Titre: Title:	Mechanical device or touchscreen widget: the effects of input device and task size on data entry on the primary flight display
Auteurs: Authors:	Charles-Antoine Lanoix, Srishti Rawal, & Philippe Doyon-Poulin
Date:	2024
Type:	Article de revue / Article
Référence: Citation:	Lanoix, C.-A., Rawal, S., & Doyon-Poulin, P. (2024). Mechanical device or touchscreen widget: the effects of input device and task size on data entry on the primary flight display. International Journal of Human–Computer Interaction, 40(22), 7481-7497. https://doi.org/10.1080/10447318.2023.2266247

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Mechanical Device or Touchscreen Widget: The Effects of Input Device and Task Size on Data Entry on the Primary Flight Display

Charles-Antoine Lanoix, Srishti Rawal & Philippe Doyon-Poulin

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2266247

Mechanical Device or Touchscreen Widget: The Effects of Input Device and Task Size on Data Entry on the Primary Flight Display

Charles-Antoine Lanoix[a], Srishti Rawal[b], and Philippe Doyon-Poulin[a]

[a]Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Québec, Canada; [b]Department of Aerospace Engineering, SRM Institute of Science and Technology, Chennai, India

ABSTRACT

Due to their customizability, touchscreens continue to advance as a device of choice when designing aircraft cockpits. Previous studies investigated the effect of turbulence on human performance when using touchscreens, but few have evaluated its performance for realistic aviation-specific tasks. In this study, we compared four touchscreen widgets and three mechanical devices during realistic data input on a primary flight display (PFD). Twenty participants took part in the experiment at a constant level of vibration, while simultaneously completing a secondary tracking task. Results indicated that virtual keypads lead to faster completion time for medium to large changes while keeping error rates low. Rotary knobs were fastest for small changes. Virtual keypads also had lower workload and discomfort compared to rotary knobs and drag-based widgets. We found the completion time to be the most important factor in tracking task performance, which translated in higher precision for keypads. These findings suggest that virtual keypads represent an efficient and secure option for numerical data input at low-to-medium vibration.

KEYWORDS

Human–computer interaction; input device; vibration; aviation; touch interaction

1. Introduction

Touchscreens have been part of the landscape in human-computer interactions for a long time (Mackenzie & Buxton, 1991). But since the turn of the century, they have become ubiquitous mainstream technology. They are found in cell phones and tablets, of course, but also in printers and refrigerators among others (Giebelhausen et al., 2014; Harvey et al., 2011). Their ease of use for people of any age (L.-Y. Lin et al., 2017) is one of their main selling points. From an engineering perspective, they present a new opportunity for renewing and improving upon existing systems, considering their drawbacks and advantages compared to other interaction devices.

Compared to conventional selection devices such as the mouse or trackball, touchscreens suffer from high error rates when targets get smaller than 25 mm or when there is vibration (Cockburn et al., 2017; C. J. Lin et al., 2010; Wang et al., 2022). Furthermore, for the same button size, physical buttons are selected more accurately than virtual buttons (Wobbrock et al., 2008). The distance of the screen from the body can also lead to more physical fatigue since the arm must remain extended for a long period of time (Stanton et al., 2013; Yau et al., 2008).

The main benefit of touchscreens is the ability for users to interact directly with the content presented on the screen. By removing the intermediary (i.e., mouse, physical button, etc.), the intuitiveness and learnability of the systems are greatly improved (Rogers et al., 2005). Touchscreens are faster than other mechanical devices for target selection (Baldus & Patterson, 2008; C. J. Lin et al., 2010; Stanton et al., 2013), data entry (Cockburn et al., 2017; Wynne et al., 2021) and drag-and-drop movements (Wang et al., 2022), even when there is turbulence. It should be noted, however, that this advantage does not translate to tasks requiring continuous and precise adjustment (Rydström et al., 2012).

Because of those reasons, and as accuracy and robustness of touchscreen technologies improved, there has been a renewed interest of integrating them to more safety-critical contexts such as transportation (Grahn & Kujala, 2020; Mayer et al., 2018), health care (Lappalainen, 2011; Yeh, 2020) and aviation (Cantu et al., 2021; Wang et al., 2018). Aviation in particular is an area of active research on this topic, as manufacturers have started migrating from legacy buttons and knobs to touchscreens. This trend can be seen in all aviation sectors: major industry players already designed or integrated touchscreens into their product line for either recreational (Garmin & subsidiaries, 2022), commercial (Rockwell Collins, 2016) or military (Lockheed Martin, 2023) purpose.

Main motivations for this change are to reduce the equipment footprint, offer direct manipulation of information and have user-reconfigurable controls (Hamon et al., 2015; Kaminani, 2011). Moreover, large touchscreens make for less cluttered and more customizable interfaces (Bhalla &

CONTACT Charles-Antoine Lanoix charles-antoine.lanoix@polymtl.ca Department of Mathematics and Industrial Engineering, Polytechnique Montréal, 2500 chemin de Polytechnique, Montréal, QC, H3T 1J4, Canada

Bhalla, 2010). Publicly available data from Boeing (2023) also points towards a small weight reduction (three pounds per two units) when replacing traditional screens with touch-based ones as a retrofit on the 737 MAX and NG. Though a small difference, it contributes to the larger effort to reduce airplane weight and save fuel.

Yet, the integration of touchscreen-based interfaces into an environment, whether for an all-new interface or as a replacement for mechanical devices, should not be taken lightly. This is especially true with safety-critical systems, as erroneous interactions can result in deadly consequences (Mallam et al., 2020; Miller et al., 2019). Orphanides and Nam (2017) identified three specific dimensions to consider for the transition to be successful. First, the choice of touchscreen technology must be appropriate for the task. Secondly, the implementation environment must be considered to ensure that the interface is designed to work well in it. Third, touch interfaces must be designed according to the profile of the users who will use them. While all three pertain to design for aviation, the one most frequently evaluated in research is the impact of the environment.

In the cockpit, the biggest environmental constraint is the presence of vibrations. These can take the form of turbulence (meteorological phenomenon) or vibrations induced by the rotation of propellers, for example in a helicopter (mechanical phenomenon). Furthermore, it is anticipated that climate change will lead to an increase in turbulence in the coming decades (Williams, 2017), which makes vibration all the more important to consider in the design of cockpits. The fixed position of the pilots and screens, the presence of other mechanical controls, and the display brightness are other environmental factors that need to be considered.

1.1. Main contributions

In this study, we evaluated the relative performance of touchscreens versus mechanical devices in a cockpit environment exposed to vibrations. Specifically, participants were presented a realistic input task, i.e., entering digital data such as speed or altitude into a primary flight display (PFD). Seven devices were tested: two virtual keypads, a virtual slider, a direct manipulation-based touch interface, two rotary knobs and a joystick. By segmenting tasks in groups depending on the distance between initial and target values, we could contrast each device's performance with more granularity than previous studies. A secondary tracking task was administered in parallel to the data entry to increase workload and exacerbate differences between devices. We analysed tracking precision in conjunction to eye tracking data, providing insight on the attentional cost of each device and its impact on piloting performance.

2. Related work

2.1. Touchscreens under vibration

Several recent studies have investigated the effect of vibration on the performance of touchscreens compared to other data input devices. Cockburn et al. (2017) tested the touchscreen and trackball for target selection tasks, among others, at several vibration levels. It was performed at two target sizes (0.8 cm and 1.6 cm). Similarly, Wang et al. (2022) performed an experiment with the same design (device × vibration × target size) for touchscreen, mouse, trackball, and remote hand controller. The results of the two papers partially overlap, as the touchscreen - and the mouse in Wang et al. (2022) - was superior to the trackball in completion regardless of vibration. Both also showed an effect of vibration on error rate, the touchscreen leading to more errors during vibration conditions than the mouse and trackball. However, the papers differed in workload results as Cockburn et al. (2017) reported higher frustration for the touchscreen, while Wang et al. (2022) instead found higher mental load and lower frustration for the touchscreen and mouse. This could be explained by the potentially lower vibration levels in the Wang et al. (2022) experiment: they are not quantified, but the variations in completion time from one vibration level to the next are much smaller than for Cockburn et al. (2017).

The most comprehensive study in the field of numerical value input is from Wynne et al. (2021). They investigated how touchscreens and rotary dials compare in inputting random three-digit values. Seven touchscreen widgets were designed for the experiment, some of them based on simple tap gestures (keypad, arrow keys for each digit), others using "tap-and-hold" (arrow keys for whole value), drag (sliders, clockface) or swipe (swipe bars) gestures. They evaluated performance at four vibration levels (static, light chop, light turbulence, moderate turbulence), and 26 participants took part in the experiment. The authors found the keypad to be the fastest of all devices across all turbulence conditions, with arrow keys also being faster than the rotary knob. Widgets based on drag or swipe performed either similarly or worse than the dial, and turbulence exacerbated the difference in completion time. The single tap devices, and particularly the keypad, also got higher workload and usability scores, which makes them the best-suited devices for the task, according to the authors. If another type of gesture was to be used, their data suggests prioritizing drag over swipe as accuracy was lower for the latter. Based on Jeong and Liu (2017), they hypothesized that the lighter movement of the swipe is more impacted by vibration while drag is more robust due to constant and deliberate contact with the screen.

2.2. Evaluation of realistic aviation tasks

The main gap that emerges from the literature is the low amount of research that evaluated data input devices using realistic tasks related to the end-use context, a finding shared by Wynne et al. (2021). Studies by Salmon et al. (2011) and Goode et al. (2012) that focused on the use of touchscreens in military vehicles are good examples of experiments where the tasks accurately represent the actions that users will need to perform in a real-world context. In aviation, there are few such studies. Previous work are mostly limited to experimental tasks such as target selection (Alapetite et al., 2018; Coutts et al., 2019; Schachner & Doyon-Poulin, 2023). Studies that focused on numerical data entry tested the input of random numbers that did not reflect the limitations of airspace parameters such as airspeed or heading (Cockburn et al., 2017; Dodd et al., 2014; Wynne et al., 2021). Avsar et al. (2016), Alapetite et al. (2012) and Rouwhorst et al. (2017) did propose new touchscreen-based interfaces to replace traditional mechanical devices, but neither included vibration level in their experimental design.

Yet, the implementation of the device interaction as well as its coherence with the task to be performed are two primordial factors to consider in the choice of the appropriate device. Consider the paper by Wynne et al. (2021) where for a single device i.e., the touchscreen, different implementations resulted in a range of performance variations for a numerical data entry task. Some implementations were superior to the rotary knob, some were equivalent, and some were inferior. Even though simple target selection tasks are useful to get insight into how devices compare in a crude way, it is only when testing the actual task that we can select which one is best in a particular context.

3. Hypotheses development

Based on previous work, we made five hypotheses regarding the use of input devices.

