A Device-Independent Interaction Framework Towards the Implementation of Reduced Crew Operations

Sterling, Millie

(2020)

DOI (TUprints): https://doi.org/10.25534/tuprints-00012430

Publikationstyp: Ph.D. Thesis

Fachbereich: 16 Department of Mechanical Engineering

Quelle des Originals: https://tuprints.ulb.tu-darmstadt.de/12430

A Device-Independent Interaction Framework Towards the Implementation of Reduced Crew Operations

Technologieunabhängiges Interaktionsmodell zur Implementierung zukünftiger Cockpitoperationen mit reduzierter Besatzung
Zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation von Millie Irene Sterling, M.Sc. aus Las Vegas, NV, USA
Tag der Einreichung: 2. Juli 2019, Tag der Prüfung: 5. November 2019

1. Gutachten: Prof. Dr.-Ing. Uwe Klingauf
2. Gutachten: Prof. Dr. Daria Kotys-Schwartz
Darmstadt - D17

TECHNISCHE UNIVERSITÄT DARMSTADT

A Device-Independent Interaction Framework Towards the Implementation of Reduced Crew Operations
Technologieunabhängiges Interaktionsmodell zur Implementierung zukünftiger Cockpitoperationen mit reduzierter Besatzung

Doctoral thesis by Millie Irene Sterling, M.Sc.

1. Review: Prof. Dr.-Ing. Uwe Klingauf
2. Review: Prof. Dr. Daria Kotys-Schwartz

Date of submission: 2. Juli 2019
Date of thesis defense: 5. November 2019

Darmstadt - D17

For my mom.

Darmstadt, Germany, on November 5^{th}, 2019
Millie Irene Sterling

What seems to represent progress is often simply the same ignorance expressed in another jargon, or, more properly, another paradigmatic language (this being the fate of most human knowledge, of course).

–P. Hancock, T. Oran-Gilad, and J. Szalma

Abstract

The aviation industry is investigating reduced crew flight deck operations as a solution to an impending pilot shortage and to reduce costs. A shift in operation requires a shift in interaction. This new interaction should be flexible, natural, and minimize task interference when managing multiple concurrent tasks. Gaze, voice, and gesture interaction techniques can provide such interaction.

As an extension to existing research, this dissertation investigates how these interaction techniques, independent of input device technology or task design, affect an operator's ability to manage concurrent tasks. To test this, an interaction framework was developed, consisting of a model of human information processing and a device-independent interaction taxonomy.

Ten human factors experiments were conducted to validate the framework and answer the research question. The experiment tasks were abstracted from flight deck tasks. The effects of the six investigated interaction techniques on the costs of concurrent task management were measured via interruption and resumption time, performance of the first and second task, and subjective workload. Interaction technique was found to have a significant effect on resumption and interruption time, but not on performance or subjective workload. Multimodal interaction, while providing flexibility, results in less effective concurrent task management when compared with unimodal interaction. It was also found that neither gender nor previous pilot experience had an effect on the operator's ability to manage concurrent task demands.

The results of this work were then applied to the construction of a mobile flight deck simulator to demonstrate the author's vision for reduced crew flight deck operations with multimodal interaction. The simulator depicts six realistic scenarios of a reduced crew flight with investigated interaction techniques. Practical demonstrations have shown that the interaction is flexible, robust,

and natural. The device-independent, (multi)modal interaction framework will provide a baseline for effective concurrent task management workflows using gaze, voice, and gesture interaction techniques towards the implementation of reduced crew operations.

Zusammenfassung

Als eine mögliche Lösung für einen bevorstehenden Pilotenmangel und um Kosten zu senken untersucht die Luftfahrtindustrie neue Arbeitsprozesse im Cockpit mit reduzierter Besatzung. Die Änderung dieser Prozesse erfordert eine veränderte Interaktion. Diese neue Interaktion sollte flexibel, robust und natürlich sein und Aufgabeninterferenzen bei der Verwaltung mehrerer gleichzeitiger Aufgaben minimieren. Blick-, Stimm- und Gesteninteraktionsmethoden sind Techniken, die solche Interaktion ermöglichen können.

