# Pilots gaze more outside while performing an auditory cognitive task

Landman, H.M.; Stuldreher, Ivo; Van der burg, Erik; Evertsen, F.W.; Reuten, Anna; Ledegang, Wietse; Houben, Mark; Groen, Eric L.

**DOI**
10.1080/00140139.2025.2500526

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Ergonomics

**Citation (APA)**
Landman, H. M., Stuldreher, I., Van der burg, E., Evertsen, F. W., Reuten, A., Ledegang, W., Houben, M., & Groen, E. L. (2025). Pilots gaze more outside while performing an auditory cognitive task. *Ergonomics*. https://doi.org/10.1080/00140139.2025.2500526

Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: www.tandfonline.com/journals/terg20

# Pilots gaze more outside while performing an auditory cognitive task

Annemarie Landman, Ivo V. Stuldreher, Erik Van der Burg, Fleur W. Evertsen, Anna J. C. Reuten, Wietse D. Ledegang, Mark M. J. Houben & Eric L. Groen

**To cite this article:** Annemarie Landman, Ivo V. Stuldreher, Erik Van der Burg, Fleur W. Evertsen, Anna J. C. Reuten, Wietse D. Ledegang, Mark M. J. Houben & Eric L. Groen (16 May 2025): Pilots gaze more outside while performing an auditory cognitive task, Ergonomics, DOI: 10.1080/00140139.2025.2500526

**To link to this article:** https://doi.org/10.1080/00140139.2025.2500526

ERGONOMICS
https://doi.org/10.1080/00140139.2025.2500526

RESEARCH ARTICLE

# Pilots gaze more outside while performing an auditory cognitive task

Annemarie Landman[a,b], Ivo V. Stuldreher[a], Erik Van der Burg[a,c], Fleur W. Evertsen[a,b], Anna J. C. Reuten[a], Wietse D. Ledegang[a], Mark M. J. Houben[a] and Eric L. Groen[a,d]

[a]TNO, Soesterberg, The Netherlands; [b]Delft University of Technology, Delft, The Netherlands; [c]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [d]Cranfield University, Bedford, UK

**ABSTRACT**

Data from two simulator experiments were examined to investigate whether performing an auditory task influences pilots' gaze behaviour. Gaze behaviour was measured while participants performed a manual flying task with an auditory task (dual-task condition) or without (single-task condition). Experiment 1 took place in a fixed-base, fixed-wing simulator with 15 novice military pilots. Experiment 2 took place in a moving-base, rotary-wing simulator with 13 experienced military helicopter pilots. Percentage dwell time outside significantly increased in the dual-task condition compared to the single-task condition in both experiments, by a factor of 1.2 and 1.5 respectively. Mean duration of fixations outside significantly increased for pilots, while it decreased for novices. In novices, altitude control performance was also significantly reduced when performing the auditory task, whereas bank angle control performance significantly increased in experienced pilots. The impact on gaze behaviour may potentially serve as a behavioural indicator of pilot auditory workload.

**Practitioner Summary:** Gaze behaviour was analysed as an indicator of auditory cognitive workload in pilots. In datasets of two simulator experiments, consistent evidence was found for a shift in gaze direction towards outside of the cockpit when pilots performed an auditory mental workload task during manual flight.

**ARTICLE HISTORY**
Received 17 September 2024
Accepted 5 April 2025

**KEYWORDS**
Eye-tracking; aviation; distraction; dual-task; attention

## Introduction

The principle 'Aviate, Navigate, Communicate' is fundamental for general, civil, and military aviation. It serves as a reminder to prioritise flying the aircraft, focus on navigation second, and engage communication only when it does not interfere with the other tasks. This principle is especially important in high-workload situations, as the additional demands of communication can negatively impact flying accuracy, as well as monitoring, troubleshooting and decision-making (Ferraro, Christy, and Mouloua 2017; Hodgetts et al. 2005; Tušl et al. 2020; Wickens and Gosney 2003). Workload management, which involves the prioritisation of tasks for time and attention allocation, was ranked as the third-most important competency by military pilots and instructors (Steinman et al. 2019). Several studies with non-military and military pilot sample groups found that performing an auditory task while flying increases susceptibility to spatial disorientation compared to flying without such additional task (Lewkowicz et al. 2018, 2019; Williams et al. 2018, 2021).

In the automotive domain, detrimental effects of auditory task load on driving performance have been well-reported. For instance, studies have shown that drivers' response times to unexpected events increased significantly when they performed a phone call conversation compared to driving in isolation (Horrey & Wickens, 2004; Recarte and Nunes 2003). One interesting finding is that performing auditory tasks have a pronounced effect on drivers' gaze behaviour (Desmet and Diependaele 2019; Hammel, Fisher, and Pradhan 2002; Nunes and Recarte 2002; Victor, Harbluk, and Engström 2005). This effect comprises increased gazing towards the future path region (i.e. driving direction), less gazing at the dashboard (e.g. speed indication) or road scene periphery (e.g. signs), and binocular dissociation (i.e. when the two eyes are not fixated on a specific point in space), suggesting that people were staring without attending to visual information (Uchida et al. 2005).

Thus far, the effect of performing an auditory task on looking outside has not yet been investigated within the aviation domain. The visual scene and

**CONTACT** Annemarie Landman annemarie.landman@tno.nl TNO, Kampweg 55, 3769 DE, Soesterberg, The Netherlands.

relevant regions therein differ between driving and flying with outside visibility, the outside region is similar in that it offers information on parameters that are most directly related to the vehicle's path. In flying, these parameters are pitch and roll, visible from the orientation of the aircraft's nose and wings relative to the horizon. Other flight path information (heading, speed and altitude) is best derived from the flight instruments, so that cross-checking these panels is required frequently. Xie et al. (2024) found that performing an auditory task made non-pilot participants look less and shorter towards task-relevant areas in a flight simulator, but included no analysis on the difference between inside and outside regions. Other studies have manipulated workload through complexity of the flying task, which itself leads to a different visual scene with more (dynamic) information being presented (see, review by Peißl, Wickens & Baruah, 2018; Ziv, 2016). Increased mental workload was thus found to be associated with *decreased* fixation durations and *increased* gaze dispersion (De Rivecourt et al. 2008; Di Nocera, Camilli, and Terenzi 2007). Mental workload has also been manipulated by varying the amount and complexity of visual information presented on the instruments (e.g. the tactical display), which was found to lead to a higher percentage of gaze directed to these instruments (Svensson, Angelborg-Thanderez, Sjöberg and Olsson, 1997). These effects are in the opposite direction to the effect found in driving experiments, which is likely caused by the use of visual instead of auditory tasks.

