The Effects of Concurrent Bandwidth Feedback on Robotics Manual Control Tasks

An Experimental and Modeling Study

SUBMITTED BY:

John Austin Karasinski

Ph.D. Dissertation Proposal

University of California, Davis

Mechanical and Aerospace Engineering

Examining Committee:

COMMITTEE CHAIR:	Professor Ron Hess Mechanical and Aerospace Engineering
COMMITTEE MEMBERS	Professor Sanjay Joshi Mechanical and Aerospace Engineering
	Professor Zhaodan Kong Mechanical and Aerospace Engineering
	Professor Stavros Vougioukas Biological and Agricultural Engineering
	Dorrit Billman, Ph.D. NASA Ames Research Center

Stephen K. Robinson

Professor Stephen K. Robinson, Advisor

Contents

1 Problem Statement

1.1 Motivation

We aim to improve performance and decrease learning times for novice operators of highly complex motor control tasks. We are specifically interested in modeling and improving human performance in robotic arm tasks, which generally require extensive training to master. The robotic arm on the International Space Station (ISS), for instance, requires hundreds of hours of training time for astronauts to reach proficiency. Being able to decrease this training time could lead to significant savings in cost, and the predictive ability provided by modeling human performance allows for safer operation of the robotic arm.

A variety of skills can be classified as motor control tasks, such as playing tuba, pole vaulting, or flying an aircraft. An individual's performance in any of these skills can change dramatically as they transition from a novice to an expert through training. We are interested in measuring and modeling this performance as it changes over the course of the training process.

Humans rely on several kinds of feedback during training to improve their performance in motor control tasks. Feedback can be largely grouped into two types: internal, or intrinsic feedback, and external, or extrinsic feedback. Intrinsic feedback is anything a person can infer using their senses: the feel of the valves of the tuba as you play, the sense of balance mid-jump, or the sound the aircraft engine makes during a climb. Extrinsic feedback, conversely, is provided by an external source, often in the form of an expert instructor. Extrinsic feedback comes in a variety of forms, and has a long history of improving performance in a large variety of motor control tasks.

We will focus on a specific type of extrinsic feedback, which is known as concurrent bandwidth feedback (CBF). Concurrent feedback is provided in real-time, as an operator is completing a task. Bandwidth feedback is provided when a objective particular value deviates outside a designated range or bandwidth. Concurrent bandwidth feedback is, therefore, feedback provided to an operator in real-time when a signal deviates out of a predefined range. This type of feedback has been shown to improve performance in many simple motor control tasks, but has not been investigated in complex, high degree of freedom tasks.

It is important to note that this feedback should be thought of as qualitative feedback, not as an additional form of quantitative guidance. We are not interested in adding additional displays or gauges to control interfaces, but would prefer to modify existing indicators, during training, to better inform an operator as to how well they are performing a task. Despite extensive evidence as to the effectiveness of this feedback, the mechanism by which performance is improved has yet to be explained, nor integrated into human performance models. We will attempt to explain why this feedback is effective in enhancing learning and integrate this explanation into a model.

1.2 Research Aims

We are interested in measuring, modeling, and predicting the effects of concurrent bandwidth feedback (CBF) on human performance in robotics manual control tasks. To this end, this proposed research includes three research aims. These aims build on each other, starting with a compensatory tracking task, extending to a robotics task, and finishing with a descriptive model describing both. The first aim is complete, and the second and third aims are in progress.

Aim One Investigate the effects of concurrent bandwidth feedback on human performance and workload effects in a three-axis manual tracking task.

Aim Two Investigate the effects of concurrent bandwidth feedback on human performance and workload effects in a robotics track and capture task.

Aim Three Extend the Hess Structural Model of the human pilot to include the effects of concurrent bandwidth feedback.

There are a number of research questions that we intend to answer by completing these aims, which include:

1. Can concurrent bandwidth feedback improve human performance in a three-axis manual tracking task?
   (a) Do 3D augmented reality displays show improved performance compared to traditional 2D displays?
   (b) Can performance be increased without increasing workload?
2. Can concurrent bandwidth feedback improve performance of simulated robotics tasks?
   (a) Can CBF reduce the required training time to peak performance?
   (b) Can CBF be removed after reaching peak performance without reducing subject performance?
   (c) Can performance be increased without increasing workload?
3. Can we develop a descriptive model of human performance which includes the effects of concurrent bandwidth feedback?
   (a) Can we use this model to estimate operational limits?

The remainder of this proposal is divided into three sections: a literature review, the proposed research, and a timeline.

2 Background

2.1 Augmented Feedback

The concept of feedback was popularized when closed-loop control systems were first developed and has since been defined many times [1]. In the context of this work, a convenient definition of feedback comes from Ramaprasad, "[f]eedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way [2]." The aforementioned "gap" is the error and can be conveyed to the operator of a system in a variety of ways. Feedback can be broadly classified into two types: intrinsic, feedback which is generated from within the context of the action itself, and extrinsic, feedback which is given from an external source [3].

Extrinsic feedback, which is also known as augmented feedback, has been extensively studied in the field motor learning [4]. In their 2013 review, Sigrist et al. write "[i]t is generally accepted that augmented feedback, provided by a human expert or a technical display, effectively enhances motor learning [4]." There are a variety of different forms of augmented feedback, that can be further classified by how, when, and by what form the feedback is provided. For this work, however, we sought to evaluate a type of feedback which was both conceptually simple and could be tied to operational requirements. With these requirements in mind, we will investigate the effects of concurrent bandwidth feedback. Concurrent, or real-time, feedback is displayed to the operator while the task is being executed, in contrast to terminal feedback, which is displayed after the task is complete. Bandwidth feedback is displayed to the operator when some parameter is inside (on-track feedback) or outside (off-track feedback) of an acceptable, predefined tolerance limit. In our implementation, concurrent bandwidth feedback describes feedback conveyed to the operator when their real-time performance drifts outside of an acceptable tolerance limit.

Experimentation with bandwidth feedback traces its origins to Thorndike's 1927 line-drawing experiment [5]. In his experiment, subjects were seated and blindfolded at a table, then asked to draw lines of 3, 4, 5, or 6 inches. The experiment was divided into two groups of subjects, one group of subjects received no verbal feedback, while the other group was told "right" if they were within an 1/8th of an inch of the desired length for the 3 inch line, or 1/4 of an inch for the other three line lengths, and "wrong" if they were outside this bandwidth. Subjects that received the verbal bandwidth feedback improved from an initial median "right" percentage of 13% to 54% after several training sessions. The feedback was then removed after these training trials, during which time subjects dropped to a median percentage of 26%. This is consistent with the guidance hypothesis (which was not formalized for another fifty years after this experiment was concluded), which states that consistent feedback during the acquisition phase of learning leads to a dependency on the feedback [6]. Subjects became dependent on the verbal feedback (extrinsic feedback) rather than their visual or proprioceptive sense (intrinsic feedback) to such an extent that they could no longer perform with the verbal feedback removed.

Payne and Hauty performed one of the first concurrent bandwidth feedback studies in 1955 [7]. In their study, subjects completed a multidimensional pursuit test, which required them to scan four simulated aircraft instruments and counter their drift by adjusting simulated aircraft controls. Subjects were placed into one of three feedback groups: a control level, where no feedback was provided, a second level, which included a single peripheral visual signal when a deviation in one of the displays occurred, but did not specify which instrument, and a third level, which provided individual indicators for each of the four instruments, and noted the locus of the deviation. They found a very significant effect between the different feedback groups, with the control group performing the worst, the second level performing better, and the third level performing better still. Subjects completed the test every hour for a four hour period. Performance dropped across all three groups as time elapsed, but the performance of the subjects in level three was superior at the end of this period compared to the subjects in the control group at the beginning of the experiment. They concluded by stating that "the increment is a positive function of the specificity of the information supplied, it can be ascribed largely to the directive properties of the cues, i.e., the cues impose a more efficient temporal and spatial organization upon [the subject's] scanning behavior [7]."

Gordon and Gottlieb performed a rotary pursuit study investigating the effects on on-track and off-track concurrent bandwidth feedback in 1967 [8]. Subjects in their study were placed into one of three groups: a control, on-track feedback, and off-track feedback. The subjects in the bandwidth feedback groups had to track a 0.75 inch by 0.75 inch target with 0.187 inch rigid stylus tip. For subjects in the on-track feedback group, a light bulb was illuminated when they were on target, and the light bulb was illuminated for subjects in the off-track group when they were not on the target. While both the on-track and off-track groups performed better than the control group, the off-target group performance was slightly superior. This finding was consistent with Williams and Briggs for subjects completing their similar task [9]. Additionally, subjects in the feedback groups completed several trials at the end of the experiment without feedback and did not experience the loss of performance which is often seen due to the guidance hypothesis. This indicates that subjects were able to use the feedback to better learn the task and were not completely dependent on the feedback. Subjects used their own intrinsic feedback to learn the task, and were able to

