Some Title Here

By

RICHARD D. JOYCE
B.S (Columbia University) 2009

THESIS

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mechanical and Aerospace Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Stephen K. Robinson, Chair

Ron A. Hess

Michael Feary

Committee in charge

2017

Contents

List of Figures

List of Tables

Richard D. Joyce
June 2017
Mechanical and Aerospace Engineering

Some Title Here

Abstract

put the abstract here

Acknowledgments

xxxxxxx

Chapter 1

Introduction

Chapter 2

Design Evaluation Experiment

2.1 Introduction

After investigating the technical approach and the benefit to including the passive haptics layer, we seek to investigate the use of the Rapidly Reconfigurable Research Cockpit in a more realistic design evaluation study. The advantages of using the R3C system would not be useful if it masked defects in a design study.

2.2 Methods

In order to perform a design evaluation study, it was first needed to have a task that the subjects would be doing using the designs.

2.2.1 Simulator Setup

The simulator workstation as configured for each group is shown and annotated in Figure ??. It was our goal to have as much as possible to be the same between the two configurations. The joystick and instrument were positioned in the same location for each group. Neither group had out the window visuals, relying only on the attitude indicator on the instrument. For the Virtual Reality (VR) group, the visuals showed a plain interior of a cockpit, but the out the window view was black. Both groups had an aural indication (a click noise of a button being pressed) when a button was activated on the instrument,

this is not done yet

using the speakers mounted behind the instrument panel.

The main difference between the two groups, beyond the VR group wearing a virtual reality headset, was the way the instrument was interacted with. The VR group used the hand tracker activated system previously described in Chapter **??**. For this experiment, the buttons were configured to highlight a blue color when the hand tracker registered a finger within the zone . After the 150 msec delay when the button was activated, the highlight would disappear and the button in the virtual world would move inwards as if it were being pushed in (of course, the physical button could not and did not move). It is also at this time that the press sound would play, as well as any response on the instrument associated with pressing that button. When the finger left the zone after a successful press, a seperate release sound would play and the button would move back to its starting position.

terrible sentence

Find the size of the zone

The Touchscreen (TS) group used a 10.1 inch capacitive touch screen with resolution of 1024x768. The two instruments were drawn in a web browser, using standard HTML elements for the buttons. Javascript press and release events were used to simulate the same behavior as described for the VR group, except for the highlighting before a button press. The visuals of the tracker were rendered on top of the browser window with the same OpenGL rendering code used for the VR group.

2.2.2 Task Design

With this simulator setup base and the goals of the study, a number of requirements were created to design the task that the subjects would perform.

- Flight task using a standard joystick
- Second task that requires use of multiple buttons on the instrument
- Able to develop simulator for both touchscreen and R3C setup
- Able to design two different layouts with one design having distinct flaws
- Simple design yet complex enough task to have sufficient workload
- Operationally relevant, or analogous to tasks required in a cockpit

could explain what this limits

Figure 2.1: Attitude Indicator Display

Ultimately, we designed a task that required number and letter inputs using the buttons, while simultaneously flying a pitch disturbance profile.

Tracking Task

The tracking task display was a standard attitude indicator display, shown in Figure 2.1. Each tick corresponds to 1 degree in the dynamics simulation, with the major ticks at intervals of 5 degrees. The attitude indicator was X.X inches square on the instrument. Thats not a real di-

Subjects controlled the one-dimensional (pitch only) task using a joystick with

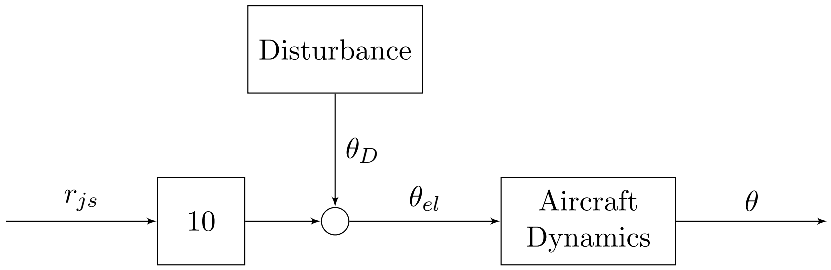

Figure 2.2: Tracking Task Dynamics Block Diagram

their left hand. The joystick is pictured in Figure **??**.