3.1. Hypothesis 1: Completion time

H1: Keypads will provide the fastest data input, regardless of the distance between initial and target values

Throughout literature, there is a consensus on the speed of the touchscreen compared to various mechanical devices (Cockburn et al., 2017; Lin et al., 2010; Stanton et al., 2013; Wang et al., 2022; Yau et al., 2008). For data entry in particular, the study by Wynne et al. (2021), whose methodology is similar to ours, found the virtual numeric keypad to be significantly faster than rotary knobs and other tactile devices. In fact, it allows constant time input, since the difference between initial and target values has no influence on the number of gestures required to enter data, unlike other devices. These results are also supported by van Zon et al. (2020) who, using Fitts' law task modeling (Fitts, 1954), concluded that a physical keyboard enabled data entry at a higher throughput than a rotary knob and random target selection on touchscreen. Although we will be using a virtual keyboard instead, its advantages (structure, equal key size) should carry over, since the literature tends to indicate that physical and virtual keyboards do not show significant differences in performance (Avsar et al., 2016; Lee & Zhai, 2009).

3.2. Hypothesis 2: Error rate and overshoots

H2: Touchscreen devices, especially keypads, will have a higher error rate, while those based on drag movements will lead to more overshoots

The negative effect of vibration on touchscreen error rates is well documented, for different types of task and vibration profiles (Cockburn et al., 2017; Goode et al., 2012; Lin et al., 2010; Yau et al., 2008). The high freedom of movement and lack of physical feel, coupled with the addition of parasitic movements to the user's arm, seem to systematically lead to lower accuracy compared to physical devices. These problems are reflected to an even greater degree in drag-based tactile interaction mechanisms (Coutts et al., 2019; Wang et al., 2022; Wynne et al., 2021): overshoot of the target value should therefore be more frequent for these. On the other hand, looking at the size of the various interactive elements that make up touch devices, we can assume that numeric keyboards will be at a disadvantage. Since they are made up of smaller keys than the buttons for direct manipulation and the slider button, participants will have to make a trade-off between speed and precision, as described by Fitts' law (Fitts, 1954).

3.3. Hypothesis 3: Discomfort and workload

H3: Mechanical devices will generate less discomfort, but a higher workload than touch devices

Since they require arm extension and do not have a dedicated support, touchscreens are most often characterized as less comfortable to use than mechanical devices (Baldus & Patterson, 2008; Stanton et al., 2013; Yau et al., 2008). Workload, meanwhile, seems to be highly dependent on the device's functional implementation and its consistency with the task. Sometimes, a touch device can lead to more frustration (Cockburn et al., 2017), but it can also be less mentally demanding overall than a mechanical device if it is well implemented. This seems to be the case with the virtual keypad for numeric data entry (Wynne et al., 2021), which explains our hypothesis.

3.4. Hypothesis 4: Tracking task

H4: Tracking task accuracy will be lower for touch devices placed in a virtual panel.

Due to the distance between the virtual panel and the center of the flight screen, participants should look less frequently at the tracking task. Thus, we believe that less sustained attention will lead to lower accuracy.

3.5. Hypothesis 5: Usability

H5: Virtual keypads will be the most usable devices.

Touchscreens, even in the presence of vibration, are considered more usable by users despite sometimes inferior performance compared to mechanical devices (Stanton et al., 2013). More specifically, users seem to regard virtual keyboards as even more usable, in part because of the data entry speed they allow (Wynne et al., 2021). We therefore believe that intuitiveness, familiarity and performance will lead to higher usability scores for these devices.

4. Method

4.1. Participants

20 participants (17 males, 3 females; M = 25.16 yrs, SD = 3.67) took part in the experiment. We determined that participants did not need to be pilots since the tasks would not

require prior knowledge of flying a plane; thus, all but one participant didn't have piloting experience. The participant with piloting experience had a private pilot's license with 200 hours of flight time. Due to the constraints of eye tracking glasses, participants could not wear glasses and needed to have a good vision of a screen at arm's length. People who had back pain or motion sickness were also rejected due to the exposure to vibrations. Both right- and left-handed people were accepted (17 right-handed, 3 left-handed; 15% of left-handed participants) to establish a representative sample of the population and to evaluate the potential effect of handedness on performance. The proportion of left-handed participants in our study is in line with the results of a recent meta-analysis that found that between 9.3% and 18.1% of people could be considered left-handed (Papadatou-Pastou et al., 2020). Participants had a mean height of 178.24 cm (SD = 8.85) and a mean right arm length of 88.14 cm (SD = 4.29). They were also asked about their education level and video game experience (beginner, intermediate, expert) in order to make finer analyses of the results. The study was approved by Polytechnique Montreal Research Ethics Board (CER-2122-55-D) and all participants signed an informed consent form. No financial compensation was given.

4.2. Tasks

4.2.1. Data entry

The main task that the participants did was the entry of numerical values on a primary flight display (PFD), the main flight display of an aircraft, see Figure 1. This display shows the aircraft's performance information: heading (HDG), altitude (ALT), vertical speed (VS), indicated airspeed (IAS), and altimeter setting (calibration of altitude value according to atmospheric pressure; Baro). Heading is traditionally displayed as a compass, while ALT, VS and IAS take the form of scrolling tapes. Baro, on the other hand, is displayed as a simple value under the ALT tape. The current value of each field is shown in the readout box with black background.

To change flight path, the crew changes the target value for the related field i.e., the value the aircraft needs to reach to follow the flight path. To do this, the crew must edit the cyan values above the tapes for IAS, ALT and VS, or left of the HDG compass. These are the four numbers that must be entered to set the target values: the tasks of the experiment consisted in setting those four values, plus the altimeter setting.

To successfully complete a task, participants had to select the correct field, enter the requested target value and confirm the entry. In case of an erroneous entry, an audio prompt was played and the entry field remained in edition mode until the correct value was entered. Figure 2 illustrates this process.

The tasks were categorized according to the distance between the current value and the target value that must be reached to complete it: from this point on, this will be referred to as the *size* of each task. This helped determine whether certain devices were faster when the change to be made was larger or smaller. For each field, participants had to successfully complete a task involving a large, medium, and small change. Since the fields had different units and range of values, we defined the intervals for each field according to boundaries that reflected a semi-realistic use of an aircraft (see Table 1).

4.2.2. Tracking task

In parallel to the PFD data entry, participants executed a tracking task using a joystick with their left hand. This task, inspired by the MATB-II software tracking task (National Aeronautics and Space Administration [NASA], n.d.), consisted of using the joystick to move a yellow cross in the middle of the screen to track a purple square moving randomly (see the center of the PFD shown in Figure 1). The position of the cross and the square were recorded 33 times per second and the distance between the two points was calculated using the root mean square (RMS) formula. During tasks, the distances were averaged to a single measurement for each segment (field selection and data entry). The use of a secondary task was intended to increase the workload felt by the participants, reflecting more realistically the divided

Figure 1. The experiment's primary flight display.

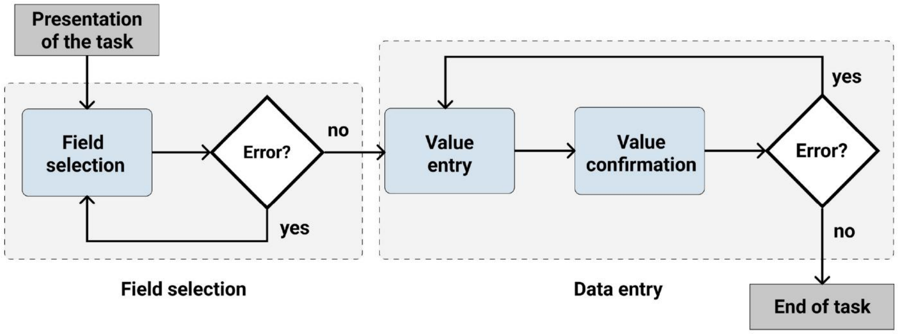

Figure 2. Sequence of data entry task completion.

Table 1. Range of values for every task size and field.

	Size of task		
Field	Small	Medium	Large
Indicated airspeed (IAS)	1–10 kts	11–75kts	76–100 kts
Altitude (ALT)	1–1000 ft	1001–15000 ft	15001–18000 ft
Vertical speed (VS)	1–500 ft/min	501–1500 ft/min	1501–2000 ft/min
Altimeter setting (Baro)	0.01–0.10 inHg	0.11–1.00 inHg	1.01–1.80 inHg
Heading (HDG)	1–10 deg	11–160 deg	161–180 deg

attention nature of data entry and flying an aircraft simultaneously. It should be noted that for this reason, the experiment is more akin to flying manually a general aviation or military aircraft, as opposed to commercial aircraft pilots who spend most of their flight time with the assistance of autopilot.

4.3. Materials

4.3.1. Flight deck rig

To simulate the turbulence of an aircraft cockpit, we used a D-Box haptic seat (model GP Pro 500) producing vibrations with 3 electric motors. The vibration profile was taken from a recording made on board a Bell-412 helicopter during a level flight at 120 kt. Most of the acceleration in the profile came from frequencies between 1–30Hz, with a magnitude of 0.715 m/s^2 [see (Schachner & Doyon-Poulin, 2023) for details]. Note that all experimental trials were done with the vibration active.

A rack made of aluminum rails was attached to the seat, on which could be installed screens and other equipment. A system of straps also allowed the part of the rack holding the screen to be moved vertically in order to adjust it to the height of each participant, as shown in Figure 3.

The screen we used was a Planar PCT2435, a 24" 16:9 capacitive touchscreen with a resolution of 1920:1080. It was fixed in front of the participant at an angle of 15° below the vertical. This corresponds to the position of flight instruments on the Main Instrument Panel (MIP) in most aircraft. Figure 4 shows the distances based on the participant's Eye Reference Point (ERP) that ensured a consistent eye-to-screen distance between participants. A Logitech Attack 3 joystick was also secured to the aluminum frame, accessible for use with the left hand. The movement of the stick was used to control the tracking task cursor.

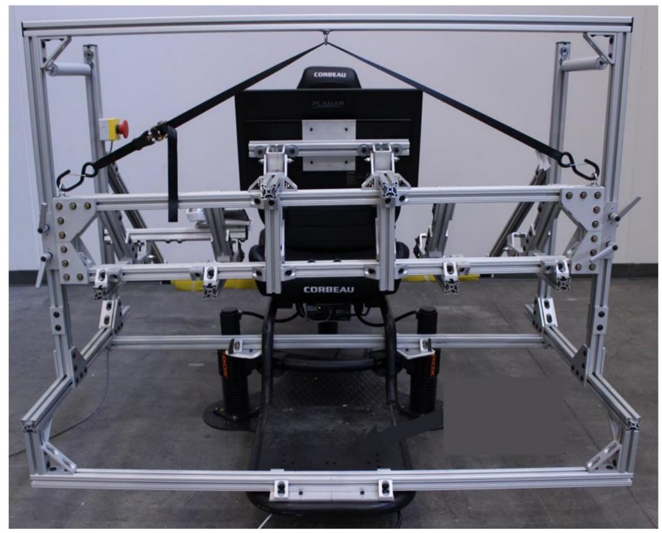

Figure 3. Vibrating seat structure.

4.3.2. Software

An application made with the Unity game engine (version 2020.3.32f1) was designed for the experiment. The screen size used was 8 inches by 20 inches to replicate the exact size of a large area display (LAD; https://www.intellisenseinc.com/products/rugged-displays/large-area-display-20-x-8/). Since the Planar display area was larger, only the lower portion of the screen was used and the top portion was covered with a plastic protector, thus ensuring the same display area as the LAD. Visually, the software reproduced a generic primary flight display. The PFD took up the entire height of the screen (8 inches) and 42% of the width (13 inches). On each side of the PFD was a black space. The task instruction was displayed at the top of the space on the right. As explained in the next part, some devices used the remaining space in the right section. The section on the left was not used. In addition to the visual interface, the software carried out the generation of the tasks, the tracking task, and the data logging.