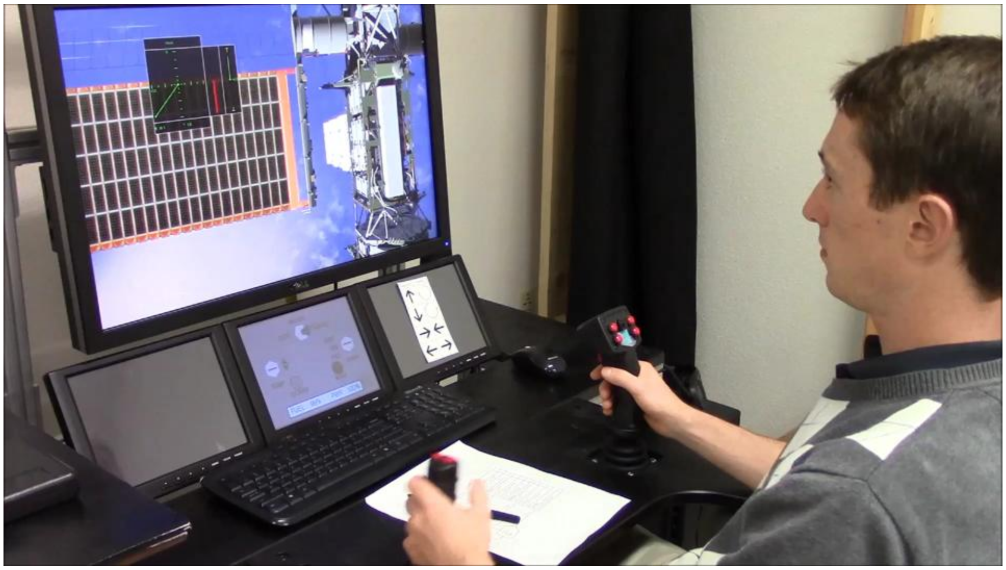


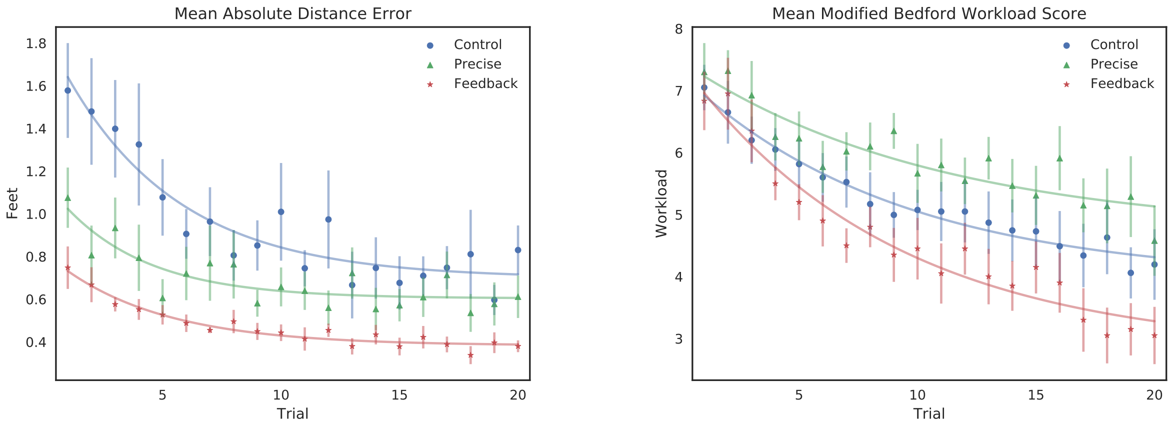
FIGURE 1: A subject from the SAFER experiment seated in the fixed-base simulator [12].

take advantage of the concurrent bandwidth feedback to both better learn and perform the task without becoming dependent on the external feedback.

In 2011, de Groot et al. investigated the effects of concurrent bandwidth feedback on learning a lane-keeping task in a driving simulator [10]. Similar to Gordon and Gottlieb, they investigated the effects of on-track and off-track feedback compared to a control group. Instead of using a visual indicator, however, de Groot et al. used haptic feedback in the form of a vibrating chair for their feedback groups. They found that on-target and off-target groups had better lane-keeping performance than the control group, and that, similar to Gordon and Gottlieb and Williams and Brigg, the off-target group performed best. Retention trials, however, showed that a majority of this performance improvement was lost during when the feedback was removed, which was in accordance with the guidance hypothesis. The off-target group, though, did still retain some minor performance improvement, which the authors partially attribute to the onset advantage [11]. The onset advantage "suggests that the sudden onset of a stimulus is a more powerful perceptual event than a stimulus offset, facilitating low-level perceptual processing and resulting in faster reaction times [10]." This effect could explain a repeated suggestion that off-track feedback is superior to on-track feedback, even if the effect is, in general, small. de Groot et al. also measured response time to a secondary task as an estimate of workload, but found no differences across groups.

2.1.1 SAFER Experiment

Our recent work includes investigations into concurrent bandwidth feedback in a four degree of freedom Simplified Aid for EVA Rescue (SAFER) task [12–14]. SAFER is a small propulsive jet pack worn during spacewalks for self-rescue [15]. Subjects were tasked with flying a SAFER simulation to perform an inspection of the International Space Station's (ISS) solar arrays. Subjects were initially placed 40 feet away from the solar array and were asked to close to 30 feet and hold this distance for the remainder of the task. They could gauge their distance from the solar array using the indicator on the guidance display and the out-the-window display. Subjects were then



(a) Mean absolute distance error, by group.

(b) Mean subjective workload rating, by group.

FIGURE 2: Subjects with CBF performed the best (a) and reported the lowest workload (b). Errors are the standard error of the mean [12].

asked to inspect four "damaged" waypoints on the solar array, and were given a guidance display for navigation to the waypoints. Two vertically arranged displays in the simulator were available to complete the task, see Figure 1. The primary display contained an out-the-window view of the solar array and, depending on which group the subject was in, one of the guidance displays. The secondary display located directly below the primary display portrayed information about the subject's current mode, remaining fuel, and a "comm light".

In our experiment, subjects were placed into one of three groups: a control, a high precision augmented feedback group, and a concurrent bandwidth feedback (CBF) group. Subjects in the high precision group were given an extra significant figure in their guidance display and an analog display which was scaled twice as large (but had half the maximum value) of their flight parameters. Subjects in the CBF group had two display elements which would change from a green to a red color when the subject's performance was outside a predefined range. Performance was measured as mean absolute error (MAE) and results across trials are shown in Figure 2a. Both treatment groups performed better than the control group, with the CBF group performing the best and having the least error. The effects that the treatments had on workload was very different than performance, however. Subjects in the high precision group reported significantly more workload than the control group, while subjects in the CBF group reported significantly less workload than the control group, see Figure 2b. The concurrent bandwidth feedback also had the added benefit of significantly reducing the amount of time required to train the subjects to their maximum skill level. Subjects with the CBF performed better on their first trial than subjects in the control group did on their last, which was after approximately two hours on the task.

2.2 Workload Measurement

Improving performance, through some kind of feedback or other technique, usually comes at the cost of increased workload, which can lead to a loss of the ability to sustain improved performance. Workload was defined by Hart and Staveland as "the perceived relationship between the amount of mental processing capability or resources and the amount required by the task" [16]. More simply put, having a low workload indicates that it would be easy to complete additional tasks, while having a high workload suggests that it would be difficult.

The NASA Task Load Index (NASA-TLX) is one of the most well known and commonly used subjective workload measures. The NASA-TLX has been in use for thirty years, and has been used and validated over a large variety of tasks [17]. The NASA-TLX is a multidimensional rating scale which uses the magnitude and ranking of six subscales to produce an overall estimate of subjective workload [16]. The six subscales are: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Each of these scales is rated from a 0 (Very Low) to 100 (Very High) scale, with the exception of Performance, which is rated from 0 (Perfect) to 100 (Failure). After marking a value for each of these subscales, subjects then make fifteen pairwise weightings, allowing them to rate each pair of subscales based on its perceived contribution to their overall workload. A final, overall workload score is computed by multiplying each subscale's score by the number of times it was chosen in the pairwise weightings, adding these values, and dividing by fifteen. As certain subscales may be more or less important than others, depending on the task being evaluated, some researchers can drop subscales or simply not compute the overall score.

In addition to subjective measures of workload, there are a variety of techniques which aim to estimate objective workload. One of the most common objective measurement techniques is the secondary task, which requires subjects to complete the primary task, then use any spare workload to respond to an additional task [18]. Secondary tasks can provide a measure more sensitive to differences in workload and performance than a single task alone and allow for a common measure between experimental conditions [19]. Care must be taken, however, to ensure that the secondary task does not intrude upon primary task performance [20]. In our previous studies, we have used a multiple choice reaction time task as an objective workload measurement. In this secondary task, subjects are presented with several different stimuli, each of which requires a different response [21]. A subject's objective workload can then be inferred by either the percentage of secondary tasks which were correctly responded to within a given time, the number of secondary tasks which were correctly responded to in a trial, or both. We have previously found this type of task to be accurately tied to subjective workload scales in the aforementioned SAFER task [13].

2.3 Pilot Modeling

In addition to popularizing the concept of feedback, the creation of control theory in the early 1940s also provided the tools required for the mathematical modeling of the human pilot. At the time, new weapons were being created for World War 2 which could only be used effectively with trained operators working in tandem with the machine. While it was thought that a human could be viewed as a unique kind of servomechanism in the control feedback loop, it was still unclear what factors affected human performance. Early work by Tustin and others extended the control theory framework and applied these theories to actual human operators [23]. Particular interest was focused on "attempt[ing] to find the laws of relationship of movement and error. In particular, it was hoped that this relationship [would] be approximately linear and so permit well developed theory of 'linear servomechanisms' to be applied to manual control in the same way as it applies to automatic following [23]." This would allow for the prediction of human performance and the ability to predict the limits of human control.