Diese Dissertation stellt eine Erweiterung existierender Forschung dar. Sie untersucht, wie diese Interaktionsmethoden, unabhängig von der Eingabegerättechnologie oder dem Aufgabendesign, die Fähigkeit eines Bedieners beeinflussen, Aufgaben gleichzeitig zu verwalten. Um dies zu testen, wurde ein Interaktionsmodell entwickelt, das aus einem Modell der menschlichen Informationsverarbeitung und einer geräteunabhängigen Interaktionstaxonomie besteht.

Um das Interaktionsmodell zu validieren und die Forschungsfrage zu beantworten, wurden zehn Versuche mit Probanden durchgeführt. Die Versuchsaufgaben wurden von zukünftigen Arbeitsaufgaben im Cockpit abstrahiert. Die Auswirkungen von sechs untersuchten Methoden auf das gleichzeitige Aufgabenmanagement wurden mittels der Unterbrechungs- und Wiederaufnahmezeiten, der Leistung in der ersten und zweiten Aufgabe und der subjektiven Arbeitsbelastung gemessen. Hierdurch wurde festgestellt, dass die Interaktionsmethode einen signifikanten Einfluss auf die Unterbrechungs- und Wiederaufnahmezeiten hat, aber nicht auf die Leistung oder die subjektive Arbeitsbelastung. Multimodale Interaktion bietet zwar Flexibilität, führt aber im Vergleich zu unimodaler Interaktion zu einem weniger effektiven gleichzeitigen Aufgabenmanagement. Es wurde auch festgestellt, dass weder Geschlecht noch Pilotenerfahrung Einfluss auf die Fähigkeit eines Bedieners hatten, gleichzeitige Aufgabenanforderungen

zu verwalten.

Mit den Ergebnissen dieser Arbeit wurde ein Cockpitprototyp gebaut, um die Vision der Autorin für zukünftige betriebliche Konzepte im Cockpit mit multimodaler Interaktion zu demonstrieren. Der Prototyp zeigt sechs realistische Szenarien für ein Cockpit mit reduzierter Besatzung mittels der untersuchten Interaktionsmethoden. Praktische Demonstrationen haben gezeigt, dass die Interaktion flexibel, robust und natürlich ist. Das geräteunabhängige, (multi)modale Interaktionsmodell wird zukünftig die Grundlage bilden, um effektive Arbeitsabläufe mit gleichzeitiger Aufgabenverwaltung mittels Blick-, Stimm- und Gesteninteraktion im Cockpit mit reduzierter Besatzung zu implementieren.

Contents

Nomenclature

Acronyms

ACROSS Advanced Cockpit for Reduction of Stress and Workload

ALIAS Aircrew Labor In-Cockpit Automation System

ALICIA All Condition Operations and Innovative Cockpit Infrastructure

ANCS aviate, navigate, communicate, manage systems

ANOVA analysis of variance

AOC airline operations center

API application programming interface

AR augmented reality

ASR automatic speech recognition

ATC air traffic control

CAD computer-aided design

CARE complementarity, assignment, redundancy, equivalence

CCD cursor control device

CDU control display unit

CE consumed endurance

CLT cognitive load theory

COTS commercial off the shelf

CRT cathode ray tube

CTM concurrent task management

DARPA Defense Advanced Research Projects Agency

DLR Deutsches Zentrum für Luft- und Raumfahrt, eng. German Aerospace Center

DOE	design of experiments
E	unimodal eye tracking/gaze
EEG	electroencephalography
EFB	electronic flight bag
EG	multimodal eye tracking and gesture
EICAS	engine indicating and crew alerting system
EMG	electromyography
FAA	Federal Aviation Administration
FFD	future flight deck
FMS	flight management system
FOV	field of view
FSR	Flight Systems and Automatic Control
G	unimodal gesture
GKS	Graphical Kernel System
GUI	graphical user interface
GV	multimodal gesture and voice
HMI	human-machine interface
HUD	head-up display
ICAO	International Civil Aviation Organization
IoT	internet of things
IR	infrared
ISO	International Organization for Standardization
LED	light emitting diode
LIDAR	light detection and ranging
LTM	long-term memory
LuFo	ger. Luftfahrtforschung, eng. aviation research
MCP	mode control panel
MM	mission manager
MMI	multimodal interface
MRT	Multiple Resource Theory
NASA	National Aeronautics and Space Administration
ND	navigation display
PFD	primary flight display
P/V	phonological/verbal
RAM	random access memory