4.3.3. Devices

Seven devices were tested: three mechanical and four touchscreen-based. The following section describes each one. Readers are referred to the video link to see every device in action (https://youtu.be/VggycbMccgM), while Figure 5 depicts a static view of each of them. The mechanical ones were selected as a baseline, as they represent the way data is normally entered in most cockpits. Keypads seemed like the most promising devices as was reported in a previous study (Wynne et al., 2021), while direct manipulation and the slider were chosen because of their drag-based interaction and intuitiveness.

4.3.3.1. Joystick. The joystick used was the same Logitech joystick as the tracking task. Using the buttons on top of the joystick similarly to arrow keys on a keyboard, participants navigated between the PFD fields and changed their value. The trigger on the back of the joystick allowed them to confirm the selection of the field as well as the value entered.

4.3.3.2. Rotary on screen. A Grayhill rotary encoder (62AG22-L5-P) was attached to the lower right side of the screen, perpendicular to its surface, similar to the LAD arrangement mentioned earlier. It had 16 pulses per rotation, a rotational torque of 3.5 ± 1.4 in-oz and a pushbutton with an actuation force of 510 ± 150 gr. The rotation allowed to navigate circularly between the PFD fields as well as to modify the value. At low rotational speed, a pulse allowed to increment or decrement the value by the smallest possible unit e.g., 1 kt. To increase the performance of the rotary, if a pulse was detected less than 0.01 seconds since the last pulse, it produced a larger increment or decrement e.g., 5 kts. In the same way as the joystick, the button integrated in the rotary was used to confirm the field and the value.

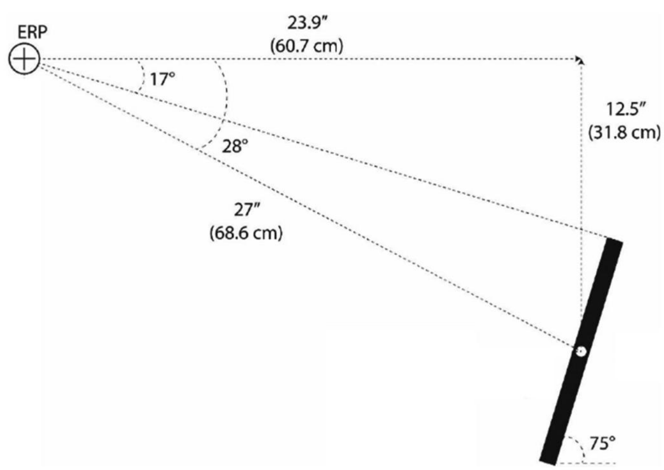

Figure 4. Diagram of eye reference point (ERP).

4.3.3.3. Rotary on pedestal. Another rotary encoder was installed on the right side of the main platform, mirroring the position the joystick, perpendicular to the ground. It was a Grayhill rotary encoder (60A18-4-RAC), 20 pulses per rotation, 2.0 ± 1.0 in-oz of rotational torque, including a pushbutton requiring 400 ± 150 gr of force to activate. In addition to being a button, the shaft could be moved in 4 directions (up-down-left-right) like a small joystick. This differed from the previous rotary in that the shaft was used to navigate between fields instead of the knob rotation. The rotation was still used to change the value, as was the button that confirmed the field and the value.

4.3.3.4. Keypad on PFD. The keypad was designed according to the layout of a telephone keypad, with a delete key and a confirmation key on the bottom row. A key to invert the sign of the displayed value was placed next to the readout of the keypad. All keys were square shaped with a 0.6 inch side, the smallest clickable size according to MIL-STD-1472 (MIL-STD-1472H, 2020), a recognized standard in cockpit design. For this device, field selection was made by clicking directly on the PFD tape associated with the field whose value needed to be changed. The keypad then appeared in the lower right corner of the PFD to enter the value i.e., co-located.

4.3.3.5. Direct manipulation. The name of this device comes from the fact that the participant had to directly manipulate the tapes on the PFD to change their value. That meant scrolling the ALT, VS and IAS tapes to change their values.

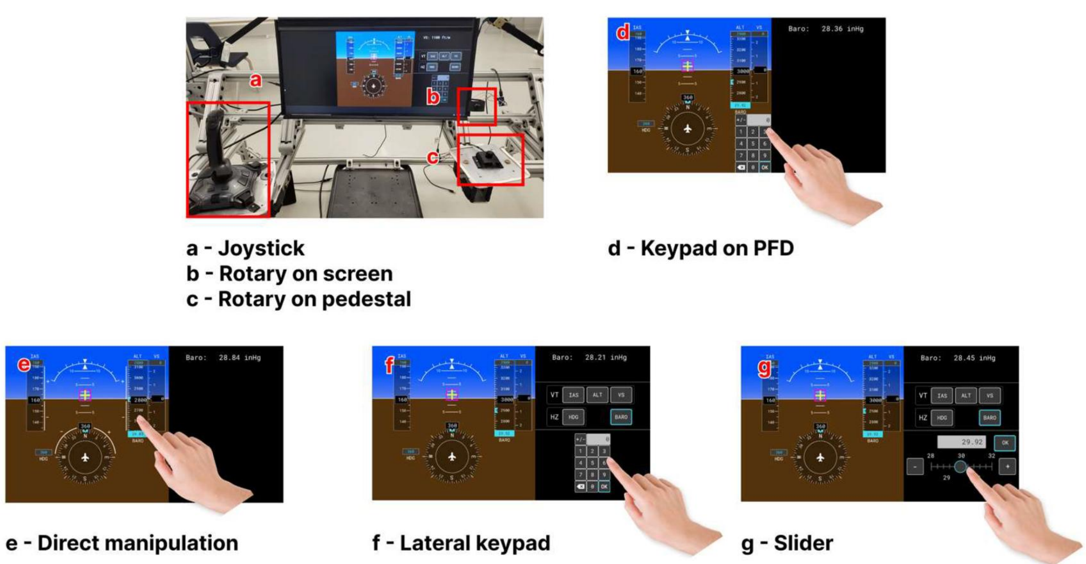

Figure 5. Visual displays of devices used for number entry.

For the BARO, they had to first click on the BARO readout, then scroll the ALT tape to change the value. For the HDG, the movement was not a scroll, but rather a rotation to change the angle of the associated compass. In all cases, a click on the field readout confirmed the value. In addition to the larger scroll or rotation movement, it was possible to click on the upper (left for HDG) or lower (right for HDG) part of the tape to increment or decrement its value by the smallest possible unit, respectively. The idea of directly manipulating elements of a flight display can be found in a previous experiment by Alapetite et al. (2012).

4.3.3.6. Side keypad. For this device, the same keypad as 2.3.3.4 was used, but was instead placed in the free space to the right of the PFD. The field selection, instead of being done by clicking on the PFD, was done through buttons located just above the keypad. This design reflected the layout of a dedicated flight mode selector on general aviation aircraft.

4.3.3.7. Slider. The slider was positioned in the same way as the lateral keypad, on the right side of the PFD, under a group of buttons to select the field. The slider could be interacted with in the same way as found on commercial electronics: the participants moved the pointer along a straight line representing the set of possible values. In addition, increment buttons were provided on each side of the slider to edit the value with more precision.

4.3.4. Questionnaires

To measure participants' workload after using each device, we used the NASA-TLX questionnaire. It is a well-established subjective method using a multidimensional rating scale (Hart & Staveland, 1988). This questionnaire assesses factors related to the nature of the task, i.e., physical demand, mental demand, and temporal demand, and related to the operator, i.e., his or her judgment of his or her performance, the effort required to perform the task, and his or her level of frustration.

The evaluation of the discomfort of the devices was performed using modified Borg CR-10 scales. This scale allows a subjective evaluation of the intensity of an individual's experience (Borg, 1998). It extends over a range of values between 0 (no sensation) and 10 (maximum sensation), enhanced by a level 11 and a level "as high as possible": these last two levels were, however, not presented to participants as we felt they were not needed. This does not change the validity of the scale as it's still self-encompassing (from no sensation to maximum): the perception of lower levels does not change, thus the scale is merely truncated. Decimal values are inserted between whole numbers from 0 to 3 in order to characterize the sensations with more granularity. After each device, participants were asked to indicate the discomfort felt at five points: right shoulder, right arm, right hand and wrist, neck, upper back, and lower back. For the joystick device, discomfort in the left shoulder, arm, and hand were requested instead.

An ad hoc form was designed to evaluate the usability of each device. The first part of the form consisted of 3 Likert scales ranging from 1 to 5. Participants were asked to indicate, for each device, the ease-of-use to make small changes (very easy to very difficult) and large changes (very easy to very difficult), as well as the perceived speed of completing tasks (very fast to very long). In the second part, participants listed the features they liked and disliked about the device, as well as any other relevant comments. Finally, a ranking of appreciation of the devices was completed as the experiment progressed until a final ranking of the seven devices was obtained. We chose to create our own usability questionnaire instead of the often-used SUS questionnaire as it includes several questions that are not related to a work-centric system. For example, we do not need to know about the intent to use the system frequently or the need for external support. Still, some of our questions were inspired by the SUS.

4.3.5. Eye tracking glasses

To record the participants' gaze during the experiment, we used Pupil Invisible eye tracking glasses from Pupil Labs. The eye-tracking technology is integrated into a pair of standard-looking glasses measuring 144 mm wide, 48 mm high and 160 mm long. A camera is set up on the outside frame of the glasses to record the participant's field of view (30 Hz), and infrared sensor is integrated into the inside of the frame to measure the direction of the user's gaze (200 Hz). Data recording was done through the Android application Pupil Invisble Companion, then imported directly to the cloud through the Pupil Cloud service.

4.4. Design

The experiment was designed in a repeated measures format, with device (joystick, rotary on screen, rotary on pedestal, keypad on PFD, direct manipulation, lateral keypad, slider) as the independent variable. For analysis purposes, the tasks were further divided by field (HDG, IAS, ALT, VS, BARO) and then by the distance between the current and target values (small, medium, large). The order of presentation of devices and tasks was randomized. The dependent variables consisted of several objective and subjective measures. The objective measures that were taken are field selection completion time and value entry time (combined into total task completion time), number of field selection errors, number of data entry confirmation errors, tracking accuracy during navigation, tracking accuracy during data entry, number of target value overshoots, as well as percentage of fixations on the tracking task. All these measures were recorded for each task, except for the percentage of fixations which was calculated for each device. Subjective measures were also calculated for each device: workload (with NASA-TLX), discomfort (with Borg CR-10) and usability (with ad hoc questionnaire). We are aware that our study doesn't take into some environmental factors into account, such as environment and display brightness, and that they should be investigated in a future study.

4.5. Procedure

Upon arrival at the test room, we measured the participant's stature (height and length of the right arm) and they answered some demographic questions. The rig was then set up to the participant's morphology by adjusting the height of the screen and the position of the seat on its rails to reach the ERP. Then, the devices were explained one by one to the participants, who could perform up to 2 minutes of practice tasks to familiarize with the device. They also received explanations on how the eye tracker app worked since they were the ones who would start and stop the recordings.