These early works were summarized in McRuer's 1957 report, "Dynamic Response of Human Operators" [24]. This work evaluated measurements for single-input/single-output (SISO) man-

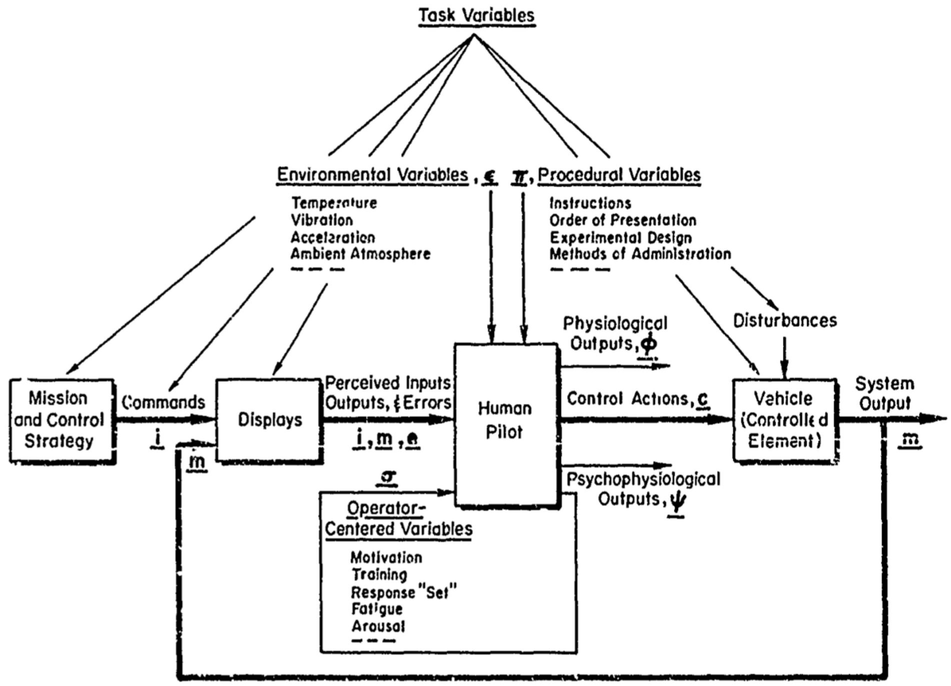


FIGURE 3: Variables affecting the pilot/vehicle system, from [22].

ual control systems and developed predictive models consistent with this data. Indeed, McRuer writes, "[i]t is possible, without doing violence to the data, to obtain describing functions which are generally applicable to the results of the many diverse experiments [24]." The report concludes by describing a hypothetical transfer function of the human operator which includes a time delay, a neuromuscular lag, and a gain. McRuer's early model of the complete pilot/vehicle system is presented in Figure 3. McRuer revisited these results in 1974, after three decades of supporting engineering and experimental psychology experiments and was able to further generalize these results to a wide variety of system dynamics [22]. In his study, McRuer completed a detailed analysis which included the human response to proportional, rate/velocity, and acceleration type controlled element dynamics, see Table 1. The result of this report was the now famous "crossover model," which relates the operator and controlled element transfer characteristics by the equation

$$Y_c(jw)Y_p(jw) = \frac{w_c e^{-jw\tau_e}}{jw} \tag{1}$$

where Y_c is the controlled element transfer function, Y_p is the approximate human operator transfer function, w_c is the crossover frequency, and τ_e is the effective time delay of the pilot. The crossover model is so named as it allows for linear behavior at approximately -20 dB/decade slope in the region of the crossover frequency. The approximate human operator response to several controlled element transfer functions and their combined open-loop transfer function are presented in Table 2. Modeling the human pilot with the crossover enabled a more complete view of the complete pilot/vehicle system, and allowed for human factors recommendations towards the design

Table 1: Example Applications of Idealized Controlled Element Forms, adapted from [22]

Controlled Element Form	Aerospace Control	Automobile Control
K_c	Attitude control with ACAH system	Speed control
$\frac{K_c}{s}$	Attitude control with a rate command system	Heading control at low to moderate speeds
$\frac{K_c}{s^2}$	Attitude control of a spacecraft with damper off	Longitudinal position control

Table 2: Summary of Human Operator Approximate Characteristics, adapted from [22]

Controlled Element Transfer Function Y_c	Approximate Human Operator Transfer Function Y_p	Open-Loop Transfer Function Y_cY_p
K_c	$\frac{K_p e^{-\tau_1 s}}{s}$	$\frac{w_c e^{-\tau_e s}}{s}$
$\frac{K_c}{s}$	$K_p e^{-\tau_2 s}$	$\frac{w_c e^{-\tau_e s}}{s}$
$\frac{K_c}{s^2}$	$K_p s e^{-\tau_3 s}$	$\frac{w_c e^{-\tau_e s}}{s}$

of new vehicles. Even today, the crossover model is used as the standard for describing pilot/vehicle systems at the crossover frequency [22, 25, 26].

The continued demand for human pilot models for use in informing vehicle design, as well predicting, preventing, and explaining accidents has led to a variety of more complex pilot models since the creation of the crossover model. A recent review by Xu et al. in 2017 surveyed the state of the art in human pilot modeling and grouped existing models into three classes of models based on: control theory, human physiology, and intelligence techniques [26]. Classical models based on control theory include the McRuer crossover model and optimal control models by Kleinman et al. developed in the early 1970s [27, 28]. Of these three overarching sets of models, the models based on human physiology are of the greatest interest here. Models based on human physiology were developed to understand human pilot perception and control behavior, and include the Hess structural model [29–31], Hosman's descriptive model [32, 33], and the biodynamic model [34]. Recent intelligence models take advantage of techniques including fuzzy control and neural networks [35, 36].

While the McRuer was very successful in predicting pilot behavior, it it did not attempt "to describe the underlying structure which contributes to human pilot dynamics [29]." For this reason, the Hess Structural Model is of particular interest due to the incorporation of multiple sensory channels and models of visual acuity and the time-varying human pilot [37]. The Structural Model includes the effects of the neuromuscular system, the force-feel characteristics of the input device, and the contributions of proprioceptive, vestibular, and visual feedback, see Figure 4. One of the key strengths of the Structural Model is the relatively few number of free parameters that need to be set to predict pilot performance. The model has been used in predicting and evaluating handling

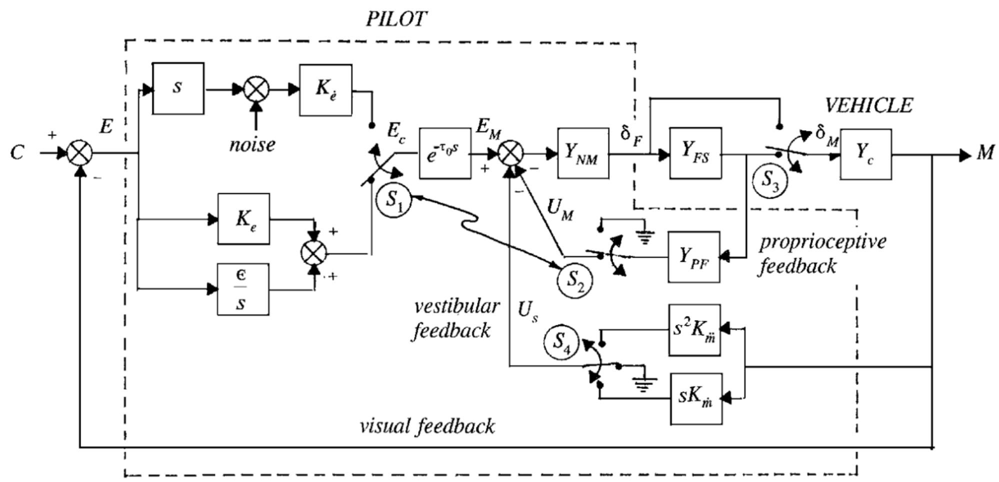


FIGURE 4: The Hess Structural Model of the Human Pilot, from [31].

qualities and pilot-induced oscillation rating levels for helicopters, Boeing 747, Lockheed C-5A, and twin ducted-fan aircraft [38–40]. Hess has also investigated how pilot control characteristics change with time due to flight anomalies, changing flight dynamics, and sudden increases in task demand [37, 41]. The results of this model have been compared to the results of a human-in-the-loop simulation for a well trained subject, and showed good comparison [41]. Recent work from Bachelder et al. has included modifications to the Structural Model to link pilot performance and workload and to enable the modeling of pulsive pilot behavior [42, 43].

3 Proposed Research

We propose to run human-in-the-loop subject testing experiments to understand the effects of concurrent bandwidth feedback, and to integrate the effects of this feedback into a human performance model. To investigate the three Aims outlined in the Problem Statement, we propose two experiments and the development of a model.

Experiment One investigates if concurrent bandwidth feedback can be used to teach novice subjects to improve performance in a three-axis manual tracking task.

Experiment Two investigates if concurrent bandwidth feedback can decrease the required learning time to peak performance in a simulated robotic arm track and capture task.

The Model will extend Professor Hess' structural model of the human pilot to include the effects of concurrent bandwidth feedback.

Concurrent bandwidth feedback has been used in a large variety of motor control tasks, and has generally been found to improve performance. Until recently, however, only simple tasks such as physical movements or low-dimensional pursuit tasks have been investigated. More recent works, including the lane-keeping task by de Groot et al., and our previous work with the SAFER task, have indicated that concurrent bandwidth feedback can also be quite effective for complex tasks. Unlike simple tasks, in which the guidance hypothesis dominates when feedback is removed, there is some evidence that concurrent bandwidth feedback can be removed after training without a loss of performance. The decrease in required learning time, improved performance, and decreased workload seen in the SAFER task show that concurrent bandwidth feedback may prove to be most useful very early in training when subjects are first exposed to complex, highly dynamic tasks. As concurrent bandwidth feedback can improve performance without an increase in workload, it may prove a useful technique for training other robotics tasks.

There has been considerable improvement in the field of pilot modeling since McRuer's crossover model, especially with models that incorporate human physiology. The Structural Model, in particular, has been very effective in predicting pilot performance, handling qualities, pilot-induced oscillation rating levels, and workload for a variety of system dynamics. None of these pilot models, however, are able to include the effects of concurrent bandwidth feedback. The performance improving effects of this feedback, seen throughout the literature, make this a compelling feature to be incorporated into a pilot model.