The flight dynamics model of the simulator was a stability derivative based model for a Boeing 747 in a low altitude landing configuration.

Include all the info on the FDM

The output of the joystick, r_{js}, varies from -1.0 to 1.0, and the gain of 10° was chosen to ensure the pilot had enough control authority to complete the task.

The transfer function of the aircraft dynamics is given as:

$$\frac{\theta}{\theta_{el}} = \frac{-0.572(s+0.553)(s+0.0396)}{(s^2+2\zeta_1\omega_1+{\omega_1}^2)(s^2+2\zeta_2\omega_2+{\omega_2}^2)} \tag{2.1}$$

$$\omega_1 = 0.0578 \qquad \zeta_1 = 0.0160$$

$$\omega_2 = 1.12 \qquad \zeta_2 = 0.798$$

The disturbance model is based off the model developed in SweetRef. It is designed to provide a broad spectrum of frequencies that the human controller needs to respond to.

$$\theta_D = K\sum_{i=1}^{12}\left[a_i\left(\frac{2\pi k_i}{240}\right)\sin\left(\frac{2\pi k_i}{240}t+\phi_i\right)\right] \tag{2.2}$$

The k_i terms are given as,

$$k_1 = 7, \qquad k_2 = 11, \qquad k_3 = 16$$

$$k_4 = 25, \qquad k_5 = 38, \qquad k_6 = 61$$

$$k_7 = 103, \qquad k_8 = 131, \qquad k_9 = 151$$

$$k_{10} = 181, \qquad k_{11} = 313, \qquad k_{12} = 523$$

The amplitude terms is $a_i = 0.5$ for $i <= 6$ and $a_i = 0.005$ otherwise. The phase terms, ϕ_i, were randomly selected on the $(-\pi, \pi)$ interval ensuring a uniform distribution. This random selection was precalculated for each trial, however the order was repeated for each subject so there was no between subjects variance in the disturbance signal. Furthermore, each subject received the same sequence of disturbance signals for each design, eliminating within subject variance as well. The disturbance amplitude, K, was chosen such that the root-mean square (RMS) of the signal was 3.5 degrees.

Prompting Task

The prompting task was designed to be both a realistic task as well as demanding to create a high workload.

The sequencing of the prompts seperated into 10 second “windows”. The prompt would appear between 2 and 3 seconds of the start of the window. From the time of appearance, a seven (7) second timer will start until timeout. When the subject presses the first button of the prompt, the prompt itself was cleared and asterisk symbols (*) were shown after each button entry by the subject. After the subject has entered 4 buttons or the timeout occurs, whichever comes first, the prompt or entry so far would clear. This process is then repeated every 10 seconds.

The prompts themselves were composed of three numbers followed by a letter or three letters followed by a number. This structure was decided upon to provide a consistent pattern. The prompts were randomly chosen but were not allowed to have repeat numbers or letters, and for the prompts with three letters, common words or acronyms were filtered out (e.g. “BAD”, “FDA”). The selection of letters or numbers as the first three characters was randomly chosen as well.

2.2.3 Instrument Designs

The two different designs used were developed to be both realistic and believable as a cockpit instrument design that would be under consideration, but still have one design with a flaw that would be found in a design evaluation.

Keypad

Edgekey

2.2.4 Experiment Design

Subjects were divided into the two groups, TS and VR. The overall sequence of the experiment started with a training session on the simulator and the task, then an evaluation session for each of the two designs, finally finishing with questionnaires asking about the designs. The timeline of the experiment was the same for each subject, except for counterbalancing the order that the designs were evaluated. The training portion started with a slide deck explaining the tasks, the simulator that the subject was using, and the functionality two designs they were to evaluate. Next, they performed practice trials with just the tracking task and then just the prompting task.