For each of the 7 devices, participants then completed all 15 possible tasks (5 fields × 3 sizes). 5 to 10 seconds after the completion of a task, a new one was randomly generated, and the instruction was displayed in the upper portion of the empty space to the right of the PFD. For all touch devices, participants were asked to use their index finder for selection and to not stabilize their hand by leaning on one side of the screen. After having completed the trial for each device, participants completed the NASA-TLX workload questionnaire and Borg CR-10 scales discomfort form, and verbally answered usability questions.

Once data collection for all devices was completed, participants were thanked and escorted out. The total duration of the test was of 2 hours.

4.6. Data analysis

We used repeated-measures ANOVAs to analyze temporal data, deviation from the target during the tracking task, and percentage of fixation time looking at the tracking task. Greenhouse-Geiser correction was applied where sphericity was violated. Post-hoc analysis was done using pairwise t-tests. We used Friedman's tests to analyze the number of errors, overshoots, NASA-TLX results, discomfort and usability questionnaires as they produced ordinal level data. Conover's tests of multiple comparisons were used for post-hoc analyses. In both parametric and non-parametric analyses, the post-hoc tests were adjusted using the Holm-Bonferroni method. The entire analysis was done on R (v. 4.2.0) and statistical significance was reported when $p < .05$.

Due to bugs in the task generation software, 46 data points out of the expected 2100 (20 participants × 7 devices × 5 fields × 3 sizes; 2.19%) were missing. The "Multivariate Imputation by Chained Equations" (MICE) R package was used to impute missing data via predictive mean matching (PMM). This only applied to variables normally logged by the software, whereas workload, discomfort and usability forms were completed by all participants. Some data was also missing for eye tracker. No data was recorded for the last participant, so the analysis was done on the data of the remaining 19 participants. Whether due to technical problems or oversight, the eye tracker recordings of 15 out of the expected 133 test sessions (20 participants × 7 devices; 11.28%) were unavailable. Data imputation was also performed by PMM using MICE. The analysis itself consisted in defining areas of interest (AOI) in the participant's field of view (specific sections of the interface, joystick, rotaries, etc.). We then had to associate the gaze position with one of these AOI at a frequency of 5 times per second. Since the analysis software did not allow for segmenting the recordings for every task, fixations that occurred between tasks were assigned to an "irrelevant" AOI whose content was subsequently removed from the analysis. We then could get the time spent looking at each AOI for the entirety of the session, minus the between-task parts.

5. Results

5.1. Completion time

5.1.1. Field selection time

The device used had a significant effect on the field selection time, $F(6, 114) = 43.24$, $p < .001$ (see Figure 6). The mechanical devices, although different in the way navigation was done, were not significantly different between each other for field selection time. However, they were all slower than the touchscreen devices (joystick: $p < .01$, rotary screen: $p < .001$, rotary pedestal: $p < .01$). Among the touchscreen devices, there were two categories of field selection: tapping on the PFD (keypad on PFD and direct manipulation) and tapping on the side panel (lateral keypad and slider). Field selection time with the slider was not significantly faster than the keypad on PFD ($p = .24$) and direct manipulation ($p = .06$), but the lateral keypad was faster than both ($p < .01$).

5.1.2. Data entry time

For data entry time, we observed that the device also produced a significant effect, $F(6, 114) = 47.27$, $p < .001$ (see Figure 7). Three categories emerged: keypads (keypad on PFD and lateral keypad), intermediate devices (joystick, rotary on screen, rotary on pedestal, slider), and direct

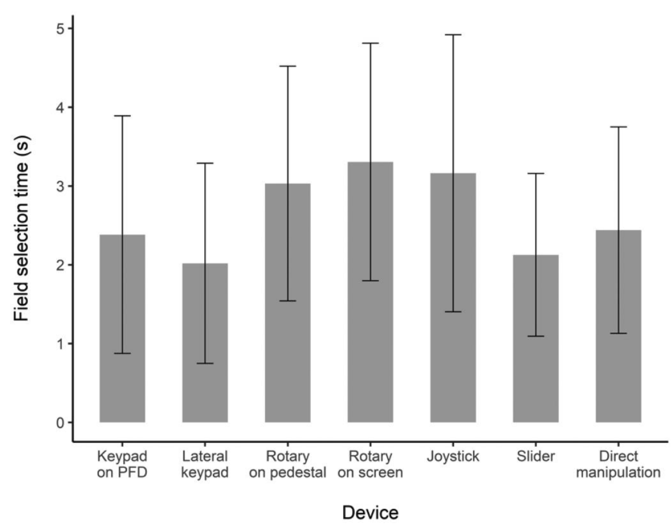

Figure 6. Field selection time across devices. Error bars represent ±1 SEM.

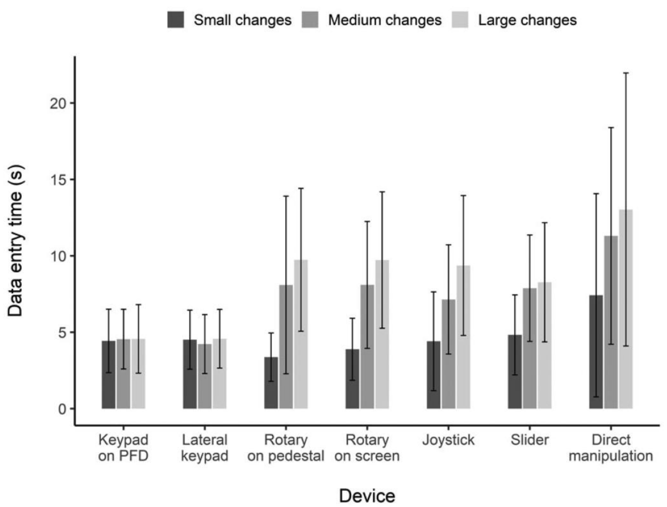

Figure 7. Completion time of data entry across devices and task size. Error bars represent ±1 SEM.

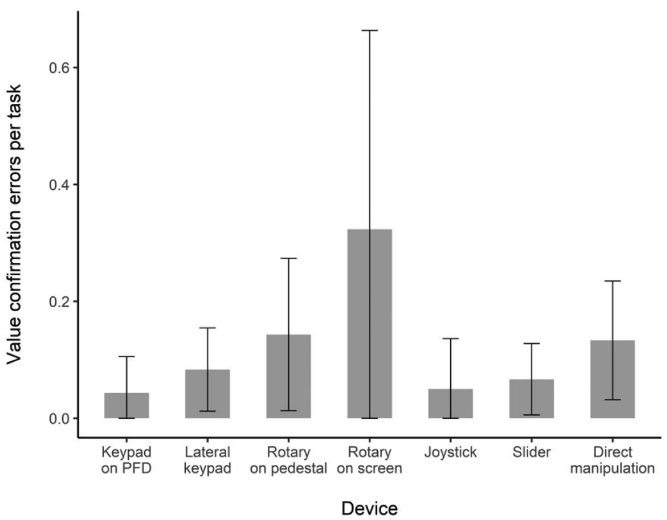

Figure 8. Number of value confirmation errors per task across devices. Error bars represent ±1 SEM.

manipulation. The latter device was found to be significantly slower than all other ones ($p < .001$). No significant difference was found between the devices in the other two categories, and the keypads were significantly faster than the intermediate devices ($p < .01$). In addition to the device effect, we investigated the effect of participants' video game experience on data entry time. They were then divided in two groups: those who self-assessed as beginner or intermediate gamers ($n = 9$), and those describing themselves as experts ($n = 11$). We then found that self-described expert gamers were significantly faster than less experienced gamers ($t(16.65) = 2.18$, $p < .05$). We also tested the effect of handedness on completion time, but found it to be not significant ($t(3.14) = 1.27$, $p = .29$).

We analyzed data entry time based on the distance between the current and target value (small, medium, large), also shown in Figure 7. When looking at small changes, we found a significant effect of device on data entry time ($F(6, 114) = 13.28$, $p < .001$). Only two devices stood out from the pack: direct manipulation was significantly slower than all other devices ($p < .01$), and the rotary on pedestal was faster than all touch devices ($p < .05$). For the medium changes, the effect of the device on data entry time was the same as for all tasks ($F(6, 144) = 48.99$, $p < .001$). Data entry time for large changes, meanwhile, also varied significantly by device ($F(6, 114) = 85.39$, $p < .001$), but with some differences. Direct manipulation was still significantly slower than all other devices ($p < .01$) and keypads significantly faster than all other devices ($p < .05$). For large changes, we also saw that the slider was significantly faster than both rotaries ($p < .05$). The combined time of selection and data entry was also significantly affected by the device ($F(6, 114) = 57.95$, $p < .001$). The results were almost identical to those presented previously, except that the slider was faster than the rotary on screen ($p < .05$).

5.2. Error rate and overshoots

The devices did not generate significant differences in field selection error rates.

For value confirmation errors, there was a significant effect of device, $X^2(6) = 40.86$, $p < .001$ (see Figure 8). The rotary on screen led to significantly more errors than the joystick, keypads, and slider ($p < .01$).

An overshoot was observed when the current value momentarily exceeded the target value before the participant confirmed the entry. Several overshoots could have occurred for one entry, for example if the current value went above, then below the target during adjustments. For this analysis, we excluded keypads as they were unable to produce overshoot. We found a significant effect of the device on the number of overshoots, $X^2(4) = 34.93$, $p < .001$ (see Figure 9). Significantly fewer overshoots were made with the rotaries than direct manipulation ($p < .001$) and the slider ($p < .01$). The joystick was also significantly more accurate than direct manipulation ($p < .05$).

5.3. Discomfort and workload

There was a significant effect of the device on discomfort, $X^2(6) = 25.53$, $p < .001$ (see Figure 10). The joystick received significantly higher overall discomfort score than the rotary on screen ($p < .01$) and direct manipulation ($p < .01$).

When looking at discomfort scores for specific body parts, we found no significant effect of the device used on discomfort to the neck ($X^2(6) = 8.94$, $p = .18$), upper back ($X^2(6) = 8.41$, $p = .21$), lower back ($X^2(6) = 5.85$, $p = .44$) and hand/wrist ($X^2(6) = 4.71$, $p = .58$). However, arm ($X^2(6) = 27.92$, $p < .001$) and shoulder ($X^2(6) = 16.88$, $p < .01$) discomfort were significantly affected by the device (see Figure 11). For the arm, the joystick was significantly more comfortable than the rotary on the screen, the keypad on PFD, and direct manipulation ($p < .001$, $p < .05$, $p < .01$, respectively). For the shoulder, the discomfort experienced with the rotary on screen was significantly higher than with the joystick ($p < .01$).