3.1 Experiment One

Experiment One investigated the effects of concurrent bandwidth feedback (CBF) and augmented reality displays on performance and workload in a three-axis manual tracking task. In our experiment, subjects were responsible for simultaneously completing three tracking tasks and a two-choice secondary task for workload measurement. Each axis of the tracking task was disturbed by a sum-of-sines, resulting in an random appearing signal that was difficult for the subjects to predict. The two-choice task appeared on a screen next to the tracking task and prompted subjects to respond to either a "LEFT" or "RIGHT" command. Subjects controlled the three-axes and responded to the two-choice task by using a Microsoft Xbox controller. The subjects used the left joystick on the controller to control the x and y axis tracking tasks. Subjects moved this joystick left and right to control the x axis, and up and down to control the y axis. The subjects controlled the z axis by moving the right joystick up and down. Subjects used the left and right triggers on the controller to respond to the two-choice task, using the left trigger to indicate "LEFT", and the right trigger to indicate "RIGHT".

3.1.1 Stereoscopic Displays

Stereoscopic displays are systems "in which two slightly different views of a scene are provided to a viewer, one image for each eye... allow[ing] the viewer's binocular visual system to extract depth information in a scene using this disparate information" [44]. Without the aid of the binocular depth cue presented by stereoscopic displays, viewers are instead reliant entirely on monocular clues such relative sizing, occlusion, and motion. One of the primary motivation for stereoscopic displays is that "[t]he visual scene of a 3D world is a more 'natural,' 'ecological,' or 'compatible' representation than that provided by 2D displays" [45]. As a result of this motivation, the effects of stereoscopic displays on human performance have been extensively studied in the literature. Several authors have attempted to classify which types of tasks may stand to benefit [44–48]. A recent review of 184 papers, for example, suggests that 60% of studies showed some benefit of 3D stereo displays, 15% of tasks showed unclear or mixed benefits, and 25% of studies showed no clear benefits [44]. In their review, tasks involving finding/identifying/classifying objects and tasks involving real/virtual spatial manipulations of objects benefited the most, while learning/training/planning tasks were the least likely to show a benefit.

Kim et al. also performed a quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking task [49]. They investigated the differences between perspective and stereoscopic displays, the elevation angle, azimuth angle, and the effects of two visual enhancements: a grid and a reference line. They found very strong relationships between elevation and azimuth angles and tracking performance, with the best performance occurring at an elevation angle of 45 degrees and an azimuth angle of 0 degrees. Tracking performance decreased rapidly as the azimuth angle varied, and decreased less rapidly as the elevation angle varied. In general, they found that the stereoscopic display allowed for better tracking performance, though the inclusion of the reference line visual enhancement greatly decreased the benefit over the perspective display. Using only two subjects, they provided some insight into intrasubject and intersubject variability. In several instances, intrasubject variability showed 50% changes within the same experimental condition, while intersubject variability also appeared quite large in some conditions. Kim et al. repeated the evaluation of these parameters on a telerobotics pick and place study [50]. They found similar results in this second study, suggesting that their results could be generalized and that three-axis tracking performance can be correlated with pick and place completion time.

Smallman at al. similarly investigated the effect of visual enhancements and 2D vs 3D displays for the development of a naval air warfare console [51]. Participants viewed naval and aircraft tracks in either a conventional 2D top-down display or a 3D display, and then attempted to reconstruct track positions. They investigated the effectiveness of drop-lines and drop-shadows, and found that they significantly improved subjects ability to localize aircraft compared to when the enhancements were not present. Furthermore, in the absence of either visual enhancement, subjects performed better with the 2D display than the 3D display. Similar to Kim et al., they ultimately recommended that 3D stereoscopic displays include the use of a reference or drop-line for optimal performance.

3.1.2 Hypotheses

This study assessed the influence of display type (perspective vs. stereoscopic), relative display attitude (zero degrees vs. thirty degrees), and concurrent bandwidth feedback (with vs. without) on performance and workload. Objective performance was measured using the root-mean-square error (RMSE) of each of the three axes individually and combined, and subjective performance was measured with the use of a questionnaire. Objective workload was measured using the response time to the secondary task, and subjective workload was measured using the NASA-TLX. It was hypothesized that:

Hypothesis 1 Concurrent bandwidth feedback will improve performance in the depth (z) axis for both display types, and will decrease workload.

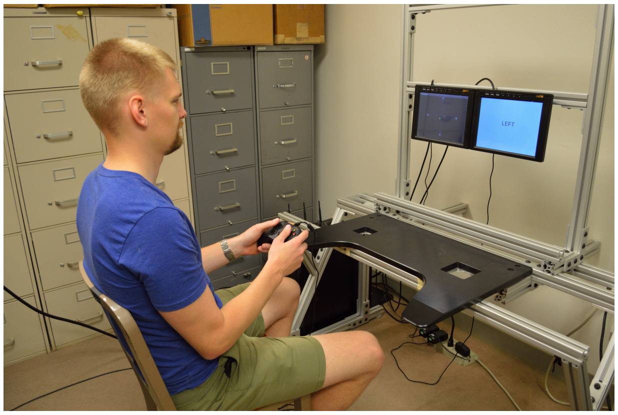
Hypothesis 2 Stereoscopic augmented reality displays improve performance in the depth (z) axis, but do not affect workload.

Hypothesis 3 Rotating the display improves performance in the depth (z) axis for both display types, and will decrease workload.

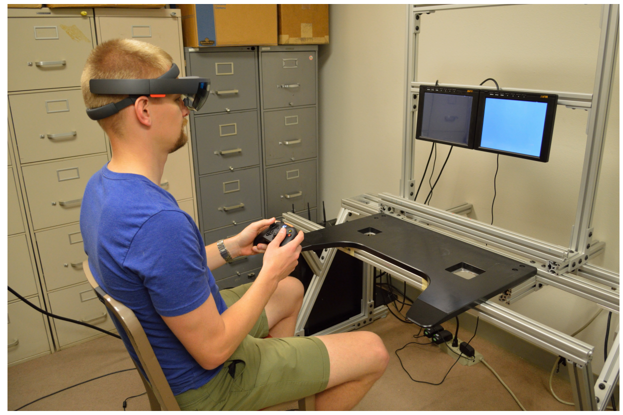
3.1.3 Procedure

A total of 24 subjects (19 males, 5 females) were recruited in accordance with the University of California, Davis Internal Review Board (IRB), and subjects signed a consent form and were not compensated. There were 12 subjects in the 2D group, and 12 subjects in the 3D group. Subjects were undergraduate and graduate students in the University of California, Davis College of Engineering. All participating subjects had normal vision (no colorblindness, eyesight correctable to 20/20 vision) and full motor control of their hands.

A human-in-the-loop simulation was conducted using a fixed-base simulator, see Figure 5. The simulator consisted of two 10.4 inch LCD displays. The tracking task was completed on the left display, while the right display showed the two-choice task. Subjects were seated for the duration



(a) 2D Group



(b) 3D Group

FIGURE 5: The fixed-based simulator used by both groups.

of the experiment and were placed one meter perpendicular from the center of the left display. For subjects in the 3D group, the left LCD monitor was turned off, and the tracking task was instead displayed on the HoloLens. For these subjects, the cross was placed at the same height as the subject's head, such that it was viewed with no relative attitude in the baseline condition. Both groups viewed the guidance cross of a width and height of 5 inches by 5 inches. For subjects in the HoloLen's group, the z motion of the cross also could move up to 5 inches in either direction away from the center of the guidance cross. Subjects in both groups used the same Microsoft Xbox controller and control scheme to complete the task.

The disturbance function was a sum of 13 sinusoids approximating a rectangular spectrum with a 2.0 rad/s cutoff frequency. The disturbing force, as a function of time, $d(t)$, was

$$d(t) = \sum_{i=1}^{13} A_i \sin\left(w_i t + \phi_i\right) \tag{2}$$

Each sine wave amplitude, frequency, and phase offset was borrowed from a similar experiment [52]. This disturbance was the same for all subjects and trials, though the subjects were naive to this. The x, y, and z axes all experienced the same disturbance force generating function, but the y axis was temporally offset by 60 seconds and the z axis was offset by 120 seconds. This allowed for a very similar generation of disturbance forces for each axis. The RMSE of the disturbance force was normalized along each axis such that all three were the same.

Three designs were presented to the subjects to evaluate: a baseline design, a color-based concurrent bandwidth feedback (CBF) design, and a rotated design. Figure 6 shows all three designs in the same error state. The three designs were very similar, having only minor differences between each other.

The baseline design consists of a flat cross with a center target point and a green sphere error indicator. This indicator also casts a green, variable-length rod perpendicular to the plane of the cross, which allows for a visual estimation of the error in the z axis. The x axis is parallel with the horizontal cross, while the y axis is parallel with the vertical cross. The color feedback display was

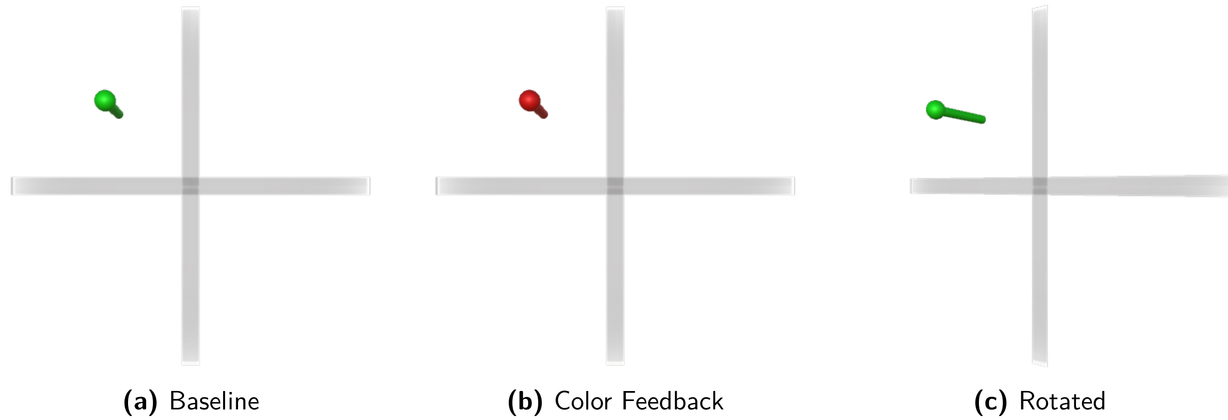


FIGURE 6: The three different designs in the same error state. (b) The color feedback has been activated here, changing the guidance target from green to red.