For the evaluation sessions with each design, they performed six trials with both tasks. The first three were a minute long, and were considered practice trials, and not included in the data analysis. The following three were two minutes each, and were the trials used for the results. Each evaluation session concluded with a two minute trial of just the tracking task. This was included to investigate if the subject had improved or fatigued at the tracking task.

2.2.5 Dependent Measures

The dependent measures were chosen to evaluate the performance of each task individually as well as the workload of the subject. For the tracking task, the root-mean square error (RMSE) was calculated for each trial. The error in this case is simply the pitch shown to the subject, the output of the flight model described above.

The prompting task has two dependent measures, for speed and accuracy. For speed we consider the *response time*, defined as the time between the prompt is first shown to the subject and when they press the first button of their response entry. The accuracy is measured by how many prompts they complete correctly. Twelve prompts are shown to the subject within each trial, and these measures are meaned per trial and then per design for each subject.

For workload, a NASA Task Load Index (TLX) survey was administered after they completed each design. The TLX survey asks for a rating of their workload between 0-100 for the following subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Our implementation allowed selection of the ratings within increments of 5, and included anchors of "Low" and "High" at the extrema of 0 and 100, respectively (except for Performance, which uses "Good" and "Bad"). The midpoint (50) was also visually indicated with a larger tick. The ranked pairs modification was used and completed for both times the subject took the survey. This modification asks the subject, for each of the combinations of pairs of subscales, which of the two they felt contributed more to their workload. The number of times they select each subscale is used a weight to calculate a weighted mean for the total TLX score.

Finally, the subjects were given a questionnaire asking for their feedback on each instrument design. For each design, the subjects were asked the following questions:

- Please comment on any difficulties you had performing the prompting task with this design especially in contrast to the other design.
- Please comment on anything you liked in this design.
- Please comment on anything you did not like in this design.
- Any other comments?

Additionally, the following questions were asked:

- Which instrument design did you prefer? Why?
- Did you experience any physical fatigue during the experiment? Where?
- Any other comments?

An open form text box was used for the response field for each of these questions.

In a standard design evaluation study, the feedback received from the users in this questionnaire (and other debriefing interviews) would often be the main source for carrying out re-design. The goal of this experiment is to determine and document in which ways does this feedback differ. For example, if most subjects in one group noted issues with the size of a button, while no one in the other group found an issue with that button, this would indicate that using this VR system may not highlight the same issues regarding button sizes. The groups were purposely left ambiguous in the example, as it does not matter which group found the flaw and which group did not comment on it. Although we could postulate as to which group are "correct", it is not a useful exercise, as the only result is to document what potential differences could arise.

To analyze these results, the sentences from the open form responses were first separated into single feedback comments, and reworded to use common language. If a subject repeated the same comment in the answers to multiple questions, they were only counted once. Each of these simplified feedback comments were assigned to a category or overall summary of their feedback. This process was completed separately for each group. To summarize the differences, we will look for feedback that is unique to a certain group, as well as the frequency of the comments that are common.

2.2.6 Statistical Tests

The quantitative dependent measures are tested with a two-way ANOVA, with one within subjects factor (Design) and one between subjects factor (Group). The Design factor contains two levels, the two designs each subject tested, Edgekey and Keypad. The Group factor also contains two levels, the VR group and the TS group. When the ANOVA showed significance in the interaction test, post-hoc repeated measured t-tests were undertaken to determine the significance of Design within each Group. All effects were considered statistically significant at the 0.0125 level. Statistical significance level was corrected using the Bonferroni correction considering the 4 different dependent measures being tested ($\alpha = 0.05/4 = 0.0125$).

2.3 Results

2.3.1 Demographics

Twenty-three subjects were recruited from the UC Davis engineering undergraduate and graduate student population. Twelve subjects were placed in the VR group, and the remaining eleven in the TS group. The mean age was $21.0(\sigma = 3.14)$, with 19 male and 4 female subjects. The female subjects were balanced between the two groups. Most subjects had no flight experience (two were student pilots), and all of the VR group subjects indicated that they had less than one hour of experience using virtual reality headsets.