In the current study, we tested whether performing an auditory task, compared to not performing an auditory task, affects the gaze behaviour of pilots (both experienced and novices) when manually flying an aircraft. This was achieved by re-analysing the datasets of two independent experiments, one with novice military pilots (Experiment 1; Stuldreher et al. 2024) and one with military helicopter pilots (Experiment 2; Evertsen et al. 2025) who performed simulated flying tasks with and without an additional auditory task. The percentage dwell time directed outside (primary outcome), mean fixation duration, and flying performance (secondary outcome) were compared between these two conditions, while complexity of the flying task and visual scene remained constant. Our gaze behaviour analysis focused on measures of gazing and fixating outside. Other metrics of gaze behaviour (e.g. dispersion, blinks, pupil diameter) were not included, as they likely correlate strongly with gazing outside, and would increase the risk of family-wise errors if not corrected for. In accordance with the above-mentioned literature from the automotive domain, we hypothesised that an additional auditory task increases the proportion dwell time directed outside and reduces flying performance. The auditory tasks required storing information in working memory in both experiments) and required processing of this information in Experiment 2. The difference in experience level of the pilot groups may offer interesting insights into more intuitive gaze behaviour (novices) or learned gaze behaviour (experienced pilots).

## Method

### *Experimental design*

Both experiments had a within-subject design with two conditions: a single task condition where participants only performed a flying task, and a dual task condition where participants performed the same flying task while performing an additional auditory cognitive task (see, tasks and conditions). Gaze behaviour and flying performance (see, dependent measures) were compared between these conditions. No comparison is made between the different sample groups, as the experimental tasks for both groups differed significantly. Both experiments featured basic single-pilot manual flying tasks without navigation, communication or tactical aspects. No other instruments were used besides the primary flight display, and no other controls were used besides stick and rudder in fixed-wing, or cyclic stick, collective lever and pedals in rotary-wing.

### *Experiment 1: fixed-wing experiment with novices*

#### *Participants*

In the fixed-wing experiment, 15 military cadets (novices) without or with minimal flight experience (3 females, mean age = 23.7 years, $SD = 2.4$ years) participated. This number allows us to determine a large effect size (Cohen's $d = 0.78$) for testing our hypothesis, with alpha = 0.05 and beta = 0.2. The average number of flight hours of the novices was 4 ($SD = 8$) and all had normal or corrected-to-normal vision. Participants confirmed that they were in good health, had refrained from using any drugs or alcohol at least 24 hours before participating, and obtained sufficient sleep the night before the experiment. This study complied with the tenets of the Declaration of Helsinki and was approved by the Ethical Review Board of TNO Soesterberg under number 2022-011.

#### *Apparatus*

The fixed-wing experiment was performed using a simulator environment developed by multiSIM B.V. (Soesterberg, the Netherlands), consisting of a fixed-base cockpit front-seat of a Pilatus PC-7 turboprop trainer aircraft, and control devices with control loading (see, Figure 1). A Varjo Aero (Varjo, Helsinki, Finland; 115° field of view) virtual reality (VR) headset with built-in eye-tracker was used to present the cockpit and a virtual environment near the Woensdrecht Air Base in the Netherlands. The flight model characteristics were validated by flight instructors of the Royal Netherlands Air Force.

***Tasks and conditions.*** Participants followed a two-day training program in the VR simulator, comprising of three sessions. Session one was performed on day one, sessions two and three on day two. In each session participants practiced three manoeuvres: first straight and level flight, then a speed change, and finally a level turn (see Figure 2). In the straight and level flight condition, participants were instructed to maintain an altitude of 5000 ft, an airspeed of 180 kts and a constant heading. The aeroplane was initialised out-of-trim and participants could use pitch and rudder trim to reduce forces on the stick and rudder. In the speed change condition, participants were instructed to decelerate from 180 to 110 kts upon an auditory command after 20 seconds, while keeping a constant heading and constant altitude of 5000 ft. In the level turn condition, participants were instructed to perform a 360-degree turn upon an auditory command after 20 seconds. The turn had to be flown with a bank angle of 30 degrees and at a constant altitude of 5000 ft. Each manoeuvre was practiced three times in sequence during 210 second runs (i.e. single-task condition), followed by a fourth run in which the manoeuvre was performed while simultaneously performing an auditory 2-back task (i.e. dual-task condition). The eye tracker was calibrated before each run. In the current study, we only analysed data of the third and fourth runs of each session to minimise the effect of training progression on results. This resulted in 31.5 minutes of gaze behaviour data per condition for each participant.

In the 2-back task, participants were presented with a randomised sequence of auditory vowels at a fixed 3.5-second interval, starting 9.5 seconds after the trial onset. Participants were instructed to press a dedicated button on their throttle – that was controlled by their left hand – when the current letter matched the letter two letters back and to withhold their response when the letter was different. The letters were randomly selected with the constraint that 25% of the 58 trials were response trials (i.e. the current letter matched the letter two trials back and should trigger a button press). Prior to the first training session, participants performed the 2-back task once without concurrently controlling the aircraft to obtain a baseline performance measure. The principle of 'Aviate, Navigate, Communicate' was instructed to participants. Here it was instructed that the 2-back task is part of communication and flying the aircraft should thus take priority.