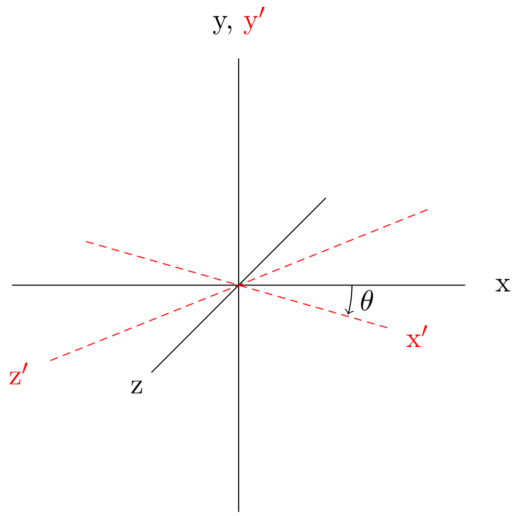


FIGURE 7: Perspective display of the coordinate frame for the tracking tasks, with the x, y, and z axes labeled. After rotating by θ around the y axis, the resulting reference frame of $x'y'z'$ is also labeled.

identical to the baseline design in every way, with the additional of visual concurrent bandwidth feedback on the z axis. When the absolute value of the error on the z axis exceeded a fixed bandwidth of 0.2 units, the color of both the spherical indicator and the cylindrical rod changed from green to red. When the absolute value of the error on the z axis was lowered back below this fixed bandwidth, the indicator changed back to a green color. The rotated display was identical to the baseline design, but the relative attitude of the display was rotated about the y axis by 30 degrees, see Figure 7.

Before entering the study, subjects were randomly placed in a display group (either the LCD monitor or HoloLens), and were then randomly placed into an order group (which consisted of Baseline-Feedback-Rotated, Feedback-Rotated-Baseline, and Rotated-Baseline-Feedback). This order group was created to remove any order effects that might arise due to training on a given display, and follows a standard Latin squares design. (The order of the designs was expected to be insignificant, but we will later discuss how this is not the case.) After entering the experiment room, the subjects were familiarized with the task, three designs, NASA-TLX, and the controller through a twenty minute training session. They were instructed to:

- Minimize the displacement of their guidance target from the center
- Respond to the two choice task as accurately and quickly as possible

Subjects in the 3D group completed a short calibration program that adjusted the display to their interpupillary distance. All subjects were then allowed to complete two familiarization trials, during which time they could ask questions about how the controls worked, or any other aspects of the task. The proctor also used this time to ensure that subjects showed basic competency with the task by responding to both the tracking and two-choice tasks appropriately. All familiarizations were done with the baseline design, regardless of which design the subjects evaluated first.

After this familiarization process, subjects completed ten trials with their first design. After completing these trials, they answered a brief survey which asked them if the design was adequate to complete the task. Subjects were also asked to subjectively rate their performance in a questionnaire after evaluating each design. Subjects were asked to rate “I found the tracking display adequate to complete the task.” on a five point scale where 1 indicated “Strongly Disagree” and 5 indicated “Strongly Agree”. After this survey, subjects then completed a NASA-TLX workload survey. Subjects then repeated this process with their second and third designs. At the conclusion of the three designs, subjects were also asked to complete a preference survey which inquired into what design the subjects preferred.

3.1.4 Results

We conducted three-way mixed ANOVAs between display (2D or 3D), design (Baseline, Feedback, or Rotated), and starting design (Baseline, Feedback, or Rotated) with repeated measures on the design factor. When significant effects were observed, post hoc comparisons using the Tukey Honest Significance Difference (HSD) test with a Bonferroni adjustment were completed to investigate which pairs of the factor were significant. In order to remove learning and fatigue effects, each subject’s best performing five trials in each design were averaged together to produce one average score for each subject and design. Additionally, the first ten seconds of each sixty second trial were not included in the analysis to remove initial transient effects.

The root-mean-square error (RMSE) of each axis was analyzed individually and combined to understand the differences between the three designs and the two devices. The RMS of the disturbance signal was calculated along each axis and used to normalize the RMSE. Under this definition, an RMSE of 1 indicates performance no better than no input, and an RMSE greater than 1 indicates quite poor performance. It was expected that the baseline design would lead to the worst performance in the z axis than the color feedback and rotated designs. It was also expected that, due to the stereoscopic nature of the display, the HoloLens would allow for better performance than the LCD monitor along z axis.

Results of the ANOVA on the z-axis RMSE showed significant effects for design ($F(2, 36) = 84.92, p < .001$), device ($F(1, 18) = 7.22, p < 0.015$), and start design ($F(2, 18) = 4.81, p < 0.021$). The ANOVA also showed a significant interaction effect between design and starting design ($F(4, 36) = 8.55, p < 0.0001$), and a three way interaction between design, device, and starting design ($F(4, 36) = 5.57, p < 0.002$). Further investigation into the effect of starting design using

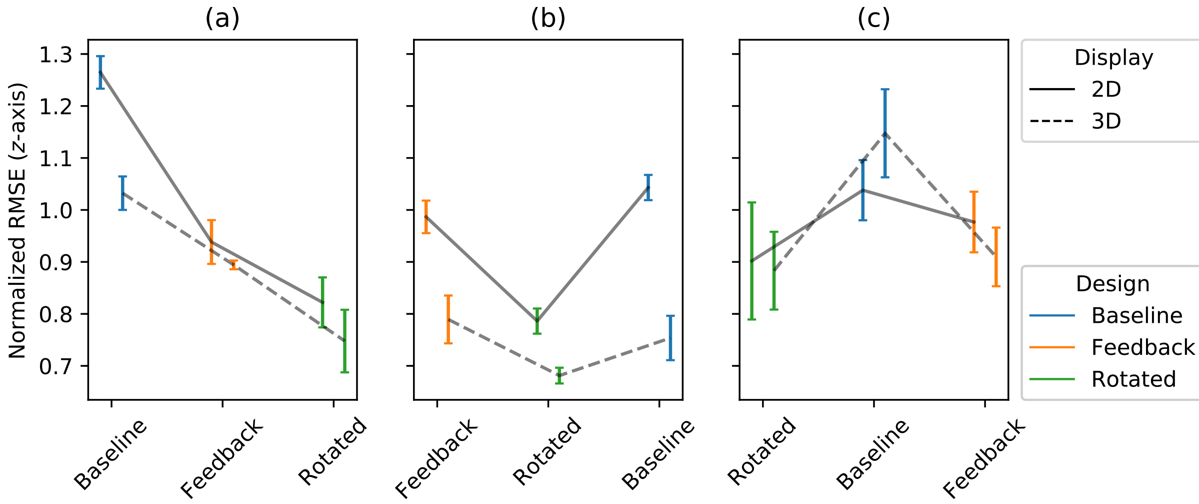


FIGURE 8: The resulting normalized RMSE along the z-axis. Subjects started in either the Baseline (a), Feedback (b), or Rotated (c) design.

a Tukey HSD comparison with a Bonferroni correction showed significance differences between subjects that started in the concurrent bandwidth feedback group and those in the baseline ($p < 0.001$) or rotated ($p < 0.001$) designs, but no difference between subjects that started in the baseline and rotated designs ($p > .38$). The difference in performance along the depth axis between design, device, and starting design can be seen in Figure 8.

Due to the presence of interaction effects in the starting design factor, the remainder of the analysis is split between subjects who started with the concurrent bandwidth feedback design and those that did not (e.g., those that started in the baseline or rotated design). For subjects that started in the CBF design, there was a significant effect of design ($F(2, 12) = 21.65, p < .0001$), a significant effect of display ($F(1, 6) = 36.73, p < 0.001$), and a significant interaction effect between design and device ($F(2, 12) = 5.42, p < 0.021$). For subjects that did not start in the CBF design, there was a significant effect of design ($F(2, 28) = 38.70, p < .0001$), but no significant effect or display ($F(1, 14) = 1.03, p > 0.32$), and no interaction effect ($F(2, 12) = 0.03, p > 0.97$). The resulting difference between subjects who started in the CBF design and those that did not is presented in Figure 9. For subjects that did not start in the CBF design, display had no significant effect, and subjects performed best in the rotated design, followed by the feedback design and finally the baseline design. Subjects that started in the CBF design performed significantly better in the 3D display than the 2D displays, and performed best in the rotated design, but comparably between the feedback and baseline designs.