2.3.2 Performance Measures

The performance of the tracking task was measured using the root-mean square average (RMSE) of the error. The effect of group yielded an F ratio of $F(1, 21) = 21.4, p < 0.001$ indicating a significant difference between VR ($M = 1.28\text{deg}, \sigma = 0.38\text{deg}$) and TS ($M = 1.97\text{deg}, \sigma = 0.38\text{deg}$). The effect of design indicated no significant difference ($F(1, 21) = 5.94, p = 0.024$) between Keypad ($M = 1.57\text{deg}, \sigma = 0.51\text{deg}$) and Edgekey ($M = 1.70\text{deg}, \sigma = 0.52\text{deg}$). The interaction effect was not significant ($F(1, 21) = 0.17, p = 0.69$).

Response time. The effect of group yielded an F ratio of $F(1, 21) = 1.61, p = 0.22$ indicating no significant difference between VR ($M = 2983\text{msec}, \sigma = 439\text{msec}$) and TS ($M = 2737\text{msec}, \sigma = 566\text{msec}$). The effect of design indicated a significant difference ($F(1, 21) = 13.9, p = 0.001$) between Keypad ($M = 2728\text{msec}, \sigma = 512\text{msec}$) and Edgekey ($M = 3002, \sigma = 488\text{msec}$). The interaction effect was not significant ($F(1, 21) = 0.17, p = 0.69$).

Number of prompts correct. The effect of group yielded an F ratio of $F(1, 21) = 43.9, p < 0.001$ indicating a significant difference between VR ($M = 6.06, \sigma = 2.90$) and TS ($M = 10.2, \sigma = 1.23$). The effect of design indicated a significant difference ($F(1, 21) = 64.1, p < 0.001$) between Keypad ($M = 9.30, \sigma = 1.83$) and Edgekey ($M = 6.78, \sigma = 3.54$). The interaction effect was significant as well ($F(1, 21) = 27.8, p < 0.001$). The post-hoc tests indicated significance between designs for the VR group ($t(11) = 8.0, p < 0.001$)

between the Keypad design ($M = 8.11, \sigma = 1.62$) and the Edgekey ($M = 4.00, \sigma = 2.37$) The post-hoc tests indicated no significant difference between designs for the TS group ($t(10) = 2.3, p = 0.05$) between the Keypad design ($M = 9.82, \sigma = 1.38$) and the Edgekey ($M = 10.6, \sigma = 0.96$)

NASA TLX scores. The effect of group yielded an F ratio of $F(1, 21) = 1.69, p = 0.21$ indicating a significant difference between VR ($M = 70.0, \sigma = 22.6$) and TS ($M = 65.3, \sigma = 8.53$). The effect of design indicated a significant difference ($F(1, 21) = 23.6, p < 0.001$) between Keypad ($M = 57.8, \sigma = 15.2$) and Edgekey ($M = 77.7, \sigma = 13.4$). The interaction effect was significant as well ($F(1, 21) = 8.25, p < 0.001$). The post-hoc tests indicated significance between designs for the VR group ($t(11) = -4.20, p = 0.001$) between the Keypad design ($M = 54.4, \sigma = 20.4$) and the Edgekey ($M = 85.6, \sigma = 11.2$) The post-hoc tests indicated no significant difference between designs for the TS group ($t(10) = -2.72, p = 0.02$) between the Keypad design ($M = 61.5, \sigma = 4.46$) and the Edgekey ($M = 69.2, \sigma = 10.1$)

2.3.3 Design Feedback

The categories of feedback and the counts of how many times they occured for each group is summarized in Table 2.1.

By far the issue that received the most feedback was the difficulty of using the switch key (Edgekey, Switch Difficult). Most of the complaints just centered around the extra difficultly of habving to press another button. Some noted that it took extra time (with no extra time given), it added to the mental demands of the task, it was difficult to see which mode the instrument was in. Both groups disliked the switch key, and mentioned it just as frequently.