In each run, in addition to manually controlling the aircraft, participants were instructed to continuously perform a systematic lookout pattern, which was not used for the current study. This pattern consisted of a scan along the horizon from the wingtip to the centre, then a check of the aircraft's nose attitude, and then a crosscheck of the cockpit instruments. This pattern is the default lookout pattern that is instructed to all students in Dutch initial military pilot training. This default lookout pattern must aid pilots in avoiding objects, while continuously maintaining aircraft control. To verify whether subjects were performing the lookout, and not just moving their head, they had to detect a visual object in the outside visuals. The visual object (i.e. a grey sphere) was shown at about 1.6 NM (3 km) distance with various inclinations at semi-random intervals (varying between 15 and 25 seconds). The subjects were instructed to fixate on the object, as measured by the eye tracker, while simultaneously

Figure 1. The experimental setup of the fixed-base simulator used in the fixed-wing experiment (source: Mediacentrum Defensie, Sjoerd Hilckmann).

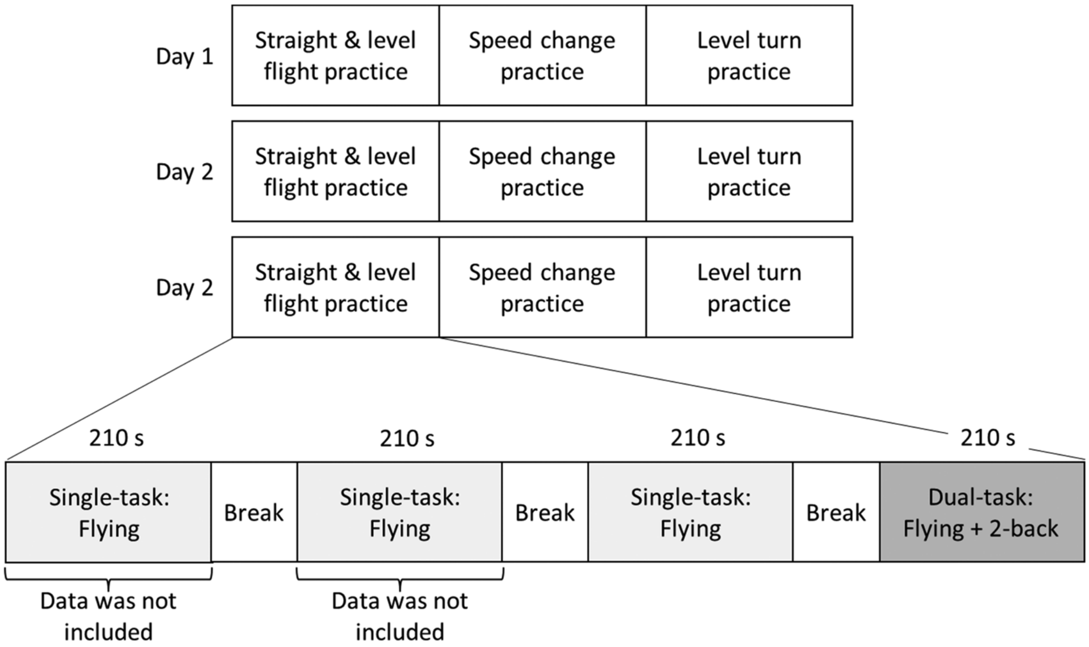

Figure 2. Overview of the tasks in the fixed-wing experiment. Data of the first two single-task flying runs was not included to reduce potential effects of learning on the outcomes.

pulling the trigger on the control stick – that was controlled by their right hand – to make the object disappear.

## Experiment 2: Rotary-wing experiment with experienced pilots

### Participants

In the rotary-wing experiment, 13 helicopter pilots participated (all male, mean age = 34.4 years, *SD* = 7.0 years, seven captains and six copilots). This number allows us to determine a large effect size (Cohen's $d = 0.85$) for testing our hypothesis, with alpha = 0.05 and beta = 0.2. Nine possessed a military pilot licence for the AS532-U2 Cougar, two for the CH-47 Chinook and two for the NH90. One more pilot initially participated but dropped out due to a simulator failure and insufficient gaze behaviour data was obtained for the current study. The average number of flight hours was 1510, SD = 1068, and all had normal or corrected-to-normal vision. None had experience with flying the simulated helicopter type used in this study. Participants confirmed that they were in good health, had refrained from using any drugs or alcohol at least 24 hours before participating, and obtained sufficient sleep the night before the experiment. This study complied with the tenets of the Declaration of Helsinki and was approved by the Ethical Review Board of TNO Soesterberg under number 2023-059.

### Apparatus

The experiment was performed in the six degrees-of-freedom motion simulator DESDEMONA, constructed by AMST-Systemtechnik GmbH (Ranshofen, Austria) and located in Soesterberg, the Netherlands (see, Figure 3). The simulator featured a high-fidelity AH-64 Apache flight model, collective and cyclic, and standard pedals. Visuals were presented in VR with the same Varjo Aero (Varjo, Helsinki, Finland) VR headset as used in Experiment 1. Head tracking relative to the simulator cabin was obtained using SteamVR and the VR headset, without using base stations. Rotations of the visual scene due to cabin motion were corrected using cabin rotation angles known within the motion cueing. Yaw drift of the visual scene was compensated using visual motion capture and two passive markers attached to the headset. Pitch and roll drift were corrected by the tracking algorithm of the VR headset. Translations of the headset were not tracked. To simulate natural head movement, head rotations were applied to the base of a rod, which simulates an offset between the neck's rotational centre and eyes of the participant. The multi-purpose displays (MPDs) in the virtual cockpit showed the Tactical Situation Display (TSD) on the left MPD and the flight page on the right MPD. The Target Acquisition and Designation Sights' Electronic Display and Control (TEDAC) in the middle presented flight information but no sensor images. No helmet mounted display (HMD) was used.

Figure 3. Left: The six degrees-of-freedom DESDEMONA motion simulator used for Experiment 2. Middle: A participant sitting in the cabin. Right: the virtual scene with gaze direction indicated by the circle.