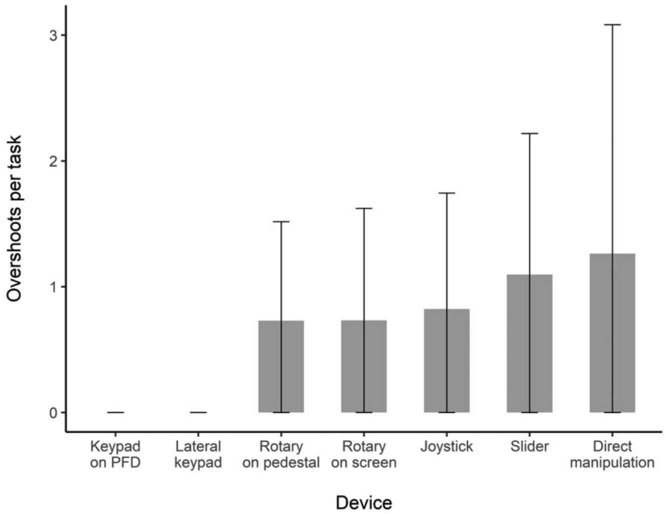

Figure 9. Number of overshoots per task across devices. Error bars represent ±1 SEM.

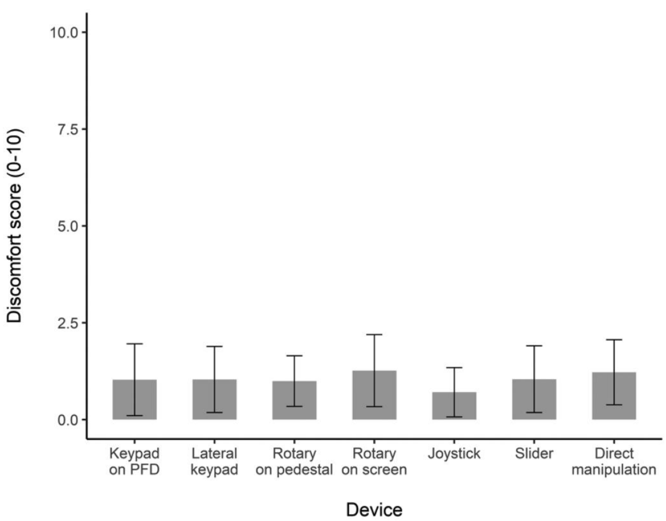

Figure 10. Overall discomfort score across devices. Error bars represent ±1 SEM.

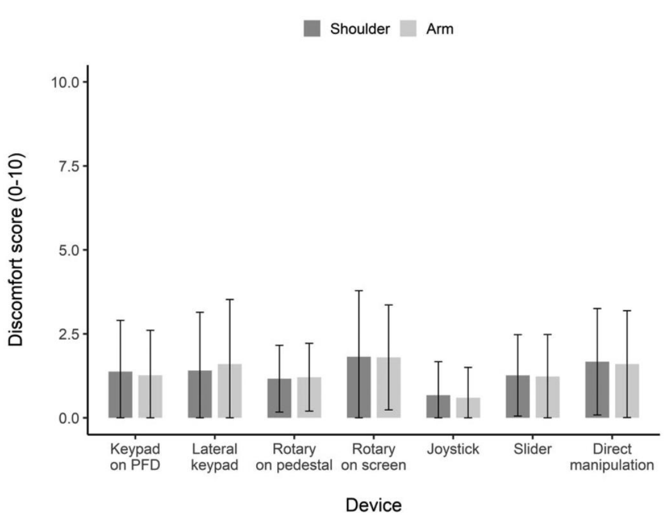

Figure 11. Discomfort score across devices and body part (shoulder and arm only), on a scale of 1 to 10 (1: No discomfort; 10: Extremely strong discomfort). Error bars represent ±1 SEM.

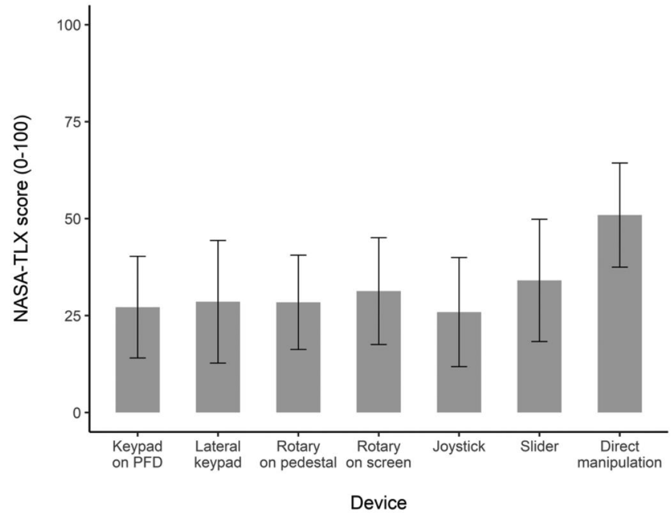

Figure 12. NASA-TLX workload score across devices. Error bars represent ±1 SEM.

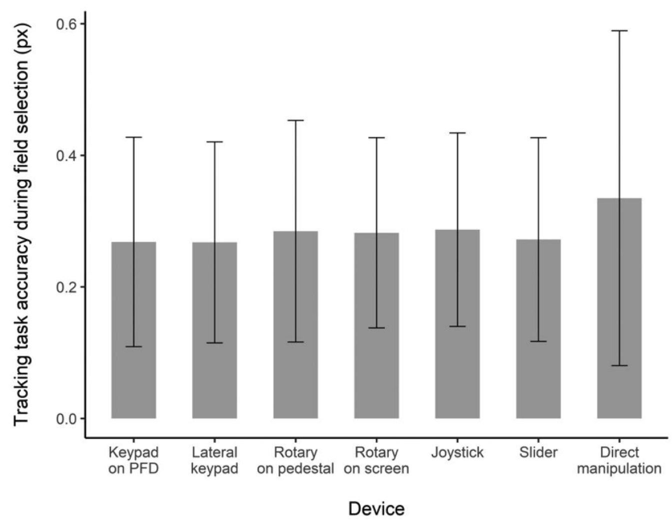

Figure 13. Tracking task accuracy during field selection across devices. Error bars represent ±1 SEM.

The effect of devices was also significant on workload, $X^2(6) = 17.96$, $p < .01$ (see Figure 12). NASA-TLX scores were significantly higher for direct manipulation compared to the joystick, rotary on pedestal, and both keypads ($p < .01$). When looking at each NASA-TLX dimensions, we found direct manipulation to impose higher demand on all dimensions, except physical demand. For this dimension, we observed that the rotary on screen was significantly more physically demanding than the joystick and the lateral keypad ($p < .05$). All other device combinations did not show significant differences.

5.4. Tracking task

During field selection, the device produced a significant effect on performance of the tracking task, $F(6, 114) = 3.08$, $p < .01$ (see Figure 13). Indeed, tracking accuracy when using direct manipulation was significantly worse than when using the slider ($p < .01$).

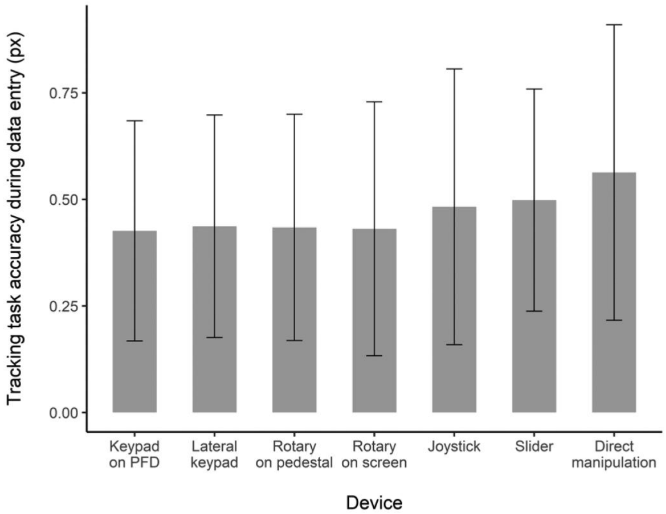

Figure 14. Tracking task accuracy during data entry across devices. Error bars represent ±1 SEM.

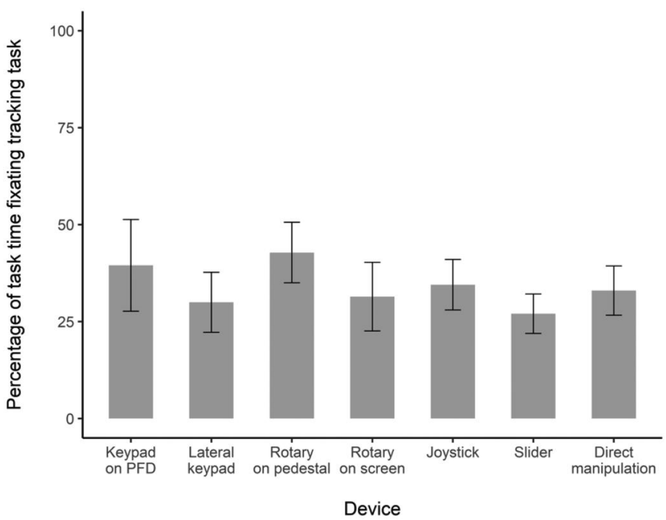

Figure 15. Percentage of task time fixating tracking task across devices. Error bars represent ±1 SEM.

The device also influenced tracking accuracy during data entry, F(3.1, 66.18) = 0.58, $p < .001$ (see Figure 14). Direct manipulation generated significantly worse tracking performance than all mechanical devices and keypads ($p < .05$), while the performance using the slider was significantly worse than the performance using the rotary on pedestal ($p < .05$).

5.4.1. Eye tracking

The device produced a significant effect on the percentage of fixation time allocated to the tracking task, F(3.49, 62.79) = 0.58. $p < .001$ (see Figure 15). Participants looked significantly less at the tracking task when using the slider compared to the joystick ($p < .05$), rotary on pedestal ($p < .001$) and keypad on PFD ($p < .01$). In addition to the slider, the rotary on pedestal led to a greater percentage of fixations on the tracking task than the joystick ($p < .05$), rotary on screen ($p < .01$), direct manipulation ($p < .01$), and lateral keypad ($p < .001$).

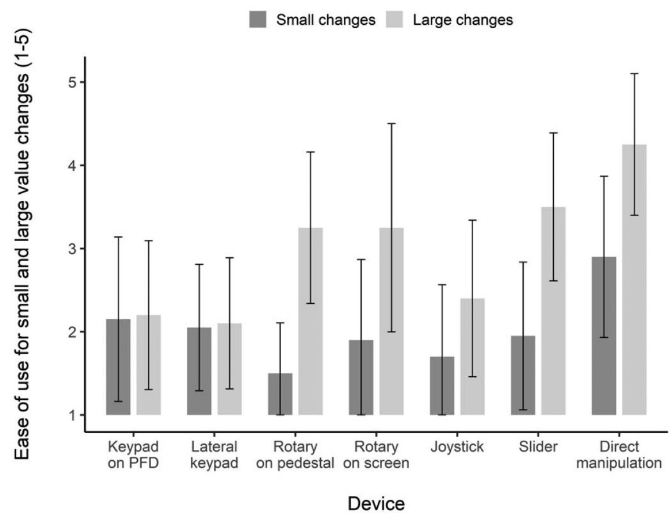

Figure 16. Ease of use score of devices on a Likert scale of 1 to 5 (1: Very easy, 5: Very difficult) for small and large value changes. Error bars represent ±1 SEM.