> Switching from numbers to letters was hard, especially if I was trying to compensate for turbulence and was struggling at the time. (TS Subject)
>
> I did not like how much extra work it was. It took so much extra focus that I forgot I was flying with the joystick (VR Subject)

Many subjects noted the familiarity of the Keypad design (Keypad, Familiar) and that having the buttons close together (Keypad, Buttons Proximal) as things they like about that design. The familiarity was noted more often for the TS Group, but both were some of the more frequent comments within each group.

One comment about the Edgekey design that got more frequent mentions from the TS Group was that they found having the flight task in the middle of the display, centered between the buttons, was preferred (Edgekey, Centered Flight Task Better). The subjects who noted they preferred the Edgekey design almost uniamonously cited this as their reason for their preference. The comments that fed into this category also included subjects who noted the difficulty of splitting their focus back and forth with the Keypad design. Interestingly, two of the TS Group subjects noted that they would have found the Keypad easier if they had tactile feedback to guide their input. It is possible that the reason the VR Group subjects did not note this as often is because with the tactile feedback they were able to keep visual focus on the left half of the screen in the Keypad design, thus not seeing benefit from the centering of the flight task display.

> [The Edgekey design] forced me to pay more attention to what I was typing, this wouldn't have been a problem if the keypad was a physical device that allowed me to locate the numbers and letters without looking, much like the dots on a computer keyboard. (TS Subject)

> I like that the flight control was centred, so you could see it even when you were looking at the buttons. (VR Subject) (VR Subject)

The most notable exceptions to providing similar feedback between groups are the categories that relate to fatigue issues. Many subjects in the TS group noted fatigue cause from using the joystick, yet none in the VR group did, even though they were using the same joystick setup, and sitting in the same location. The VR group did note more fatigue caused by using their other arm for the prompting task. This fatigue seemed to be caused by the additional effort needed to have the hand tracker recognize the hand. For example, one subject wrote:

> My right wrist was somewhat fatigued. Though I think this is mostly from positioning my hand for the simulator to recognize my input. (VR Subject) (VR Subject)

Similarly to the fatigue issues being different, there were some comments that were due to the technology being used moreso than the designs themselves. The obvious ones are the subjects who noted difficulty using the hand tracker, but some of the other categories had comments that may have been caused by this. For example, the keypad design was noted as causing more mistakes for some subjects. For the TS Group, this was due to the touchscreen being so quick to use:

> since I was able to go more quickly with this layout, I had more mistakes in the entry. (TS Subject)

One subject in VR noted a common problem caused by the hand tracker which caused more mistakes in the Keypad design. When the hand tracker was having registration issues it would sometimes place the other fingers mistakenly in the activation zone of the buttons underneath the one being targeted, causing multiple buttons to be pressed in a short period of time.

> There's more unintended register since other fingers might trigger the buttons (VR Subject)

Although only one subject noted this, it was observed happening to many subjects.

2.4 Discussion

2.5 Conclusion

Topic	Feedback Summary Category	VR Group	TS Group
Edgekey	Switch Difficult	14	12
Keypad	Familiar	6	11
Edgekey	Centered Flight Task Better	4	13
Keypad	Buttons Proximal	6	7
Keypad	Buttons Always Visible	5	5
Other	Hand Tracking Issues	9	0
Fatigue	Fatigue from Joystick	0	8
Edgekey	Hand Blocks View	3	4
Edgekey	Clean Design	3	2
Fatigue	Prompting Arm	4	0
Edgekey	Easier	0	4
Keypad	Buttons Confusable	0	3
Fatigue	Eye Fatigue	3	0
Keypad	Easy Focus Switch	2	1
Keypad	More Mistakes	1	2
Edgekey	Accuracy Worse	1	2
Keypad	Buttons Bad Layout	2	0

Table 2.1: Counts of Design Feedback Comments per Group

Appendices