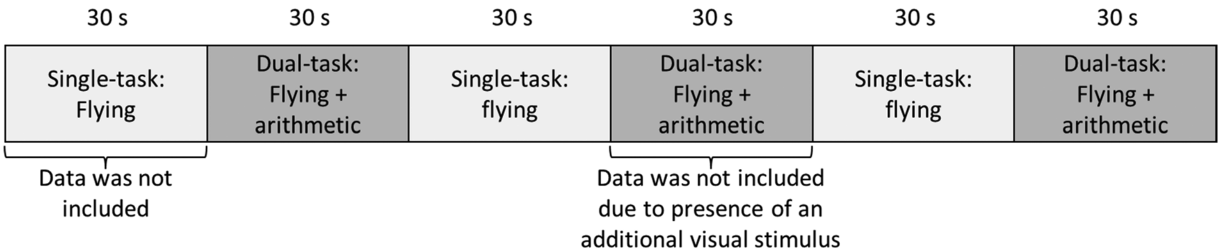

Figure 4. Overview of one run in the rotary-wing experiment. Data of the first single-task flying interval was not included as the helicopter needed to be stabilised during this interval.

### *Task and conditions*

Participants performed a single session on one day, with a 15-min break half-way in. The eye-tracker was calibrated at the start of the session and again after the break. Participants first familiarised themselves with the helicopter model by flying a traffic pattern, then practiced the cognitive task once on the ground, and then practiced both tasks simultaneously two times. Participants were instructed that the flying task took precedence over the cognitive task and received no special instructions on gaze behaviour. After familiarisation, they performed six different 3-minute scenarios, each in two conditions, and repeated twice (i.e. 6×2 x 2), in a randomised order. The scenarios considered different spatial disorientation stimuli, and the conditions consisted of either the presence or absence of a spatial disorientation stimulus. The scenarios were: straight-and-level flight without outside visibility with/without a leans stimulus, straight-and-level flight with/without a false (sloped) horizon, making two turns and flying straight-and-level above featureless/featured terrain, monitoring a flight with/without somatogyral yaw stimulus, monitoring a flight with/without brownout, and monitoring a flight with/without loss of visual horizon while wearing night vision goggles. Only the false horizon scenario (both conditions) was suitable for gaze analysis, as it featured a constant flying task and outside visibility. In this scenario, participants had to maintain straight-and-level flight above a cloud layer while flying at speed 95 kt and altitude 4,800 ft and be aware that other traffic may be present. Other scenarios were not included here because they featured a monitoring task instead of a flying task, the flying task was not constant between conditions, or there was no outside visibility. No instructions were given on the focus of gaze or attention in this experiment.

In the included scenario, an additional cognitive task (dual-task condition) was performed at three 30-second intervals (see, Figure 4). This was a mental arithmetic task consisting of single-digit, single-operator addition and subtraction problems, presented auditorily on the headset with 3-second intervals (i.e. 10 problems per 30-second interval). The pilots answered whether the solution was above or below five using a switch button on the cyclic. The second 30-second interval of dual-task performance could not be included for gaze analysis, as a visual stimulus (sloped cloud deck) was presented during this interval. The two 30-second intervals between dual-task performance were used for the single-task condition. The 30-second interval at the start of each run was not included in the single-task condition due to participants needing to stabilise the helicopter. This resulted in four minutes of gaze behaviour data per condition for each participant.

***Eye-tracking data collection.*** In both experiments, gaze direction (eye theta and psi, respectively vertical and horizontal deviation from looking straight ahead) was measured at 200 Hz using the Varjo Aero VR headset. The position of the eyes remained fixed in the virtual world; except that small translations were induced based on rotations of the head (see, Apparatus section). In both experiments, two areas of interest (AOIs) were defined: the inside of the cockpit and the outside environment. Based on inspection of the virtual scene, in the fixed-wing experiment theta < −15° was used to define the inside the cockpit and theta ≥ −15° the outside. In the rotary-wing experiment theta < −20° was used to define the inside of the cockpit and theta > −20° the outside. For the rotary-wing experiment, data was inspected to confirm that participants did not look outside of the side windows during the tasks, as these extend below −20°.

## Dependent measures

### % Dwell time outside

In each condition, the proportion of time (samples) with gaze located within the AOI 'outside environment' compared to the total time (samples) was obtained. Samples where the gaze tracker missed data (e.g. due to blinks or saccades) were excluded from this total. The mean was taken over all single-task segments and over all dual-task segments in each experiment, and expressed as percentage.

### Mean fixation duration

The mean duration of fixations was obtained in each condition. Fixations were identified using the Python implementation of EyeMMV's fixation detection algorithm (PeyeMMV, Krassanakis 2023). This is a dispersion based algorithm, in which fixations are identified based on spatiotemporal criteria. The algorithm has three free parameters, being spatial parameters $t_1$ and $t_2$ and minimum duration for fixation identification. In our analyses, $t_1$ and $t_2$ were set to 1 deg, and the minimum fixation duration was 150 ms. We separately report fixation duration of fixations inside the cockpit and fixations in the outside environment.

### Flying performance

As a secondary outcome measure, manual flying performance was measured. For this, the mean absolute error of the roll angle from the instructed roll angle (i.e. 0 or 30 degrees), and the mean absolute error of the altitude from the instructed altitude, were obtained. The mean was taken over all single-task segments and over all dual-task segments in each experiment. Roll was used instead of heading, because some tasks did not include an instructed heading.

### Auditory task performance

To provide information on the level of performance on the auditory tasks, the average response times and percentage of errors – either a miss (incorrect withhold of response) or a false alarm (incorrect response) – were obtained for Experiment 1, and average response times (from presentation of the last number until button press) and the proportion of correct answers were obtained for Experiment 2.

### Motion sickness

To monitor problems with motion sickness, participants of Experiment 1 rated the severity of motion sickness symptoms after each four consecutive runs for each manoeuvre. They answered the question 'Did you experience any discomfort (headache, tired eyes, nausea, etc.) during the exercise?' on a visual-analogue scale ranging from 0 to 100. Participants of Experiment 2 rated the severity of motion sickness symptoms half-way into the test session and at the end, on the Dutch version of the Misery scale (MISC), which ranges from 0 ('no problems') to 10 ('vomiting'; Bos et al., 2005).