The NASA-TLX was used to measure differences in subjective workload, and the reaction time to the secondary task was used to measure differences in objective workload between design, device, and starting design. There were no significant effects, nor interaction effects, found in the ANOVA for design, device, or starting design for the NASA-TLX measurements. There was a significant effect of design ($F(2, 36) = 7.93, p < 0.0014$) for the reaction time to the secondary task, though the magnitude of this effect was very small between designs (less than 100 ms difference) and was not

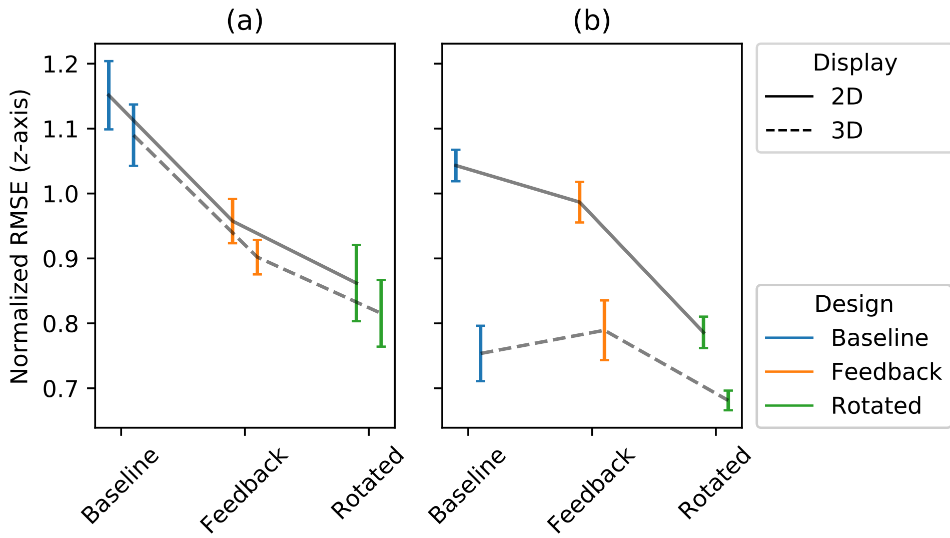


FIGURE 9: The resulting normalized RMSE along the z-axis, grouping by subjects that started without (a) or with (b) feedback.

significant during Tukey HSD tests. In general, there were no significant effects found for workload measurements.

To summarize these results, there were significant effects found in the z-axis RMSE for design, with subjects generally performing the best using the angled design, followed by the CBF design, and performing worst with the baseline design. There were significant effects found for the factors of device and starting design, though these must be interpreted carefully. Subjects that started in the design with concurrent bandwidth feedback performed better than subjects that did not. We believe that this result resembles that found in our prior SAFER experiment, where subjects that were exposed to the CBF early on learned the task better than those that were not exposed. An interesting effect of this exposure is that, after learning the task with CBF, subjects continued on to perform significantly better in the baseline condition than those subjects that did not start in the CBF design.

Subjects that started in the CBF design and that were wearing the HoloLens appear to have used the CBF to better learn the depth cue presented in the stereoscopic display. These subjects continued to perform significantly better than subjects who started with the CBF design but without the stereoscopic display when they continued to the baseline design. This indicates that even a brief exposure to the concurrent bandwidth feedback was sufficient to induce improved performance in the baseline design. Additionally, this also suggests that well trained subjects could perform better using the stereoscopic display compared with the traditional display, but that subjects who were still learning the task could not take advantage of the additional depth cues provided by the display. Finally, there were no significant effects found for the NASA-TLX measurements, and there were significant effects found between the designs for reaction time.

In summary, we find partial agreement with all of our hypotheses in respect to the performance aspects of our experiment, while the workload was essentially unaffected by all of our experimental

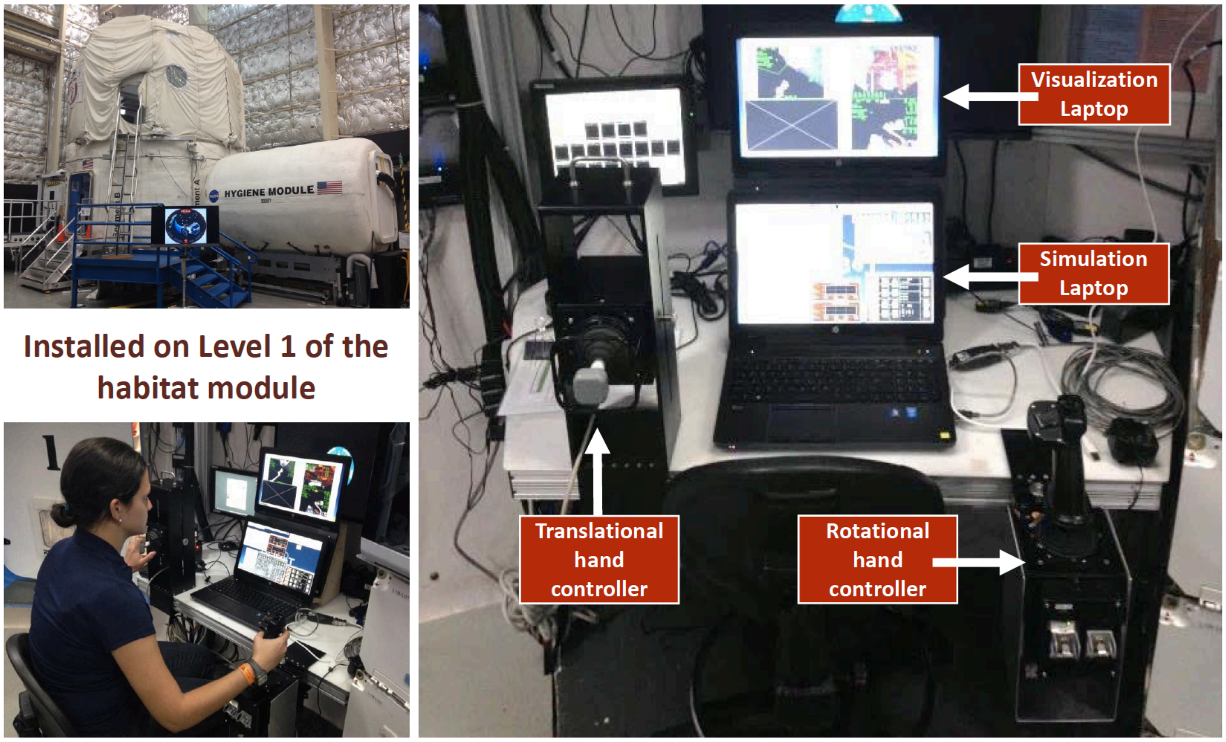


FIGURE 10: The Robotic On-board Trainer (ROBoT) station set up in the NASA HERA Analog, from [53].

factors. For Hypothesis 1, subjects that completed the baseline design before the CBF design performed better in the CBF design, while subjects that completed the CBF design before the baseline performed approximately the same in both designs. This indicates that CBF can both improve performance compared to a baseline design, and better train subjects such that, even after brief exposure, the feedback is no longer required. Workload was unaffected by the CBF. For Hypothesis 2, subjects that started in the CBF design appear to have better learned the task and used the CBF to learn to interpret the depth cue provided by the stereoscopic display. Subjects that were not initially exposed to this feedback were unable to exploit the display, and did not perform significantly better than subjects without the display. For Hypothesis 3, subjects in the rotated design did perform better than those in the baseline design and appeared to be able to use this view to better interpret the depth of the target.

3.2 Experiment Two

The second Aim of the proposal is addressed by Experiment Two. Experiment Two will focus on if concurrent bandwidth feedback can decrease the required learning time to peak performance in a simulated robotic arm track and capture task. We will compare task performance throughout training and workload between two groups: a control group, which receives no feedback, and a treatment group, which receives concurrent bandwidth feedback on one or more sensor readouts. Subjects in both groups will complete the same task, and it is hypothesized that subjects in the treatment group will perform the task better and require less training time to reach peak performance. This hypothesis is based on the results of the SAFER experiment, as well as the results of Experiment One.

In this task, subjects will command a robotic arm to track an approaching vehicle and grapple with it. This is motivated by the primary use of the robot arm on the International Space Station, which grasps visiting vehicles when they arrive at the station and then attaches them to a separate fixed location on the station. Subjects will be trained on this task and will then repeat the task for one to two hours through a variety of slightly different start and end conditions.

NASA's Robotic On-board Trainer (ROBoT) will be used for this experiment. ROBoT is a package of simulation software which includes a dynamic model of the robotic arm on the space station and presents the user with multiple camera angle views and the instrument panel required to operate the robotic arm, see Figure 10. In addition to these displays, ROBoT also includes the two hand controllers required to control the arm. NASA trainers at Johnson Space Center developed metrics for performance which are included in ROBoT and include the time to capture, the alignment measurements during approach, the amount of wobble in the arm, the number of times the grapple fixture contacted the structure, the overall path efficiency, and the number of capture attempts. The preceding list presents candidate metrics from which we may observe human performance.

3.2.1 Hypotheses

This study will assess the influence of concurrent bandwidth feedback (with vs. without) on performance and workload. Objective performance data will be measured using time to capture, alignment error metrics, and other metrics available by the ROBoT system. Subjective workload will be measured using the NASA-TLX at several points throughout the experiment. It is hypothesized that:

Hypothesis 1 Concurrent bandwidth feedback will improve performance in the track-and-capture task.

Hypothesis 2 Concurrent bandwidth feedback will cause subjects to more quickly reach their peak performance in the track-and-capture task.

Hypothesis 3 Concurrent bandwidth feedback will decrease workload in the track-and-capture task.

3.2.2 Previous Work

We have experience using the ROBoT simulation software from participating in a study investigating the effects of sleep loss and circadian misalignment on performance on the ROBoT simulator at NASA Ames [54]. In this study, subjects trained during one-hour long sessions each day for five days, then spent a twenty four hour period in the lab while they continued to perform robotic tasks. We did not find evidence of performance loss during sleep deprivation or circadian misalignment on any of the ROBoT performance metrics. In fact, participants continued to show improvement over time, which indicated that they had continued to learn the task despite the sleep loss [54]. This finding reinforces the need for enhanced learning techniques, as the current training strategy requires a tremendous amount of time to reach peak performance.