RCO	reduced crew operations
RFID	radio frequency identification
ROI	region of interest
SCR	stimulus/code/response
S/M	visuospatial/manual
SPO	single pilot operations
STM	short-term memory
TAF	terminal aerodrome forecast
TCAS	traffic collision avoidance system
TLX	task load index
TUDA	Technische Universität Darmstadt
V	unimodal voice
VE	multimodal voice and eye tracking/gaze
WIMP	windows, icons, menus, and pointers
WM	working memory

1. Introduction

This introductory chapter begins with the motivation for developing an interaction framework for a reduced crew operations (RCO) flight deck. An improved human-machine interface (HMI) is one of many technological, legal, and sociological enablers for RCO. The chapter then describes RCO research to date and how the current study aligns with the existing research field. The objective of this research is to determine how flexible and natural interaction techniques affect the human's ability to communicate with increasingly automated systems, especially in situations of concurrent task management (CTM). The approach taken was to validate the developed interaction framework with human factors experiments and then apply it to an RCO vision demonstrator. The dissertation structure is provided at the end of this chapter to aid the reader in navigating this document.

1.1. Motivation

Boeing estimates that 635,000 new airline pilots will be needed to fly commercial aircraft globally in the next two decades [The19]. Given that there are only 290,000 active pilots today [CAE17], the industry either needs to produce more cockpit crew or find a way to reduce the number of crew required [Par17]. One of the ways the industry is seeking to reduce the number of required crew, while taking advantage of a significant potential profit opportunity[1], is investigating a new type of flight deck operation: one feasible for a single pilot or reduced crew [Har07; Joh+12; Bla+14; Bai+17; Cas+17; Ask+17; Boy14].

[1] A recent global research study identified a profit opportunity of over $15bn. for commercial and cargo operations to move to RCO, and an over $35bn. opportunity for the entire aviation industry related to pilotless aircraft [Cas+17].

Starting in the 1950s, commercial flight decks began gradually reducing the number of required crew, from five, to three, to two [Joh+12]. This historical de-crewing trend leads many in the industry to believe that some form of RCO or single pilot operations (SPO) is inevitable [Ask+17]. Executives from major aircraft manufacturers and airlines have conceded that the technology for single-pilot planes is available today; the main obstacles are perception and regulation [Cas+17]. The human element, however, remains a topic of heated debate [Bla+14; Bai+17]. Specifically, maintaining situation awareness, ensuring a manageable workload, task reallocation between human and autonomy (including cross-checking checklists), and monitoring pilot performance and health have been identified as the main challenges to flight deck operation by a reduced crew [Joh+12; Fab13; Bai+17].

As autonomy becomes increasingly prevalent in modern flight decks, new technologies need to be designed with effective human-autonomy teaming in mind [Bai+17; SNS18]. A 2017 National Aeronautics and Space Administration (NASA) study found that simply replacing the second pilot with autonomy, without redefining the function and responsibility of the remaining pilot, is not possible with current flight deck design and airspace operations [Bai+17]. The complexity of modern flight decks and the desire to find better (natural, intuitive, flexible, and error-tolerant) pilot-aircraft interfaces for RCO drives research on alternative control devices [MS00; Hol+17]. Direct and human-like interaction can simplify operations on future flight decks [RCO10; Hol+17], while interface flexibility is necessary for effective human-autonomy teaming in increasingly autonomous systems [Bai+17].

1.2. Reduced Crew Operations

Executives from aircraft manufacturers [Sha17] as well as aircraft operators [The10] have been quoted acknowledging that reduced crew operations are an inevitable next step in flight deck evolution [Ask+17] in addition to a possible solution for the coming pilot shortage and a promising profit opportunity [Har07; Wil+13; Gra+14; CSC16; Bai+17; Cas+17]. European research projects have been ramping up their focus on RCO and SPO research, as evidenced by the All Condition Operations and Innovative Cockpit Infrastructure (ALICIA) project [Sev14], the Advanced Cockpit for Reduction of Stress and Workload (ACROSS) project [Sev16b], and the 2020 ger. Luftfahrtforschung, eng. aviation research (LuFo) VI call for proposals, which specifically names SPO as a targeted research topic [Bf18]. In the United States, Section 744 of the 2018 Federal Aviation Administration (FAA) Reauthorization Act calls for research and development activity in support of single-pilot cargo aircraft [11518]. NASA is looking at distributed crew operations, with an airborne pilot and a ground operator supporting multiple aircraft with the ability to serve as a dedicated ground-based first officer in off-nominal events [Bai+17; Lac+14]. The Defense Advanced Research Projects Agency (DARPA), together with Aurora Flight Sciences, developed the Aircrew Labor In-Cockpit Automation System (ALIAS), a tailorable, drop-in robotic arm that can automate certain pilot workflows in existing aircraft [Def16].