## Data analysis

For each participant, we calculated the mean % dwell time outside, the mean fixation duration, the mean roll error and the mean altitude error in the single-task and dual-task condition. For each dependent variable, we examined whether the dual task condition differed compared to the single task condition using two-tailed paired-samples *t*-tests. If not normally distributed, Wilcoxon signed rank tests were used, and medians and interquartile ranges (IQRs) are reported instead of means and *SD*s. Alpha was set at 0.05. Effect sizes (Cohen's *d*) of 0.2, 0.5 and 0.8 will be discussed as small, medium and large, respectively.

# Results

## Experiment 1: Fixed-wing experiment with novices

Figure 5 shows an overview of the gaze behaviour results, and Figure 6 shows the flying performance results, in Experiment 1.

### % Dwell time outside

The % dwell time outside was significantly higher in the dual-task, mean = 71%, SD = 6.6, than in the single-task

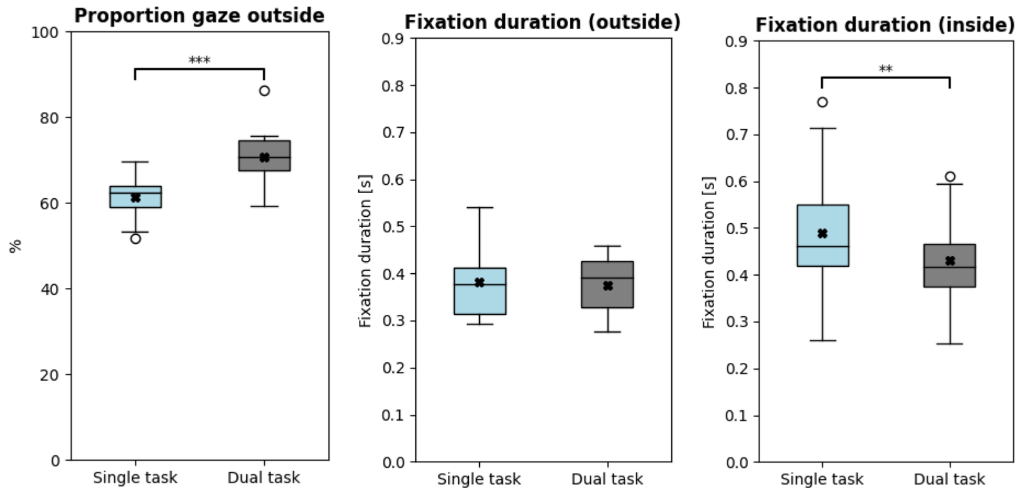

Figure 5. Tukey boxplots with gaze behaviour results in the single-task and dual-task conditions of Experiment 1 (novices). × = mean, whiskers indicate 1.5 IQR.

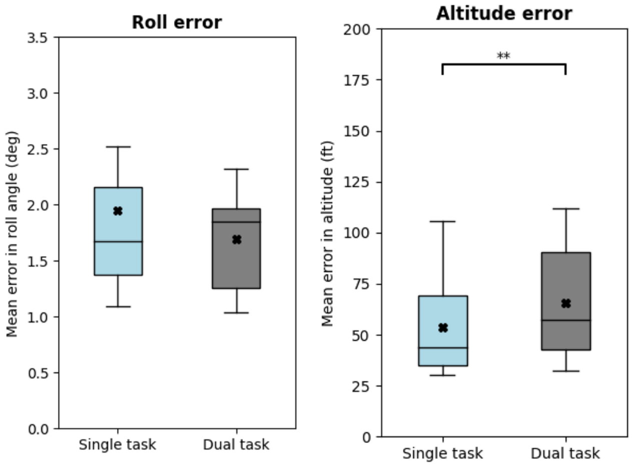

Figure 6. Tukey boxplots with flying performance results in the single-task and dual-task conditions of Experiment 1 (novices). × = mean, whiskers indicate 1.5 IQR.

condition, mean = 61%, $SD=5.1$, $t(14) = 7.01$, $r=0.650$, $p<0.001$. The effect size was large, $d=1.84$.

### Fixation duration

The mean fixation duration of fixations in the outside environment did not significantly differ between single-task, mean = 0.38 s, $SD=0.08$ s, and dual-task, mean = 0.37 s, $SD=0.06$ conditions, $t(14) = 0.54$, p = .060. The mean fixation duration of fixations in the cockpit was significantly lower for the dual-task, mean = 0.43 s, $SD=0.10$ s, than for the single-task, mean = 0.49 s, $SD=0.13$ s condition, $t(14) = 4.01$, $p = .001$.

### Auditory task performance

The average response time was 1.33 s, $SD=0.16$. The average error rate was 9.3%, $SD=3.8\%$.

***Flying performance.*** Flying performance results are shown in Figure 6. The mean roll error was not significantly higher in the dual-task, mean = 1.53°, $SD=0.33°$, than in the single-task condition, mean = 1.72°, $SD=0.71°$, $t(14) = -0.97$, $r=0.041$, $p = .351$.

The mean altitude error was significantly higher in the dual-task, mean = 65.8 ft, $SD=27.6$ ft, than in the single-task condition, mean = 53.8 ft, $SD=23.1$ ft, $t(14) = 3.94$, $r=0.906$, $p = .001$.

### Motion sickness

Self-reported symptoms of motion sickness were generally low for all manoeuvres, including straight and level flight, $M=6.6/100$, $SD=4.7$, speed change, $M=6.5/100$, $SD=7.5$, and level turn, $M=6.5/100$, $SD=5.2$.

## Experiment 2: rotary-wing experiment with experienced pilots

Figure 7 shows the gaze behaviour results, and Figure 8 shows the flying performance results, of Experiment 2.

### % Dwell time outside

The % dwell time outside was significantly higher in the dual-task, mean = 44%, $SD=13$, than in the single-task condition, mean = 29%, $SD=12$, $t(12) = 7.96$, $r=0.868$, $p<0.001$. The effect size was large, $d=2.21$.