5.5. Usability

5.5.1. Ease-of-use

The perceived ease-of-use of the devices showed significant differences for both small changes ($X^2(6) = 25.57$, $p < .001$) and large changes ($X^2(6) = 55.01$, $p < .001$), see Figure 16. Participants found small changes significantly more difficult to make using direct manipulation compared to joystick ($p < .01$) and rotary on pedestal ($p < .001$). For large changes, direct manipulation was similarly significantly less easy to use than the joystick ($p < .001$), keypad on PFD ($p < .001$), and lateral keypad ($p < .001$). At the other end of the spectrum, the lateral keypad was also rated as being significantly easier to use than the rotary on screen ($p < .05$), rotary on pedestal ($p < .05$), and slider ($p < .01$). The keypad on PFD was also significantly easier to use than the slider for large changes ($p < .05$).

5.5.2. Perceived task completion speed

We found a significant effect of the device used on the perceived speed of task completion, $X^2(6) = 49.09$, $p < .001$ (see Figure 17). The perceived speed when using direct manipulation was significantly slower than all other devices except the slider ($p < .05$), which was in turn perceived as significantly slower than the lateral keypad ($p < .01$).

5.5.3. Appreciation ranking

There also was a significant difference between the devices in the general appreciation ranking, $X^2(6) = 40.82$, $p < .001$ (see Figure 18). The pattern of significant differences was the same as for perceived speed, with direct manipulation significantly less liked than the joystick ($p < .001$), rotary on screen ($p < .05$), rotary on pedestal ($p < .001$), keypad on PFD ($p < .001$), and lateral keypad ($p < .001$), while the slider was also significantly less liked than the lateral keypad ($p < .05$).

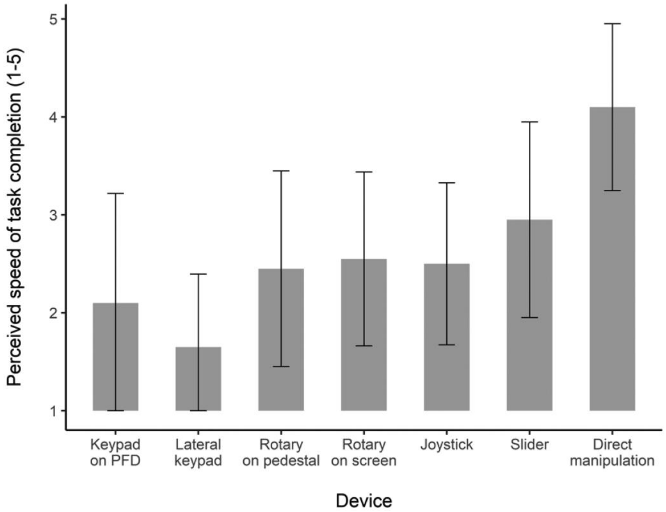

Figure 17. Perceived task completion speed by devices on a Likert scale of 1 to 5 (1: Very fast, 5: Very slow). Error bars represent ±1 SEM.

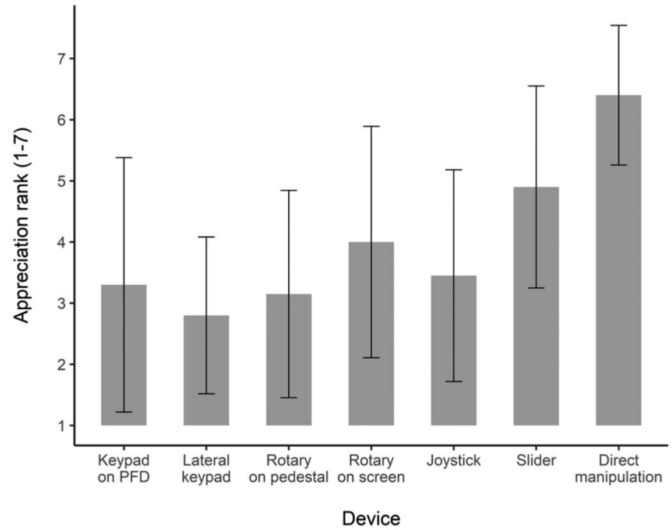

Figure 18. Ranking of devices on a scale of 1 to 7 (1: Most appreciated, 7: Least appreciated). Error bars represent ±1 SEM.

6. Discussion

The goal of this experiment was to evaluate the performance of touch devices compared to mechanical devices historically used in cockpits for digit entry. Our study was intended to be a continuation of the Wynne et al. (2021) paper: the aim of the experiment was the same, but the method was adapted according to the recommendations they issued. First, the tasks performed by participants were realistic and closely replicated the action of entering values into a PFD. Second, tasks were divided by the distance between the current value and the target value, which allowed for a finer-grained analysis of device performance. Finally, the tests were performed with right- and left-handed participants to test the impact of handedness on performance; on this point, the results tended to indicate that being right- or left-handed had no influence on completion time, workload, or discomfort. For ease of understanding, the rest of the discussion will be divided according to the variables of the experiment.

6.1. Completion time

6.1.1. Field selection time

We found that the field selection time was significantly faster for touch devices than for mechanical devices. This appears to be because touchscreen selection was always done with a single movement (a tap on the PFD or lateral panel) as opposed to mechanical devices which potentially required multiple cursor movements to reach the desired field, depending which field was initially highlighted.

For touch devices, those that used buttons on a lateral panel generally appeared to be faster than those requiring direct interaction with the PFD. This can be explained by the fact that the buttons on the lateral panel were located directly below the task statement and were grouped together. This made the eye-travel and arm movement shorter and consistent from task to task. However, the decision to integrate a lateral panel into a cockpit display should not be made solely based on this aspect: the results also showed that participants looked less at the tracking task with the lateral panel devices than with the PFD-based ones. This will be discussed in section 6.4.

6.1.2. Data entry time

H1 was mostly confirmed. Indeed, keypads were generally faster, consistent with previous studies (Coutts et al., 2019; Wynne et al., 2021). They were also clearly faster in the case where the distance between the target value and the initial value was large. However, the assumption that keypads and rotaries would be the fastest for small value changes did not hold: rotaries were significantly faster in this case. The natural explanation is that the time to enter a number with a keypad does not depend on the initial value, while rotaries do. It implies that rotaries start with an advantage for small changes (10 increment steps or less), which erodes as the target value moves away from the initial value i.e., larger task size. In the case of large changes, even the slider was faster than the rotaries since it could cover a larger range of values in a shorter time. In this study, we implemented a rate-aiding feature to the rotaries to accelerate the increments with large changes. However, 9 out of 20 participants found that its use was not reliable enough and could be improved. Our results suggested that even with additional calibrations, keypads would retain their speed advantage over rotaries for large value changes.

It's also worth noting that keypads provided the fastest data entry despite the vibration, the prohibition on stabilizing the hand, and the minimal key size. As past research showed how vibration degraded keypad performance (Coutts et al., 2019; M. A. Smith et al., 2020), we can hypothesize that the performance gap between keypads and other devices would be even bigger in a stable, non-vibrating environment.

6.2. Errors and overshoots

First, it should be noted that the increase in errors we expected for keypads did not occur, as the number of errors per task was very low (on average less than 0.4% for all devices). This comes from our definition of an error: an attempt to confirm an erroneous value. This happened as a result of misreading the task or the field value, or from an accidental activation of the confirmation mechanism. As such, it did not include all keypad missed presses, since an erroneous entry could be corrected before confirmation. The reasoning behind choosing such a definition was to monitor serious mistakes with real consequences on the flight safety. Conversely, inputs made before confirmation could be corrected and did not impact the movement of the aircraft. This confirmation mechanism was suggested in several previous articles (Hamon et al., 2014; M. Smith et al., 2018; Wang et al., 2018; Wynne et al., 2021) to prevent errors when using touchscreens under vibration.

Second, it is important to qualify the results obtained for the number of errors. The larger number of errors observed with the rotary on screen compared to the other devices does not reflect a generalized failure of all rotaries; it mainly shows a hardware problem with our setup. Indeed, the 510 ± 150 gr actuation force of the button was insufficient to prevent inadvertent actuations during data entry. One interesting point is that despite its lower actuation force (400 ± 150 g), fewer errors were made with the rotary on pedestal. This means that the position of the rotary had an effect on the number of accidental clicks. Whether this was due to the extension of the arm and therefore the weaker support, or simply the angle of the device which involved a different finger grip, or a combination of both, we saw that the rotary on screen led to more unwanted clicks. Still, the rotary on pedestal was the device that led to the second-most errors. Future work should investigate higher actuation force for rotary to prevent inadvertent actuation.

While the errors produced with the rotary could be greatly reduced with another model, the problem of direct manipulation is more fundamental. In order to keep the PFD interface uncluttered, the value readouts of each field acted as confirmation buttons. Since they were located at the end of the scrolling tapes (except for HDG), participants would sometime inadvertently press them while trying to scroll. To eliminate this source of error, we suggest adding a confirmation button placed elsewhere on the screen e.g., in the lower right corner. With such a mechanism, it is likely that the error rate for direct manipulation would be similar to that of other touch devices.

We used the number of overshoots per task i.e., the number of times the current value exceeded the target value before being confirmed, to analyze the device accuracy for data entry. Keypads were not included in the calculation of this variable since they could not produce overshoot. Of the devices analyzed, the direct manipulation produced the most overshoots, with the difference between the slider being not significant. This confirms our original hypothesis, H2, as we found that using dragging motions (slide or scroll) on a touchscreen in a vibrating environment led to lower precision. Our results showed that the rotaries allowed for more accurate and therefore less time-consuming and frustrating input.

6.3. Discomfort and workload

Overall, the NASA-TLX scores obtained were low. Compared to the scores obtained in other tracking experiments, as surveyed in a recent meta-analysis (Grier, 2015), the workload scores for each device were in the first (both keypads, both rotaries, joystick and slider) or second (direct manipulation) quartile. This suggests that our tracking task was easier than 50% of the other 70 experiments surveyed by Grier (2015). Although tracking was integral to our experiment, we emphasized to participants to focus on the data entry task. It also implies that data entry did not have a great impact on workload when combined with tracking.

The higher workload scores received by direct manipulation reflected its poor usability. Indeed, participants reported frustration when using it due to the time required to complete the tasks, and that the tape scrolling felt less responsive. The vibration made the scrolling movements more imprecise.

That said, H3 stated that the rotaries would generate the lowest workload, which turned out to be partially wrong. The rotary on pedestal was indeed in the group of devices with the lowest workload, but the rotary on screen was not. In fact, it was the one that had the highest physical demand scores. This is consistent with the results of the discomfort questionnaire, which points to the same finding, as the rotary on screen had the highest discomfort scores for shoulder and arm. During the interaction tasks, participants had to keep their arm extended for a significant period of time. Since the rotary was placed to the right of the touchscreen, the angle of the arm and the distance to reach it were greater than those required to interact with the screen itself. This added additional pressure on the shoulder and arm.

At the opposite end of the spectrum, the joystick received significantly lower scores than the rotary on screen for physical demand as well as arm and shoulder discomfort. Having the arm resting on a support closer to the body certainly explained this low physical demand, whereas all the other devices required active maintenance of the arm's position in space. Note that the differences in discomfort between the devices were small since the average scores for each ranged from 0.5/10 (extremely weak) to 2.0/10 (weak).