The DARPA method of RCO replaces today's human operator with a robot [Def16]. NASA's research redefines the role of the pilot by reallocating the tasks of the pilot, airline operations center (AOC), and automation [Bai+17; Lac+14]. ACROSS investigated specific scenarios of SPO in today's aircraft in off-nominal conditions [Sev16b]. ALICIA investigated disruptive technologies in support of any aircraft operation that would lead to greater overall efficiency [Sev14]. The current study aligns most closely with the NASA approach and seeks to support an operational scenario that is manageable by a single person.

The author makes a contribution to the field of RCO research by exploring the potential of more human-like interaction methods to create a natural and flexible interface. It assumes that the operation, while being different than flight deck operation today, retains situations of CTM. In this RCO scenario, the traditional aviate, navigate, and communicate responsibilities of a pilot today will largely be taken over by automation, and the job title and description of a future pilot will evolve to be that of a systems administrator, or mission manager. In this redefined role, a more natural, intuitive, and flexible interface may aid an operator in managing highly autonomous systems. The developed device-independent, (multi)modal interaction framework provides a baseline for creating effective CTM workflows using gaze, voice and gesture interaction techniques. It is assumed that at least one technological solution for each interaction technique will achieve a state of maturity suitable for commercial aviation adoption (if this is not already the case) by the time of widespread RCO adoption.

1.3. Research Goals and Approach

Aviation personnel are being required to interact and team with increasingly autonomous and complex systems [Shi+16; Ho+17]. The goal of this research is to determine how flexible and natural interaction techniques affect the human's ability to communicate with increasingly automated systems, especially in situations of CTM. Knowing feasible applications of these interaction techniques prepares the aviation industry for the changes they could bring to information and service provision. The interaction framework proposed herein is one building block of a larger ongoing research effort at the Technische Universität Darmstadt (TUDA), which seeks to provide aircraft manufacturers and airlines with feasible concepts towards RCO implementation [Ins18].

The current study assesses the effects of gaze, voice, and gesture interaction (and combinations thereof) on the ability of humans to perform in situations of CTM, independent of input device technology. These techniques were chosen specifically because they represent human-like communication while being flexible and robust [Ber08; RCO10; DK15; Qva+17]. First, each technique (and combination thereof) was evaluated by means of extensive literature review. Initial laboratory experiments, both uni-modally (see [Mor15; Dee16]) and multi-modally (see [Cha17]), were conducted to identify feasible applications of each technique. The literature review and preliminary experiments were then used to develop a (multi)modal interaction framework, consisting of a model of human information processing, and a device-independent interaction taxonomy, both tailored for an RCO flight deck. A human-centric, rather than user-centric, design approach was chosen to develop the interaction framework.

The interaction framework was validated with human factor experiments, which tested the interaction techniques' effects on a human's ability to manage multiple concurrent tasks. The framework and experiment results were then applied to the construction of a mobile prototype which is used to demonstrate probable multimodal interaction on an RCO flight deck.

1.4. Document Structure

This dissertation is structured into six chapters and three appendices.

Chapter 2 begins with theories of human information processing in aviation. It then provides a brief history of human-machine interaction and its anticipated evolution into multimodal interaction. Gaze, voice, and gesture as device-independent interaction techniques are then reviewed. Existing interaction taxonomies are presented. CTM, and its contribution to the operational complexity of real-world flight deck operations, is described. At the end of the chapter, the research gap that is addressed by this dissertation is presented.

Chapter 3 defines the role of the pilot in today's two-crew complement and then describes the high-level differences of a mission manager under RCO. The interaction framework, consisting of a model of human information processing, and a device-independent interaction taxonomy, is then presented.

Chapter 4 describes a series of 10 human factors experiments that were conducted to test the developed interaction framework.