### Mean fixation duration

The mean fixation duration in the outside environment was significantly greater in the dual-task condition, mean = 0.37 s, $SD=0.05$ s, than in the single-task condition,

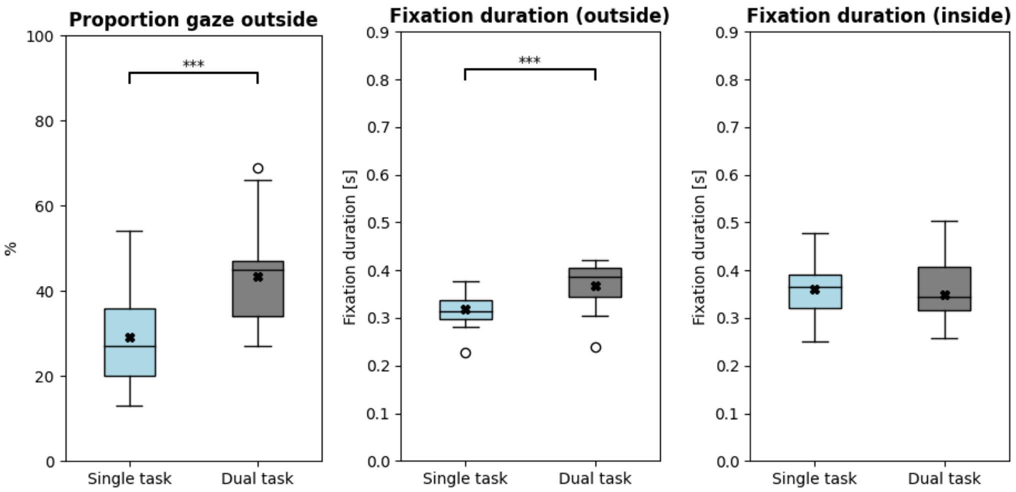

Figure 7. Tukey boxplots with gaze behaviour results in the single-task and dual-task conditions of Experiment 2 (experienced pilots). × = mean, whiskers indicate 1.5 IQR.

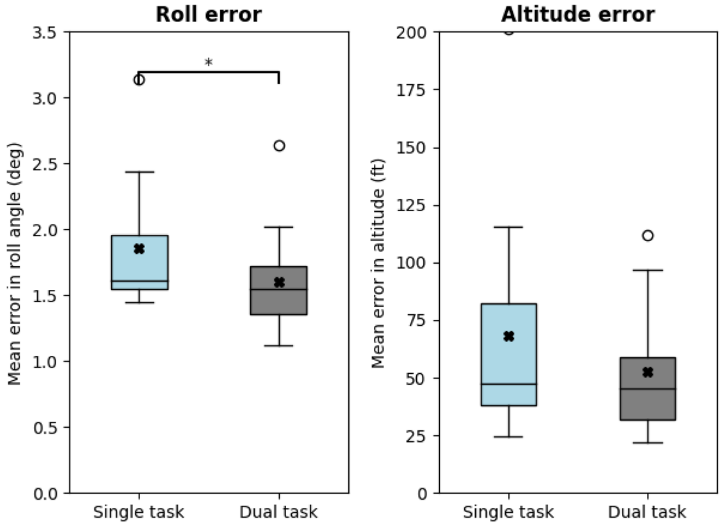

Figure 8. Tukey boxplots with flying performance results in the single-task and dual-task conditions of Experiment 2 (experienced pilots). × = mean, whiskers indicate 1.5 IQR.

mean = 0.32, *SD*=0.04 s, condition, *t*(12) = 5.12, *p* < .001. The mean fixation duration of fixations in the cockpit did not significantly differ between the single-task, mean = 0.36 s, *SD*=0.07 s, and dual-task, mean = 0.35 s, *SD*=0.07 s conditions, *t*(12) = 0.93, *p* = .369.

### Flying performance

The flying performance results are shown in Figure 8. The mean roll error was significantly greater in the dual-task, mean = 1.86°, *SD*=0.51°, than in the single-task condition, mean = 1.61°, *SD*=0.41°, *t*(12) = 2.69, *r*=0.749, *p*=0.020. The effect size was large, *d*=0.75.

The mean altitude error was not significantly higher in the dual-task, mean = 52.6 ft, *SD*=28.2 ft, than in the single-task condition, mean = 56.3 ft, *SD*=29.6 ft, *t*(12) = 0.62, *r*=0.722, *p*=0.547.

### Auditory task performance

The average response time was 0.81 s, *SD*=0.33. The median of the proportion of correct answers was 95%, IQR = 10%.

### Motion sickness

Self-reported symptoms of motion sickness were generally low halfway into the test session, *M*=1.4, *SD*=1.9, as well as at the end, *M*=1.7, *SD*=1.9.

## Discussion

In both experiments, the gaze data show a shift towards the outside environment when the pilots performed an auditory task. This increase was by a factor of 1.2 in novices and 1.5 in experienced pilots. This effect was observed despite differences in simulated aircraft (fixed-wing versus rotary-wing), simulator motion (fixed-base versus moving-base simulator), flight experience level (novices versus experienced military helicopter pilots), and the type of cognitive task (n-back versus mental arithmetic). The shift towards outside also persisted despite the lookout task which the novices were performing throughout the scenarios. For the novices, the auditory task induced a shorter mean fixation duration inside, while the mean fixation duration outside did not significantly differ. Thus, when looking outside, they looked around more. This can be explained by the lookout task, performed for a different experiment, requiring them to look at different areas in the outside scene. For the experienced pilots, the auditory task induced a longer mean fixation duration outside, which could indicate staring behaviour and an interference of the auditory task with their usual cross-checking of the instruments.

The increased looking (and staring) outside we observed might be explained by the auditory task occupying a large part of the limited amount of cognitive resources. This might reduce the availability of resources for perceptual-motor processing needed for the flying task (see, Wickens 1980). With reduced cognitive resources available for visual processing, the participants may have resorted to ‘resting their eyes’ on the outside scene, which did not present a high information density as the instruments. Staring at the outside scene may also have been preferable over staring at the inside of the cockpit, as the horizon possibly presents information on the aircraft attitude most intuitively, allowing for efficient management of the immediate flight path while performing the auditory task. According to the multiple resource model (see, Wickens, 2024), the extend of auditory task interference would increase if the auditory task is more spatial instead of verbal, is more complex, and has higher priority. Future research could vary one of more of these aspects to test whether % dwell time outside is indeed indicative of reduced availability of cognitive resources.