6.4. Tracking task

Our hypothesis (H4) that the slider and the lateral keypad, i.e. the devices on the virtual panel, would produce the worst tracking task performance turned out to be partially wrong. The slider was indeed one of the two worst performing devices for tracking, but the worst was, by far, direct manipulation. Based on the results, we realize that tracking performance is a result of two factors: task completion time and attention dedicated to tracking. Our hypothesis was based only on the second factor, and in that sense was more reflective of the eye tracking results. Indeed, participants looked less at the tracking task when using the slider and the lateral keypad, which was also confirmed by the participants themselves as half of them said they felt they looked at the tracking task less when using these devices. We had not considered that given the faster completion time of the tasks with the lateral keypad, the deviation between the cursor and the target would be smaller. Conversely, completing data entry tasks with the slider took more time. Combined with lower visual attention dedicated to the tracking task, this led to poor tracking performance.

Our theory also held up for direct manipulation, as long completion time led to poor tracking performance despite participants looking at the tracking task more than with lateral panel devices. In addition, a more serious problem occurred with this device. The participants' arm completely blocked their view of the tracking task when they scrolled the IAS tape because it was located on the left side of the PFD and the participants had to use their right hand. Under these conditions, it was therefore harder to track the target with the joystick.

The only result difficult to explain was the lack of performance difference between the two rotaries. The rotary on the pedestal generated significantly more fixations on the tracking task than the rotary on the screen, and their completion time was not significantly different. We believe that a more difficult tracking task requiring more sustained concentration would have generated a significant difference in performance between the two.

6.5. Usability

It was observed that the ease-of-use score for each device roughly followed data entry time. This impression was confirmed using a Pearson correlation test, as a negative correlation was found between data entry time and ease-of-use score for small changes ($r(138) = -.42$, $p < .001$). For large changes, a negative correlation was also detected ($r(138) = -.46$, $p < .001$). Thus, it is understood that the speed of completion played an important role in the user perception of what makes a device usable. It can be argued that it is not exactly the actual speed that played a role on usability, but rather the perceived speed. Indeed, the correlation between the mean ease-of-use score for small and large changes and perceived speed was even stronger ($r(138) = .66$, $p < .001$). This finding holds true for all devices, as will be demonstrated in the following sections. As keypads were the fastest devices, they received the best usability scores, which confirms H5.

6.5.1. Keypads

The most common comment about the keypads was that it was very quick to select a field and to enter a value, as reported by 13 participants. 11 participants also found the keys too small, but this did not seem to have as big an influence on ease-of-use scores as the task completion time. It seems that despite the potential frustration of having to be precise in pressing small keys, being able to do the tasks quickly impacted usability more strongly.

6.5.2. Rotaries

Consistently with what was previously reported, participants appreciated the tactile feel of the rotaries for making small changes quickly and accurately, as claimed by 10 participants. In contrast, 9 participants found that the rate-aid mechanism did not work well enough and expressed that making large changes took too long, further making the link between speed and usability explicit.

In the case of rotaries, another element that may have affected usability was the way in which the field was selected. Indeed, 15 out of 20 participants preferred the rotary on pedestal to the rotary on screen for this segment of the task. This preference is explained by the way the selection cursor moved for each device. The movement of the cursor for the rotary on pedestal was always the same: a click upwards moved the cursor upwards, etc. On the other hand, the rotary on screen allowed a circular tabbing navigation, with the cursor moving between fields in the direction of the rotary's rotation. However, the movement of the cursor depended on its position: if it was in the upper part of the screen and the rotary was turned to the right, the cursor moved to the right as expected. If the cursor was in the lower part of the PFD, a rightward rotation of the rotary caused the cursor to move to the left as it "rotates" around a central axis. This referential movement required an additional reflexive step and slowed down the reaching of the desired field, hence the stronger interest of the participants in selecting the field with the rotary on pedestal.

6.5.3. Joystick and slider

While for the other devices the ease-of-use and appreciation of a device seemed to depend on its ability to enter a value quickly, this was not clearly the case for the joystick and the slider. As a reminder, data entry time was similar between the two devices for small and medium changes, but the slider was slightly faster for large changes. Surprisingly, the joystick got higher scores for ease of use and perceived speed for large changes. For the joystick, participants noted the improvement in precision they felt during the test (8 participants). As for the slider, it was the difficulty of being precise and the fact of having committed several overshoots that stood out for 11 participants.

The difference in the assessment of performance between the joystick and the slider had an impact on the participants' perception of speed. Even though the slider was in fact slightly faster, the large number of adjustments needed to reach the correct value and the frustration this caused made the tasks longer to their eyes. This confirms the hypothesis posed earlier that it is perceived speed that is related to the usability of a device, not actual speed.

6.5.4. Direct manipulation

The slowness of the device, especially for large changes, was the most frequent comment (13 participants). With 9 participants, the difficulty of being precise when scrolling with vibration was the next most common remark.

6.6. Limitations

A first limitation of our study stems from our definition of an error as an erroneous confirmation. The idea came from a previous study by Cockburn et al. (2017) that defined errors in a similar way, trying to define errors only as actions with undesired consequences. This also applies in our case, as any interaction before confirmation has no impact on airplane security. Still, a calculation of the number of interactions for each task compared to the minimum number of interactions would have allowed a better analysis of each device accuracy. Another limitation was the rate-aid calibration of the rotaries that would require more investigation to improve its usability. The fact that participants did not wear a harness limiting their position as in a real cockpit was also an issue. Although we asked them to remain seated without leaning forward, it was not possible to control their position precisely. Finally, one participant had piloting experience, which implies a potential difference in the results, especially in the appreciation of the devices. Indeed, as presented above, participants liked the keypads the most. However, in a previous experiment (Avsar et al., 2016), pilots reported having little interest in virtual keypads. This is also consistent with a discussion we had with a pilot during a preliminary test of our experiment. Thus, it is likely that the observations on device usability would have been different if more pilots had been selected as participants. Still, we believe our results to be valid with the sample we had: the tasks were simple enough so that anybody could execute them with just a few minutes of practice, and no prior knowledge was required. Because we used a confirmation mechanism to input data and registered very low error rates, we do not think pilots would have interacted with the system differently even though they might have a different attitude toward mistakes.

6.7. Future research

As the pros and cons of the devices tested in this study are becoming well documented, a way to deepen our knowledge in the field would be to repeat a similar study and proposing combinations of devices, for example the touchscreen for field selection and a rotary for data entry. Participants could also be offered two devices simultaneously to observe whether they use more one over the other. Also, since keypads seemed to be promising, it would be interesting to evaluate the performance of keypads located elsewhere on the screen e.g., keypads that can be dragged and dropped onto the PFD by the user. Several participants would also have preferred a physical keypad to a virtual keypad. Intuitively, this would retain the speed advantages of the virtual keypad while enhancing it with a better physical feel. However, research shows that physical keypads do not perform better than virtual ones (Avsar et al., 2016; Lee & Zhai, 2009), and there is little point in testing a mechanical device when the trend is toward virtualization. More promising avenues such as digital input on a touchscreen with haptic feedback, as proposed by Wynne et al. (2021), or using a speech-to-text system could be explored.

7. Conclusion

With the arrival of touchscreens in aircraft cockpits, it is necessary to ensure the performance and safety of touch devices that may replace mechanical ones. In this study, we compared both types of devices in the realistic setting of digital data entry on a flight display. The results showed that traditional rotary knobs provided faster data entry time than keypads when the target value was close to the initial value, but that keypads were faster when the difference between the two values exceeded ten increments. The keypads also produced fewer errors, had higher usability scores, and did not show significant differences in workload and comfort. The realistic nature of the tasks, as well as the way they were divided in size categories, allowed this study to bring more nuance to previous research which placed the keypad as invariably faster than the rotary knob. Our results will allow for a more informed and safe design of touch devices in the cockpit and serve as a baseline for other comparative analyses of innovative data entry techniques.

Acknowledgments

We would like to thank Adam Schachner who built the flight deck rig and assisted us with the experimental design, as well as Nami Bae from CMC Electronics for her input in designing the interface. We also thank all participants who took part in this study, as well as the two reviewers for having challenged us and helped improve the quality of the article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Consortium de Recherche et d'innovation en Aérospatiale au Québec (CRIAQ) under Grant 051768; Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant ALLRP 567177-21; and CMC Electronics Inc.

ORCID

Philippe Doyon-Poulin http://orcid.org/0000-0003-2563-4600

References

Alapetite, A., Møllenbach, E., Stockmarr, A., & Minakata, K. (2018). A rollercoaster to model touch interactions during turbulence. *Advances in Human-Computer Interaction*, *2018*, 1–16. https://doi.org/10.1155/2018/2698635

Alapetite, A., Fogh, R., Zammit-Mangion, D., Zammit, C., Agius, I., Fabbri, M., Pregnolato, M., & Becouarn, L. (2012). Direct tactile manipulation of the flight plan in a modern aircraft cockpit. In *Proceedings of International Conference on Human-Computer Interaction in Aerospace, HCI Aero 2012*. http://orbit.dtu.dk/en/projects/onedisplay-for-a-cockpit-interactive-solution(aba6247c-90a9-4568-8e5d-fde-c20a2d5dc).html

Avsar, H., Fischer, J. E., & Rodden, T. (2016). *Designing touch screen user interfaces for future flight deck operations* [Paper presentation]. 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), 1–9. https://doi.org/10.1109/DASC.2016.7777976

Baldus, T., & Patterson, P. (2008). Usability of pointing devices for office applications in a moving off-road environment. *Applied Ergonomics*, *39*(6), 671–677. https://doi.org/10.1016/j.apergo.2008.01.004

Bhalla, M. R., & Bhalla, A. V. (2010). Comparative study of various touchscreen technologies. *International Journal of Computer Applications*, *6*(8), 12–18. https://doi.org/10.5120/1097-1433

Boeing. (2023). *Touchscreen interface for flight management*. Boeing Global Services. https://services.boeing.com/maintenance-engineering/modifications/avionics/touchscreen-control-display-unit

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Human Kinetics.