3.3 Model Extension

The third and final Aim of the proposal is addressed by developing a model of the human pilot which includes the effects of concurrent bandwidth feedback. The proposed Model will extend Professor Hess' structural model of the human pilot to include the effects of concurrent bandwidth feedback.

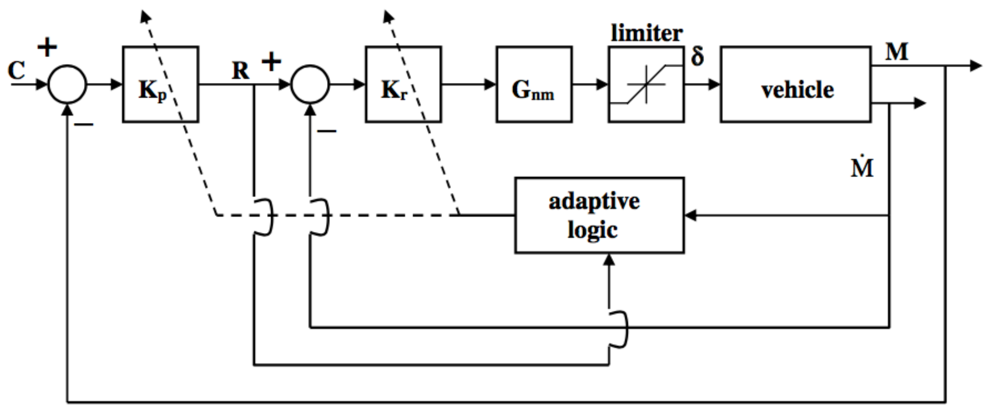


FIGURE 11: Hess' model of the adaptive human pilot, from [37].

The Structural Model has been extremely successful in predicting human performance through a variety of system dynamics and can predict how performance changes during a pilot's adaptation to changing dynamics. Hess has developed adaptive logic for the human pilot in a pursuit task which triggers when the pilot notices that vehicle dynamics have changed [37]. This logic is based off several criteria, which "must be predicated upon information available to the human [and] the postadapated pilot models must follow the dictates of the crossover model of the human pilot [37]." The primary result of the adaptive logic is to increase the resulting crossover frequency of the pilot, effectively making them more responsive, which could be interpreted as more focused on the task. Our initial approach to adding concurrent bandwidth feedback into the Structural Model will be based off of Hess' approach to modeling human adaptation in pursuit tasks, which is currently ad-hoc in nature, see Figure 11.

While this model of the adaptive pilot has been successful in predicting changes in performance for a well trained subject, it does not consider how a pilot would behave when they are still in the early stages of training. Our modified model will include two major changes to Hess' current model:

- The adaptation logic will be changed to focus on concurrent bandwidth feedback
- The timescale of the adaptation will be significantly longer

We propose to modify the adaptive logic to trigger when the pilot is receiving concurrent bandwidth feedback, rather than when a change in system dynamics occurs. This will require the addition of a feedback loop onto the Structural Model which triggers when the bandwidth feedback is activated. This loop will likely be based around the K_e gain, which is currently the primary way of setting the crossover frequency in the Structural Model. This implies that the subjects in our experiments do their primary learning when they are receiving qualitative feedback that their current level of aggressiveness is not sufficient to complete the task. While there must be a separate loop that adjusts the crossover frequency as learning progresses over the course of several hours, the change in performance we see when subjects use the concurrent bandwidth feedback happens relatively rapidly, within a few minutes. This is reflected in the delta of performance between subjects in the different groups of our SAFER experiment, even on the first trial, see Figure 2a. This relates to the second required change, the amount of required adaptation time. Professor Hess' model requires that pilots adapt within a very short time period, on the order of 5 seconds [55]. The results of our experiment with SAFER and the three-axis tracking task also suggest relatively short adaptation times, though they are on the order of a few minutes, again see Figure 2a.

While it should be noted that this work is in preliminary stages and not scheduled to begin until the Fall Quarter, some early work has been started in preparation for the qualifying examination. An effort has been made to begin to replicate some of Hess' results. Working with Professor Hess, we have been able to replicate some of the existing adaptation logic for a two-axis tracking task, and begun exploratory research into modifying the model.

4 Timeline and Risks

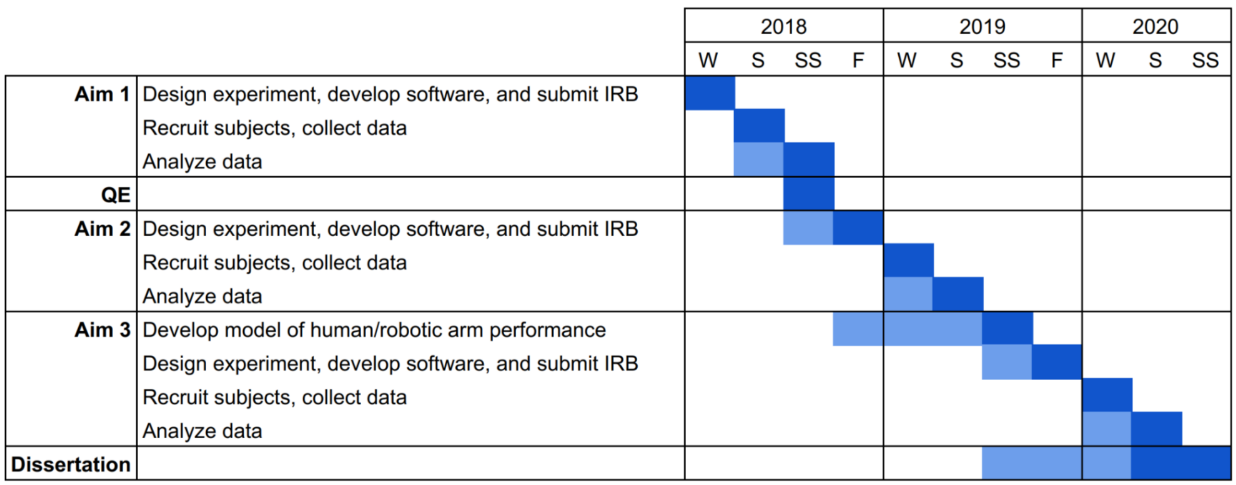


FIGURE 12: The proposed timeline for this research.

The proposed timeline for the remainder of this research is available in Figure 12. This timeline shows the past few months spent accomplishing the first aim and outlines our estimated timeline to complete the remaining aims, with an estimated graduation date of Summer Quarter 2020. While not present on the timeline, we also plan to write and submit several conference papers and journal articles throughout the completion of this research.

A conference paper, "Evaluating Augmented Reality in a Three-Axis Manual Tracking Task", J. Karasinski and S. Robinson has recently been submitted for consideration at AIAA SciTech 2018. A draft journal article on the SAFER experiment is being prepared for submission in the journal of Human Factors. We expect to publish an additional conference paper to an AIAA or IEEE conference and journal article for the results of the robotic arm experiment, likely in Human Factors. If we are successful in creating a model which closely follows human behavior, we would also publish this as a journal article, likely in the Journal of Guidance, Control, and Dynamics.

This schedule is flexible to changes in plans, and was designed with the full awareness that we may face setbacks. There are a number of places where we may suffer delays or other issues:

- While the author has experience subject testing over one hundred subjects over three hundred hours of experiment time, it is difficult to predict how long subject testing will take to complete. Under the assumption that we will run approximately twenty to thirty subjects in the robotic arm experiment, and that the experiment will last for approximately two to three hours per subject, we have budgeted one quarter to recruit subjects and run them through the experiment. This could, however, easily stretch to two quarters, depending on subject availability and success rates.

- NASA Ames has agreed to allow the Human/Robotics/Vehicle Integration and Robotics Lab to use the ROBoT simulator for this proposed experiment at UC Davis. If, for whatever reason, we are unable to use ROBoT, we could either run the experiment at Ames or attempt to build our own robotic arm simulation.
- While we believe we have sufficient evidence that our feedback techniques will improve performance, it is also possible that we will not see significant effects in the robotic arm task. If this is the case we will need to further investigate how this task is different from our previous successes. This would allow us, for instance, to make begin to make recommendations for which types of tasks can and cannot benefit from this feedback.
- It is possible that we may be unable to create a model which incorporates feedback and that accurately mirrors the effects we have seen in our experiments. This may require us to investigate other types of performance models or to create a novel technique.

References

[1] N. Wierner, *Cybernetics (or the control and communication in the animal and the machine)*, vol. 1. 1948.

[2] A. Ramaprasad, "On the Definition of Feedback," *Systems Research and Behavioral Science*, vol. 1, no. 1, pp. 4–13, 1983.

[3] D. Laurillard, *Rethinking University Teaching : A Conversational Framework for the Effective Use of Learning Technologies*. Routledge, 2002.

[4] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review," *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 21–53, 2013.

[5] E. L. Thorndike, "The Law of Effect," *American Journal of Psychology*, vol. 39, no. 1/4, pp. 212–222, 1927.

[6] A. W. Salmoni, R. A. Schmidt, and C. B. Walter, "Knowledge of results and motor learning: A review and critical reappraisal," *Psychological Bulletin*, vol. 95, no. 3, pp. 355–386, 1984.

[7] R. B. Payne and G. T. Hauty, "Effect of psychological feedback upon work decrement," *Journal of Experimental Psychology*, vol. 50, no. 6, pp. 343–351, 1955.

[8] N. B. Gordon and M. J. Gottlieb, "Effect of supplemental visual cues on rotary pursuit.," *Journal of Experimental Psychology*, vol. 75, no. 4, pp. 566–568, 1967.

[9] A. C. Williams and G. E. Briggs, "On-target versus off-target information and the acquisition of tracking skill," *Journal of Experimental Psychology*, vol. 64, no. 5, pp. 519–525, 1962.