Tole et al. (1982) found a similar effect of auditory task load on pilots’ focus on the primary compared to secondary instruments. Our findings align also with findings from driving tasks with hands-free phone conversations (Desmet and Diependaele 2019; Nunes & Recarte, 2002; Uchida et al. 2005; Victor, Harbluk, and Engström 2005). In these studies, increased dwell time on the future path region was not only associated with increased auditory task demands, but also with increased visual task and driving task demands.

With respect to flying performance, outcomes of Experiment 1 partly confirmed our hypothesis, as the mean altitude error was significantly larger in the dual-task condition compared to the single-task condition. Outcomes of Experiment 2 were in contrast to our hypothesis, as the mean roll error was smaller and no significant effects on mean altitude error were found. The sample group of experienced pilots in Experiment 2 may have been less affected by auditory task load (see also Bałaj et al. 2019). Perhaps roll angle control performance was facilitated in this group by gazing more outside, as military helicopter pilots are trained to use outside information for attitude control (see also Ho et al. 2016). Only basic flight path management performance could be measured in the current study, due to either short durations of time of measuring (i.e. 30 second periods in Experiment 2) or lack of navigation tasks (headings). It remains to be investigated if gaze behaviour and flying performance are more affected by an auditory task if the flying task is more demanding in terms of navigation and speed management.

The current study makes use of the datasets of two experiments which were not originally intended for the purpose of testing the effect of cognitive task load on gaze behaviour. This leads to the following limitations. First, in both experiments the order of the conditions was not counterbalanced within and between participants. For the novices a dual-task condition was always performed after the single-task condition. However, a learning effect would likely *reduce* the impact of the dual-task on mental workload and thus on gaze behaviour, meaning that the effect could be underestimated instead of overestimated. Second, task performance durations were not very long, reducing the general accuracy of results and preventing accurate analyses of low-frequency performance variables such as speed and altitude control. Third, tasks for which the instruments are the only source of information (e.g. navigation, system management) were not included in the current study. This suggests that our results underestimate the detrimental effect of performing an auditory task on flying performance. Fourth, sample groups consisted of novices and military pilots, who were both not trained on the simulated aircraft. Mental workload required for flying was likely higher than it would be for trained pilots, limiting the generalisability of our results to operational practice. Fifth, the limited sample sizes and measuring durations resulted in limited statistical power for analysing a wide range of gaze behaviour measures.

Motion sickness symptoms were unlikely to affect the results, as these were very low throughout the experiment (i.e. Experiment 1: <10 on a 0-100 scale; Experiment 2: <2 on a 0-10 scale). There was no indication that participants ignored the auditory task, as mean error and miss rates remained below 10%. This suggest a possible ceiling effect (at least for error rates), and future research on this topic may benefit from increasing the difficulty of the auditory task.

Despite the exploratory nature of this study, the observed effect on gaze behaviour may warrant further research to investigate whether it has implications for flight safety or workload management training. Increased looking outside during auditory tasks could interrupt adequate scanning of the instruments in pilots and could impair pilot situation awareness and lead to spatial disorientation. On a more practical side, the analysis of staring behaviour may possibly be used as an indicator of pilot mental workload in a non-invasive way. Follow-up research on this effect could use more focused experiment designs and include more complex flying tasks.

## Disclosure statement

The authors report there are no competing interests to declare.

## Funding

This study was funded by the Dutch Ministry of Defense, under research program V2306 'NextGen Aircrew Performance'.

## References

Bałaj, B., R. Lewkowicz, P. Francuz, P. Augustynowicz, A. Fudali-Czyż, P. Stróżak, and O. Truszczyński. 2019. "Spatial Disorientation Cue Effects on Gaze Behaviour in Pilots and Non-Pilots." *Cognition, Technology & Work* 21 (3): 473–486. doi:10.1007/s10111-018-0534-7.

Bos, J. E., S. N. MacKinnon, and A. Patterson. 2005. "Motion Sickness Symptoms in a Ship Motion Simulator: effects of inside, outside, and no View." *Aviation, Space, and Environmental Medicine* 76 (12): 1111–1118.

De Rivecourt, M., M. N. Kuperus, W. J. Post, and L. J. M. Mulder. 2008. "Cardiovascular and Eye Activity Measures as Indices for Momentary Changes in Mental Effort during Simulated Flight." *Ergonomics 51* (9): 1295–1319. doi:10.1080/00140130802120267.

Desmet, C., and K. Diependaele. 2019. "An Eye-Tracking Study on the Road Examining the Effects of Handsfree Phoning on Visual Attention." *Transportation Research Part F: Traffic Psychology and Behaviour 60*: 549–559. doi:10.1016/j.trf.2018.11.013.

Di Nocera, F., M. Camilli, and M. Terenzi. 2007. "A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload." *Journal of Cognitive Engineering and Decision Making 1* (3): 271–285. doi:10.1518/155534307X255627.

Evertsen, F., A. Landman, E. L. Groen, M. M. J. Houben, M. M. (René) Van Paassen, O. Stroosma, and M. Mulder. 2025. "Quantifying the Impact of Spatially Disorienting Stimuli on Pilot Mental State Using Eye-Tracking and Secondary Task Performance." *Human Factors*. doi:10.1177/00187208251323116.

Ferraro, J., N. Christy, and M. Mouloua. 2017. "Impact of Auditory Interference on Automated Task Monitoring and Workload." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61 (1): 1136–1140.

Hammel, K. R., D. L. Fisher, and A. K. Pradhan. 2002. "Verbal and Spatial Loading Effects on Eye Movements in Driving Simulators: A Comparison to Real World Driving.". In Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 2174–2178.

Ho, H. F., H. S. Su, W. C. Li, C. S. Yu, and G. Braithwaite. 2016. "Pilots' Latency of First Fixation and Dwell among Regions of Interest on the Flight Deck.". In Engineering Psychology and Cognitive Ergonomics: 13th International Conference, EPCE 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings 13, 389–396. Springer International Publishing.