Chapter 5 applies the results of the framework and human factors experiments to a mobile prototype that is used to demonstrate realistic use cases for multimodal interaction on an RCO flight deck and its implications for the future of flight deck information solutions.

Chapter 6 describes the lessons learned from the current study, and provides an outlook for its significance towards RCO implementation.

Appendix A provides supplementary information for a more detailed understanding of the material presented in Chapter 2.

Appendix B contains detailed information on the experimental setup and methods described in Chapter 4 for the sake of reproducibility.

Appendix C presents additional considerations towards the construction of an RCO prototype, described in Chapter 5, including the dimensions of the final demonstrator.

2. State of the Art in Human Information Processing, Human-Machine Interaction, and Concurrent Task Management

This chapter provides the background knowledge required to understand the interaction framework proposed by this dissertation. The chapter begins with theories of human information processing in aviation. It then provides a brief history of human-machine interaction and its anticipated evolution into multimodal interaction. Gaze, voice, and gesture as device-independent interaction techniques are then reviewed in Section 2.4, Section 2.5, and Section 2.6. Existing interaction taxonomies are presented. Concurrent task management (CTM), and its contribution to the operational complexity of real-world flight deck operations, is described. At the end of the chapter, the research gap that is addressed by this dissertation is presented.

2.1. Attention Theories in Aviation

Flight deck information management issues suggest that researchers need to better understand how pilots use and process information by studying the cognitive mechanisms routinely used by flight crews [JR95; Vid+10]. Attention ties together the many components of human cognition and information processing [KWK07, chap. 17]. Since the early 1950s, researchers have been conducting multidisciplinary and cross-industry research in an effort to apply this human resource effectively in the design of human-machine systems [Bro54; Kah73; WSV83; HSO05; LBC18].

Many terms are related to attention, but are not quite the same in meaning: e.g.

situation awareness, mental or cognitive workload, vigilance, fatigue, drowsiness, alertness, activation, distraction [KWK07, chap. 2]. The following sections describe the most common theories used to evaluate attention, namely working memory, cognitive load, and limitations on mental resources [Vid+10; Rui11]. Attention theory is an inexact science, however, and different nomenclature is used to describe elements of similar concepts. An agreed upon definition of "attention" and "workload" does not exist, and as such, no agreed upon method of measuring or modelling it [HM79; KWK07].

2.1.1. Theory of Working Memory

Working memory (WM) stems from the concept of short-term memory (STM) and refers to the the temporary storage of information and its manipulation[1] [Bad12]. It states that a set of cognitive processes keep a limited amount of items available for immediate use in STM so as to forego the difficult task of recalling information from long-term memory (LTM) [Bad83; Bad12]. WM research began with Broadbent's filter theory, which described a human's information processing capability as single-channel, selective attention [KWK07, chap. 1]. In his seminal split-span experiment, Broadbent showed that subjects would filter and process information according to which ear the information was presented [Bro54]. In 1956, Miller suggested that STM can only retain seven pieces of information, plus or minus two [Mil56]. He claims that we can increase the accuracy of our judgements by chaining or chunking related pieces of information together, or by differing the dimensions along which the stimuli occur [Mil56].

In 1974, Baddeley & Hitch developed a model which describes WM as a combination of two temporary storage systems (visual-spatial and phonological) which are coordinated by an executive component (central executive) [Bad83]. The visual-spatial sketchpad interprets visual stimuli, such as pictures and diagrams, and the phonological loop interprets auditory or verbal stimuli [Ovi06]. In 1977, Anderson, Spiro, and Anderson showed strong empirical evidence of WM benefitting from schema in LTM [ASA78]. The authors described schema as mental structures that summarize information abstractly, and are generic

[1] If the human brain was a computer, WM would be analogous to random access memory (RAM).

characterizations of things and events [ASA78]. Baddeley developed a modified model, shown in Figure 2.1, which adds schema accessing LTM and an episodic buffer to integrate the components of STM [Bad00]. In addition to Miller's chunking and stimuli differentiation, the ability to call upon schemas located in LTM is a way to increase the accuracy and capacity of WM [PV94; AK07].