Cantu, A., Vinot, J.-L., Letondal, C., Pauchet, S., & Causse, M. (2021). *Does folding improve the usability of interactive surfaces in future airliner cockpits An evaluation under turbulent conditions and varying cognitive load: Apport de la physicalité et du changement de forme pour pallier les faiblesses de l'interaction tactile dans les cockpits d'avions* [Paper presentation]. 32e Conférence Francophone Sur L'Interaction Homme-Machine, 1–10. https://doi.org/10.1145/3450522.3451246

Cockburn, A., Gutwin, C., Palanque, P., Deleris, Y., Trask, C., Coveney, A., Yung, M., & MacLean, K. (2017). *Turbulent touch: Touchscreen input for cockpit flight displays* [Paper presentation]. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 6742–6753. https://doi.org/10.1145/3025453.3025584

Coutts, L. V., Plant, K. L., Smith, M., Bolton, L., Parnell, K. J., Arnold, J., & Stanton, N. A. (2019). Future technology on the flight deck: Assessing the use of touchscreens in vibration environments. *Ergonomics*, *62*(2), 286–304. https://doi.org/10.1080/00140139.2018.1552013

Dodd, S. R., Lancaster, J., Grothe, S., DeMers, B., Rogers, B., & Miranda, A. (2014). *Touch on the Flight Deck: The Impact of Display Location, Size, Touch Technology & Turbulence on Pilot Performance* [Paper presentation]. 2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC), 2C3–1. https://doi.org/10.1109/DASC.2014.6979570

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/h0055392

Garmin, & subsidiaries, G. L. or its (2022). *Garmin G3X Touch™ Flight Displays for Certificated Aircraft*. Garmin. https://www.garmin.com/en-US/p/682215

Giebelhausen, M., Robinson, S. G., Sirianni, N. J., & Brady, M. K. (2014). Touch versus tech: When technology functions as a barrier or a benefit to service encounters. *Journal of Marketing*, *78*(4), 113–124. https://doi.org/10.1509/jm.13.0056

Goode, N., Lenné, M. G., & Salmon, P. (2012). The impact of on-road motion on BMS touch screen device operation. *Ergonomics*, *55*(9), 986–996. https://doi.org/10.1080/00140139.2012.685496

Grahn, H., & Kujala, T. (2020). Impacts of touch screen size, user interface design, and subtask boundaries on in-car task's visual demand and driver distraction. *International Journal of Human-Computer Studies*, *142*, 102467. https://doi.org/10.1016/j.ijhcs.2020.102467

Grier, R. A. (2015). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1727–1731. https://doi.org/10.1177/1541931215591373

Hamon, A., Palanque, P., André, R., Barboni, E., Cronel, M., & Navarre, D. (2014). *Multi-Touch interactions for control and display in interactive cockpits: Issues and a proposal* [Paper presentation]. Proceedings of the International Conference on Human-Computer Interaction in Aerospace - HCI-Aero, *14*, 1–10. https://doi.org/10.1145/2669592.2669650

Hamon, A., Palanque, P., & Cronel, M. (2015). *Dependable multi-touch interactions in safety critical industrial contexts: Application to aeronautics* [Paper presentation]. 2015 IEEE 13th International Conference on Industrial Informatics (INDIN), 980–987. https://doi.org/10.1109/INDIN.2015.7281868

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Harvey, C., Stanton, N. A., Pickering, C. A., McDonald, M., & Zheng, P. (2011). To twist or poke? A method for identifying usability issues with the rotary controller and touch screen for control of in-vehicle information systems. *Ergonomics*, *54*(7), 609–625. https://doi.org/10.1080/00140139.2011.586063

Jeong, H., & Liu, Y. (2017). Effects of touchscreen gesture's type and direction on finger-touch input performance and subjective ratings. *Ergonomics*, *60*(11), 1528–1539. https://doi.org/10.1080/00140139.2017.1313457

Kaminani, S. (2011). *Human computer interaction issues with touch screen interfaces in the flight deck* [Paper presentation]. 2011 IEEE/AIAA 30th Digital Avionics Systems Conference, 6B4-1–6B4-7. https://doi.org/10.1109/DASC.2011.6096098

Lappalainen, M. (2011). *Touchscreen in a safety-critical medical device*. Aalto University.

Lee, S., & Zhai, S. (2009). *The performance of touch screen soft buttons* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 309–318. https://doi.org/10.1145/1518701.1518750

Lin, C. J., Liu, C. N., Chao, C. J., & Chen, H. J. (2010). The performance of computer input devices in a vibration environment. *Ergonomics*, *53*(4), 478–490. https://doi.org/10.1080/00140130903528186

Lin, L.-Y., Cherng, R.-J., & Chen, Y.-J. (2017). Effect of touch screen tablet use on fine motor development of young children. *Physical & Occupational Therapy in Pediatrics*, *37*(5), 457–467. https://doi.org/10.1080/01942638.2016.1255290

Martin, L. (2023, January 26). *Ctrl + P: 3D Printing an F-35 Cockpit*. Lockheed Martin. https://www.lockheedmartin.com/en-us/news/features/2023/ctrl-p-3d-printing-an-f-35-cockpit.html

Mackenzie, I. S., & Buxton, W. (1991). *A comparison of input devices in element pointing and dragging tasks* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 161–166. https://doi.org/10.1145/108844.108868

Mallam, S. C., Nordby, K., Johnsen, S. O., Bjørneseth, F. B. (2020). The digitalization of navigation: examining the accident and aftermath of US Navy destroyer John S. McCain. *Proceedings of the Royal Institution of Naval Architects Damaged Ship V*.

Mayer, S., Le, H. V., Nesti, A., Henze, N., Bülthoff, H. H., & Chuang, L. L. (2018). *The Effect of Road Bumps on Touch Interaction in Cars* [Paper presentation]. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 85–93. https://doi.org/10.1145/3239060.3239071

Miller, T. C., Rose, M., Faturechi, R., Chang, A. (2019, December 20). *The navy installed touch-screen steering systems to save money. Ten Sailors Paid With Their Lives*. ProPublica. https://features.propublica.org/navy-uss-mccain-crash/navy-installed-touch-screen-steering-ten-sailors-paid-with-their-lives/

MIL-STD-1472H. (2020). Department of Defense Design Criteria Standard Human Engineering.

National Aeronautics and Space Administration. (n.d). *MATB-II*. MATB-II. Retrieved October 25, 2022, from https://matb.larc.nasa.gov/

Orphanides, A. K., & Nam, C. S. (2017). Touchscreen interfaces in context: A systematic review of research into touchscreens across settings, populations, and implementations. *Applied Ergonomics*, *61*, 116–143. https://doi.org/10.1016/j.apergo.2017.01.013

Papadatou-Pastou, M., Ntolka, E., Schmitz, J., Martin, M., Munafò, M. R., Ocklenburg, S., & Paracchini, S. (2020). Human handedness: A meta-analysis. *Psychological Bulletin*, *146*(6), 481–524. https://doi.org/10.1037/bul0000229

Rockwell Collins. (2016). *Boeing 777X to feature touchscreen flight displays from Rockwell Collins*. Skies Mag. https://skiesmag.com/press-releases/boeing-777x-feature-touchscreen-flight-displays-rockwell-collins/

Rogers, W. A., Fisk, A. D., McLaughlin, A. C., & Pak, R. (2005). Touch a screen or turn a knob: Choosing the best device for the job. *Human Factors*, *47*(2), 271–288. https://doi.org/10.1518/0018720054679452

Rouwhorst, W., Verhoeven, R., Suijkerbuijk, M., Bos, T., Maij, A., Vermaat, M., & Arents, R. (2017). *Use of touch screen display applications for aircraft flight control* [Paper presentation]. 2017 IEEE/

AIAA 36th Digital Avionics Systems Conference (DASC), 1–10. https://doi.org/10.1109/DASC.2017.8102060

Rydström, A., Broström, R., & Bengtsson, P. (2012). A comparison of two contemporary types of in-car multifunctional interfaces. *Applied Ergonomics*, *43*(3), 507–514. https://doi.org/10.1016/j.apergo.2011.08.004

Salmon, P. M., Lenné, M. G., Triggs, T., Goode, N., Cornelissen, M., & Demczuk, V. (2011). The effects of motion on in-vehicle touch screen system operation: A battle management system case study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(6), 494–503. https://doi.org/10.1016/j.trf.2011.08.002

Schachner, A., & Doyon-Poulin, P. (2023). Avionic touchscreen interaction under vibration: Supported versus freehand target selection in cockpit conditions. *International Journal of Human–Computer Interaction*, 1–30. https://doi.org/10.1080/10447318.2023.2212225

Smith, M. A., Plant, K. L., Parnell, K. J., Wynne, R. A., & Stanton, N. A. (2020). Investigating the usability of touchscreen interfaces in a turbulent flight deck – for panning and numeric data entry tasks. *SID Symposium Digest of Technical Papers*, *51*(1), 1438–1441. https://doi.org/10.1002/sdtp.14158

Smith, M., Coutts, L. V., & Plant, K. L. (2018). *Investigating the Usability of Touchscreens in a Turbulent Flight Deck* [Paper presentation]. 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), 1–10. https://doi.org/10.1109/DASC.2018.8569749

Stanton, N. A., Harvey, C., Plant, K. L., & Bolton, L. (2013). To twist, roll, stroke or poke? A study of input devices for menu navigation in the cockpit. *Ergonomics*, *56*(4), 590–611. https://doi.org/10.1080/00140139.2012.751458

van Zon, N. C. M., Borst, C., Pool, D. M., & van Paassen, M. M. (2020). Touchscreens for aircraft navigation tasks: Comparing accuracy and throughput of three flight deck interfaces using Fitts' law. *Human Factors, 62*(6), 897–908. https://doi.org/10.1177/0018720819862146

Wang, L., Wang, Y., & Chen, Y. (2018). Survey on introducing touchscreen into civil aircraft cockpit: Opinions of aircraft designers and pilots. In *CSAA/IET International Conference on Aircraft Utility Systems (AUS 2018)* (pp. 1260–1265). Guiyang. https://doi.org/10.1049/cp.2018.0250

Wang, H., Tao, D., Cai, J., & Qu, X. (2022). Effects of vibration and target size on the use of varied computer input devices in basic human-computer interaction tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *32*(2), 199–213. https://doi.org/10.1002/hfm.20938

Williams, P. D. (2017). Increased light, moderate, and severe clear-air turbulence in response to climate change. *Advances in Atmospheric Sciences*, *34*(5), 576–586. https://doi.org/10.1007/s00376-017-6268-2

Wobbrock, J. O., Cutrell, E., Harada, S., & MacKenzie, I. S. (2008). *An error model for pointing based on Fitts' law* [Paper presentation]. Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08, 1613. https://doi.org/10.1145/1357054.1357306

Wynne, R. A., Parnell, K. J., Smith, M. A., Plant, K. L., & Stanton, N. A. (2021). Can't touch this: Hammer time on touchscreen task performance variability under simulated turbulent flight conditions. *International Journal of Human–Computer Interaction*, *37*(7), 666–679. https://doi.org/10.1080/10447318.2021.1890492

Yau, Y.-J., Chao, C.-J., & Hwang, S.-L. (2008). Effects of input device and motion type on a cursor-positioning task. *Perceptual and Motor Skills*, *106*(1), 76–90. https://doi.org/10.2466/pms.106.1.76-90

Yeh, P.-C. (2020). Impact of button position and touchscreen font size on healthcare device operation by older adults. *Heliyon*, *6*(6), e04147. https://doi.org/10.1016/j.heliyon.2020.e04147

About the authors

Charles-Antoine Lanoix is a Master's student at the Department of Mathematical and Industrial Engineering at Polytechnique Montreal, where he studies human-computer interaction, human factors and user experience. This article is his first authored work.

Srishti Rawal is an undergraduate student at the Aerospace Engineering at the SRM Institute of Science and Technology in Chennai, India. She will start her graduate studies at Polytechnique Montreal under the supervision of Prof. Doyon-Poulin.

Philippe Doyon-Poulin is an assistant professor in the Department of Mathematical and Industrial Engineering at Polytechnique Montreal. His research focuses on human factors in aviation, error prevention and decision-making with automated systems. He authored more than 25 aircraft certification reports to show compliance on flight deck usability and pilot error.