[10] S. de Groot, J. C. F. de Winter, J. M. L. García, M. Mulder, and P. A. Wieringa, "The Effect of Concurrent Bandwidth Feedback on Learning the Lane-Keeping Task in a Driving Simulator," *Human Factors*, vol. 53, no. 1, pp. 50–62, 2010.

[11] R. Fischer and J. Miller, "Differential redundancy gain in onset detection versus offset detection," *Perception & Psychophysics*, vol. 70, no. 3, pp. 431–436, 2008.

[12] J. A. Karasinski, "Real-Time Performance Feedback for the Manual Control of Spacecraft," Master's thesis, University of California, Davis, 2016.

[13] J. Karasinski, S. Robinson, P. Handley, and K. Duda, "Real-Time Performance Feedback in a Manually-Controlled Spacecraft Inspection Task," in *AIAA Modeling and Simulation Technologies Conference*, 2017.

[14] J. A. Karasinski, S. K. Robinson, K. R. Duda, and Z. Prasov, "Development of real-time performance metrics for manually-guided spacecraft operations," in *IEEE Aerospace Conference Proceedings*, vol. 2016-June, 2016.

[15] K. Vassigh, D. Weigel, and T. Mack, "USA simplified aid for EVA rescue (SAFER) operations manual," *NASA Johnson Space Center, Systems Division, EVA Systems Group, JSC-26283 Rev. A*, 1998.

[16] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, vol. 52, no. C, pp. 139–183, 1988.

[17] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[18] V. J. Gawron, *Human Performance, Workload, and Situational Awareness Measures Handbook, Second Edition*. CRC Press, 2008.

[19] G. K. Slocum, B. H. Williges, S. N. Roscoe, and H. Stanley, "Meaningful shape coding for aircraft switch knobs," *Aviation Research Monographs*, vol. 1, no. 3, pp. 27–40, 1971.

[20] R. C. Williges and W. W. Wierwille, "Behavioral Measures of Aircrew Mental Workload," *Human Factors*, vol. 21, no. 5, pp. 549–574, 1979.

[21] R. J. Lysaght, S. G. Hill, a. O. Dick, B. D. Plamondon, P. M. Linton, W. W. Wierwille, a. L. Zaklad, a. C. Bittner Jr, and R. J. Wherry, "Operator workload: Comprehensive review and evaluation of operator workload methodologies," tech. rep., ANALYTICS INC WILLOW GROVE PA, 1989.

[22] D. T. McRuer and E. S. Krendel, "Mathematical Models of Human Pilot Behavior," tech. rep., ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT NEUILLY-SUR-SEINE (FRANCE), 1974.

[23] A. Tustin, "An investigation of the operator's response to manual for the human operator as a controller and decision maker in control of a power driven gun," in *Metropolitan Vickers Electrical Manned Aerospace Systems," in Proc. AGARD Conf*, no. 114, 1944.

[24] D. T. McRuer and E. S. Krendel, "Dynamic Response of Human Operators," tech. rep., KELSEY-HAYES CO INGLEWOOD CA CONTROL SPECIALISTS DIV, 1957.

[25] D. T. McRuer and E. S. Krendel, "Human Pilot Dynamics in Compensatory Systems," tech. rep., 1965.

[26] S. Xu, W. Tan, A. V. Efremov, L. Sun, and X. Qu, "Review of control models for human pilot behavior," *Annual Reviews in Control*, vol. 44, pp. 274–291, 2017.

[27] D. Kleinman, S. Baron, and W. Levison, "An optimal control model of human response part I: Theory and validation," *Automatica*, vol. 6, no. 3, pp. 357–369, 1970.

[28] S. Baron, D. Kleinman, and W. Levison, "An optimal control model of human response part II: Prediction of human performance in a complex task," *Automatica*, vol. 6, no. 3, pp. 371–383, 1970.

[29] R. A. Hess, "Structural Model of the Adaptive Human Pilot," *Journal of Guidance, Control, and Dynamics*, vol. 3, no. 5, pp. 416–423, 1980.

[30] R. A. Hess, "Model for human use of motion cues in vehicular control," *Journal of Guidance, Control, and Dynamics*, vol. 13, no. 3, pp. 476–482, 1990.

[31] R. A. Hess, "Unified Theory for Aircraft Handling Qualities and Adverse Aircraft-Pilot Coupling," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 6, pp. 1141–1148, 1997.

[32] R. J. A. W. Hosman, *Pilot's Perception and Control of Aircraft Motions*. PhD thesis, 1996.

[33] R. Hosman and H. Stassen, "Pilot's perception in the control of aircraft motions," *Elsevier*, vol. 7, pp. 1421–1428, 1999.

[34] M. J. Griffin, "The validation of biodynamic models," *Clinical Biomechanics*, vol. 16, no. 1, pp. S81–S92, 2001.

[35] K. B. Zaychik, F. M. Cardullo, and G. George, "A Conspectus on Operator Modeling: Past, Present and Future," in *AIAA Modeling and Simulation Technologies Conference and Exhibit*, (Reston, Virigina), pp. 1–17, American Institute of Aeronautics and Astronautics, 2006.

[36] M. Gestwa and J. Bauschat, "On the Modelling of a Human Pilot Using Fuzzy Logic Control," *IGI Global*, pp. 148–167, 2003.

[37] R. A. Hess, "Modeling Pilot Control Behavior with Sudden Changes in Vehicle Dynamics," *Journal of Aircraft*, vol. 46, no. 5, pp. 1584–1592, 2009.

[38] R. A. Hess and R. Joyce, "Analytical Investigation of Transport Aircraft Handling Qualities," in *AIAA Atmospheric Flight Mechanics (AFM) Conference*, (Reston, Virginia), American Institute of Aeronautics and Astronautics, 2013.

[39] A. I. Afloare and A. Ionita, "Prediction of the handling qualities and pilot-induced oscillation rating levels," *Incas Bulletin*, vol. 6, pp. 3–13, 2014.

[40] O. L. Grant, K. A. Stol, A. Swain, and R. A. Hess, "Handling Qualities of a Twin Ducted-Fan Aircraft: An Analytical Evaluation," *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 6, pp. 1126–1131, 2015.

[41] R. A. Hess, "Modeling Human Pilot Adaptation to Flight Control Anomalies and Changing Task Demands," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 3, pp. 655–666, 2016.

[42] E. Bachelder, R. Hess, M. Godfroy-Cooper, and B. Aponso, "Modeling Pilot Pulse Control," 2017.

[43] E. N. Bachelder, R. A. Hess, M. Godfroy-Cooper, and B. L. Aponso, "Linking the Pilot Structural Model and Pilot Workload," in *2018 AIAA Atmospheric Flight Mechanics Conference*, (Reston, Virginia), American Institute of Aeronautics and Astronautics, 2018.

[44] J. McIntire, P. Havig, and E. Geiselman, "Stereoscopic 3D displays and human performance: A comprehensive review," *Displays*, vol. 35, no. 1, pp. 18–26, 2014.

[45] C. D. Wickens, "Three-dimensional stereoscopic display implementation: Guidelines derived from human visual capabilities," in *Stereoscopic displays and applications*, vol. 1256, pp. 2–12, International Society for Optics and Photonics, 1990.

[46] C. D. Wickens, S. Todd, and K. Seidler, "Three-dimensional displays: Perception, implementation, and applications," tech. rep., CREW SYSTEM ERGONOMICS INFORMATION ANALYSIS CENTER WRIGHT-PATTERSON AFB OH, 1989.

[47] N. Naikar, "Perspective displays: A review of human factors issues," tech. rep., DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION CANBERRA (AUSTRALIA), 1998.

[48] S. Dixon, E. Fitzhugh, and D. Aleva, "Human factors guidelines for applications of 3D perspectives: a literature review," in *Display Technologies and Applications for Defense, Security, and Avionics III*, vol. 7327, p. 73270K, International Society for Optics and Photonics, 2009.

[49] W. S. Kim, S. R. Ellis, M. E. Tyler, B. Hannaford, and L. W. Stark, "Quantitative Evaluation of Perspective and Stereoscopic Displays in Three-Axis Manual Tracking Tasks," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 17, no. 1, pp. 61–72, 1987.

[50] W. S. Kim, F. Tendick, and L. W. Stark, "Visual Enhancements in Pick-and-Place Tasks: Human Operators Controlling a Simulated Cylindrical Manipulator," *IEEE Journal on Robotics and Automation*, vol. 3, no. 5, pp. 418–425, 1987.

[51] H. Smallman, E. Schiller, and M. Cowen, "Track location enhancements for perspective view displays," tech. rep., SPACE AND NAVAL WARFARE SYSTEMS CENTER SAN DIEGO CA, 2000.

[52] R. A. Hess, "Effects of Time Delays on Systems Subject to Manual Control," *Journal of Guidance, Control, and Dynamics*, vol. 7, no. 5, pp. 640–640, 1984.

[53] V. Ivkovic, G. Alexander, D. Cefaratti, B. Sommers, D. F. Dinges, and G. E. Strangman, "Bevahioral Core Measures: Robotic On-Board Trainer (ROBoT) Task Performance," Presented at the Human Research Program's Investigators' Workshop, Galveston, TX, 2017.

[54] E. Flynn-Evans, L. Wong, S. Pradhan, J. Karasinski, C. Hu, G. Strangman, V. Ivkovic, M. Anderson, and L. Stone, "Performance on the Robotic On-Board (ROBoT) Simulator during Sleep Loss and Circadian Misalignment," tech. rep., NASA, 2018.

[55] D. H. Weir and A. V. Phatak, "Model of Human-Operator Response to Step Transitions in Controlled Element Dynamics," in *Proceedings of the Second Annual NASA-University Conference on Manual Control*, no. SP-128, p. 65–83, NASA, 1996.