Hodgetts, H., E. Farmer, M. Joose, F. Parmentier, D. Schaefer, P. Hoogeboom, … D. Jones. 2005. "The Effects of Party Line Communication on Flight Task Performance." In *Human Factors in Design, Safety, and Management*, edited by D. de Waard, K. Brookhuis, R. A.van Egmond, and T. Boersema, 1–12. Maastricht, the Netherlands: Shaker Publishing.

Horrey, W. J., and C. D. Wickens. 2004. "Cell Phones and Driving Performance: A Meta-Analysis.". In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 48, No. 19*, 2304–2308. Sage CA: Los Angeles, CA: SAGE Publications. doi:10.1177/154193120404801921.

Krassanakis, V. 2023. "PeyeMMV: Python Implementation of EyeMMV's Fixation Detection Algorithm." *Software Impacts* 15: 100475. doi:10.1016/j.simpa.2023.100475.

Lewkowicz, R., P. Stróżak, B. Bałaj, and P. Francuz. 2019. "Auditory Verbal Working Memory Load Effects on a Simulator-Induced Spatial Disorientation Event." *Aerospace Medicine and Human Performance 90* (6): 531–539. doi:10.3357/AMHP.5277.2019.

Lewkowicz, R., P. Stróżak, B. Bałaj, P. Francuz, and P. Augustynowicz. 2018. "Selective Auditory Attention and Spatial Disorientation Cues Effect on Flight Performance." *Aerospace Medicine and Human Performance 89* (11): 976–984. doi:10.3357/AMHP.5153.2018.

Nunes, L., and M. A. Recarte. 2002. "Cognitive Demands of Hands-Free-Phone Conversation While Driving." *Transportation Research Part F: Traffic Psychology and Behaviour 5* (2): 133–144. doi:10.1016/S1369-8478(02)00012-8.

Peißl, S., C. D. Wickens, and R. Baruah. 2018. "Eye-Tracking Measures in Aviation: A Selective Literature Review." *The International Journal of Aerospace Psychology* 28 (3-4): 98–112. doi:10.1080/24721840.2018.1514978.

Recarte, M. A., and L. M. Nunes. 2003. "Mental Workload While Driving: effects on Visual Search, Discrimination, and Decision Making." *Journal of Experimental Psychology. Applied 9* (2): 119–137. doi:10.1037/1076-898x.9.2.119.

Steinman, Y., M. H. van den Oord, M. H. Frings-Dresen, and J. K. Sluiter. 2019. "Flight Performance Aspects during Military Helicopter Flights." *Aerospace Medicine and Human Performance 90* (4): 389–395. doi:10.3357/AMHP.5226.2019.

Stuldreher, I. V., W. D. Ledegang, E. Van der Burg, M. M. J. Houben, D. van der Horst, G. Almekinders, and E. L. Groen. 2024. "Measuring Progression in Manual Flying Skills and Gaze Behavior in a Virtual Reality Flight Simulator." Manuscript Submitted for Publication.

Svensson, E., M. Angelborg-Thanderz, L. Sjoberg, and S. Olsson. 1997. "Information Complexity-Mental Workload and Performance in Combat Aircraft." *Ergonomics* 40 (3): 362–380. doi:10.1080/001401397188206.

Tole, J. R., A. T. Stephens, R. L. S. Harris, and A. R. Ephrath. 1982. "Visual Scanning Behavior and Mental Workload in Aircraft Pilots." *Aviation, Space, and Environmental Medicine* 53 (1): 54–61.

Tušl, M., G. Rainieri, F. Fraboni, M. De Angelis, M. Depolo, L. Pietrantoni, and A. Pingitore. 2020. "Helicopter Pilots' Tasks, Subjective Workload, and the Role of External Visual Cues during Shipboard Landing." *Journal of Cognitive Engineering and Decision Making 14* (3): 242–257. doi:10.1177/1555343420948720.

Uchida, N., Y. Asano, T. Ueda, and A. Iihoshi. 2005. "Effects of Hands-Free Phone Conversation on Visual Behavior: dissociation of Binocular Gaze Point as an Index of Inattention." *SAE Transactions114* (6): 481–488.

Victor, T. W., J. L. Harbluk, and J. A. Engström. 2005. "Sensitivity of Eye-Movement Measures to in-Vehicle Task Difficulty." *Transportation Research Part F: Traffic Psychology and Behaviour 8* (2): 167–190. doi:10.1016/j.trf.2005.04.014.

Wickens, C. D. 1980. "The Structure of Attentional Resources". In *Attention and Performance VIII*, edited by R. Nickerson, 239–257. Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C. D. 2024. "The Multiple Resource Theory and Model. Some Misconceptions in Data Interpretations." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Sage CA: Los Angeles, CA: SAGE Publications.

Wickens, C. D., and J. L. Gosney. 2003. "Redundancy, Modality, and Priority in Dual Task Interference." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 47 (13), 1590–1594.

Williams, H. P., D. S. Horning, B. D. Lawson, C. R. Powell, and F. R. Patterson. 2018. "Effects of Various Types of Cockpit Workload on Incidence of Spatial Disorientation in Simulated Flight."

Williams, H. P., D. S. Horning, C. Etgen, and C. R. Powell. 2021. "Effects of Cockpit Workload and Motion on Incidence of Spatial Disorientation in Simulated Flight." Technical Report NAMRU-D-21-034. Wright-Patterson Air Force Base, Dayton, OH: Naval Medical Research Unit-Dayton.

Xie, X., T. Li, S. Xu, Y. Yu, Y. Ma, Z. Liu, and M. Ji. 2024. "The Effects of Auditory Working Memory Task on Situation Awareness in Complex Dynamic Environments: An Eye-Movement Study." *Human Factors* 66 (7): 1844–1859. doi:10.1177/00187208231191389.

Ziv, G. 2016. "Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation." *The International Journal of Aviation Psychology* 26 (3-4): 75–104. doi:10.1080/10508414.2017.1313096.