller-pilot voice communication. Federal Aviation Administration, Washington DC, USA.
Cardosi, K.M., 1993. An analysis of en route controller-pilot voice communications. Federal Aviation Administration, Cambridge, MA, USA.
CASA. (2019). Aviate-navigate-communicate-administrate. *Flight Safety Australia*. Retrieved from https://www.flightsafetyaustralia.com/2019/05/aviate-navigate-communicate-administrate/.
Cohen, J., 1988. Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale, N.J.
Darling, K.A., Helton, W.S., 2014. Dual-task interference between climbing and a simulated communication task. Exp. Brain Res. 232 (4), 1367–1377. https://doi.org/10.1007/s00221-014-3855-7.
Estival, D., Molesworth, B.R.C., 2016. Native English speakers and EL2 pilots : an experimental study. *Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers*. Routledge, London, UK, pp. 140–181.
Estival, D., Molesworth, B.R.C., 2020. Errors in air-ground pilot communication: an experimental study. The ESPecialist, 41(23, Special issue on Aviation English). doi:10.23925/2318-7115.2020v41i3a1.
EUROCONTROL, 2006b. Air-ground Communication Safety Study: Causes and Recommendations (Report # DAP/SAF 2006-09). (DAP/SAF 2006-09). European Organisation for the Safety of Air Navigation, Brussels, Belgium.
EUROCONTROL, 2014. Use of 'AT PILOTS DISCRETION' and 'WHEN READY' – verbal climb and descent clearances. Retrieved from http://us1.campaign-archive1.com/?u=0fa4593636877e9fd022bcdc1&id=16953fee95&e=baba26755a.
FAA. (2018). Fly the Aircraft First. FAA Safety Briefing. Retrieved from https://www.faa.gov/news/safety_briefing/2018/media/SE_Topic_18-07.pdf.
Helton, W.S., Funke, G.J., Knott, B.A., 2014. Measuring workload in collaborative contexts: trait versus state perspectives. Hum. Factors 56 (2), 322–332. https://doi.org/10.1177/0018720813490727.
Helton, W.S., Russell, P.N., 2011. Working memory load and the vigilance decrement. Exp. Brain Res. 212 (3), 429–437. https://doi.org/10.1007/s00221-011-2749-1.
Helton, W.S., Russell, P.N., 2013. Visuospatial and verbal working memory load: effects on visuospatial vigilance. Exp. Brain Res. 224 (3), 429–436. https://doi.org/10.1007/s00221-012-3322-2.
Jeffreys, H., 1961. Theory of Probability, 3rd ed. Oxford University Press, Oxford, UK.
Kahneman, D., 1973. Attention and Effort. Prentice Hall, Englewood Cliffs, NJ.
Kantowitz, B.H., 1985. Channels and stages in human information processing: A limited analysis of theory and methodology. J. Math. Psychol. 29 (2), 135–174. https://doi.org/10.1016/0022-2496(85)90014-8.
Lachman, R., Lachman, J.L., Butterfield, E.C., 2015. Cognitive Psychology and Information Processing: An Introduction. UK Psychology Press, East Sussex.
Molesworth, B.R.C., Estival, D., 2015. Miscommunication in general aviation: The influence of external factors on communication errors. Saf. Sci. 73, 73–79.
Morrow, D., Lee, A., Rodvold, M., 1993. Analysis of problems in routine controller-pilot communication. Int. J. Aviation Psychol. 3 (4), 285–302.
Morrow, D., Rodvold, M., 1993. The influence of ATC message length and timing on pilot communication (NASA Contractor Report 177621). NASA Ames Research Center, Moffett Field, CA, USA.
Morrow, D., Rodvold, M., 1998. Communication Issues in Air Traffic Control. In: Smolensky, M.W., Stein, E.S. (Eds.), Human Factors in Air Traffic Control. Academic Press, San Diego, USA, pp. 421–456.

NASA, 2020. NASA Task Load Index. Retrieved from humansystems.arc.nasa.gov/groups/tlx/.
NTSB, 1973. Aicraft Accident Report: Eastern Airlines Flight 401. (NTSB-AAR-73-14). National Transportation Safety Board, Springfield, VA, USA.
Orlady, H.W., Orlady, L.M., 1999. Human factors in multi-crew flight operations. Ashgate, Aldershot, USA.
Rhodes, S., Cowan, N., 2018. Attention in working memory: attention is needed but it yearns to be free. Ann. N. Y. Acad. Sci. 1424 (1), 52–63. https://doi.org/10.1111/nyas.13652.
Rothman, K.J., 1990. No Adjustments Are Needed for Multiple Comparisons. Epidemiol. Commun. Health 1 (1), 43–46.
Said, H., 2011. Pilots/Air Traffic Controllers Phraseology Study. IATA, Montréal, Québec.
Sellers, J., Helton, W.S., Näswall, K., Funke, G.J., Knott, B.A., 2014. Development of the Team Workload Questionnaire (TWLQ). Proc. Human Factors Ergonom. Soc. Ann. Meet. 58 (1), 989–993. https://doi.org/10.1177/1541931214581207.
Sexton, J.B., Helmreich, R.L., 2000. Analyzing cockpit communications: the links between language, performance, error, and workload. J. Human Perform. Extreme Environ. 5 (1), 689–695.
Stets, A., Smith, S.L., Helton, W.S., 2019. Dual-task interference between swimming and verbal memory. Hum. Factors 62 (7), 1132–1140. https://doi.org/10.1177/0018720819871743.
Taylor, J.L., Yesavage, J.A., Morrow, D.G., Dolhert, N., III, J.O.B., Poon, L.W., 1994. Effects of information load and speech rate on young and older aircraft pilots' ability to read back and execute simulated ATC instructions. J. Gerontol.: Psychol. Sci. 49, 191–200.
Verghese, J., Kuslansky, G., Holtzer, R., Katz, M., Xue, X., Buschke, H., Pahor, M., 2007. Walking while talking: effect of task prioritization in the elderly. Arch. Phys. Med. Rehabil. 88 (1), 50–53. https://doi.org/10.1016/j.apmr.2006.10.007.
Watanabe, K., Funahashi, S., 2014. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. Nat. Neurosci. 17 (4), 601–611. https://doi.org/10.1038/nn.3667.
Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50 (3), 449–455. https://doi.org/10.1518/001872008x288394.
Wickens, C.D., Goh, J., Helleberg, J., Horrey, W.J., Talleur, D.A., 2003. Attentional models of multitask pilot performance using advanced display technology. Hum. Factors 45 (3), 360–380. https://doi.org/10.1518/hfes.45.3.360.27250.
Woodham, A., Billinghurst, M., Helton, W.S., 2016. Climbing with a head-mounted display: dual-task costs. Hum. Factors 58 (3), 452–461. https://doi.org/10.1177/0018720815623431.
Yogev-Seligmann, G., Rotem-Galili, Y., Dickstein, R., Giladi, N., Hausdorff, J.M., 2012. Effects of explicit prioritization on dual task walking in patients with Parkinson's disease. Gait Posture 35 (4), 641–646.
Yogev-Seligmann, G., Rotem-Galili, Y., Mirelman, A., Dickstein, R., Giladi, N., Hausdorff, J.M., 2010. How does explicit prioritization alter walking during dual-task performance? Effects of age and sex on gait speed and variability. Phys. Ther. 90 (2), 177–186.