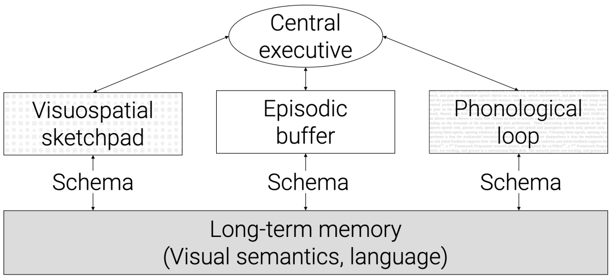


Figure 2.1.: Author rendition of Baddeley's revised model of WM from [Bad00], describing two separate systems for the storage and manipulation of information in STM [Bad12] and including links to LTM

Due to its complexity, Baddeley & Hitch focused on perception and described the central executive control of attention as a homunculus [Bad83]. Norman and Shallice proposed that attention is controlled in two different ways: (1) habitually, which relies more on LTM schema to complete a task and, (2) supervisory, which responds to abnormal situations that deviate from learned processes[NS80]. An example of the former is taxiing to a familiar runway. An example of the latter would be taxiing to the runway in extremely low visibility and heavy traffic. In places where no schema exist (e.g. learning something new or a never before seen situation) WM becomes a scarce resource and performance slows [Rui11].

Goodstein and Rasmussen's Step Ladder Model [GR85] describes Norman and Shallice's habitual and supervisory methods to control attention as a continuum between conscious and automatic behavior [Emb05], as depicted in Figure 2.2. With each increase in level, the demand for mental resources increases [GR85;

Emb05]. The more experience a human has, the less dependent the human is on resource-heavy levels of information processing [Vid+10]. Experience can therefore be seen as the creation of LTM schemata which enable the user to move from demanding, knowledge-based processes to automated, rule- or even skill-based processes [GR85; Bad00; Emb05; Vid+10].

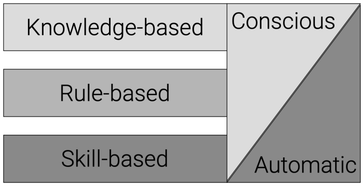


Figure 2.2.: Author rendition of the levels of information processing and demand for mental resources on a continuum between conscious and automatic behavior, based on [GR85] and [Emb05]

Subsequent variations of the model have added other human senses (haptics, smell, and taste) to the processing methods, but the theory's foundation argues that at least two slave systems (visual and auditory) are employed by a human's working memory, with the central executive coordinating the external stimuli with schemas, when available [RCO10; Bad12].

2.1.2. Cognitive Load Theory

Cognitive Load Theory (CLT) describes how LTM schemata are built for later retrieval from working memory[2] [Swe94; Lep+15]. It focuses on the coordination and interaction of the WM perception modules between a limited short-term memory [Mil56; PV94] to a comparatively unlimited long-term memory [Paa+03].

CLT, visualized in Figure 2.3, states that WM is made of three types of cognitive load. The first, intrinsic load, is the load inherent to a task and cannot easily be

[2]Revisiting the computer analogy in Section 2.1.1, if WM is RAM, CLT is analogous to how efficiently information is written to a storage device and how quickly it can be accessed from storage.

changed [Paa+03; Qva+17]. Extraneous and germane load, on the other hand, can be manipulated by the system designer [Swe94]. Extraneous load, often referred to as ineffective load, refers to the mental workload due to the poor design of a task, but not inherent to the task itself [RCO10; Qva+17]. Especially in the flight deck, inappropriate ways of presenting information can quickly lead to cognitive overload [LV15]. Germane load is the mental effort a learner experiences while converting new information into LTM schema [RCO10]. The quality of the schema that are built from germane load help to automate similar tasks in the future [Swe94; PRS04].

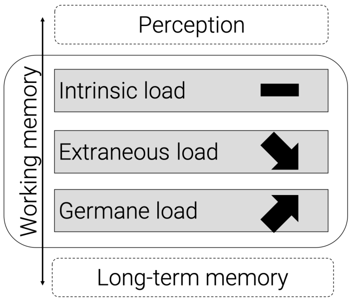


Figure 2.3.: Author rendition of the three components of Sweller's CLT from [Swe94] and the recommendation to decrease extraneous load in task design and increase germane load (when learning), while intrinsic load is inherent to the task itself and cannot be influenced

CLT, specifically through increasing germane load, is primarily used in creating interfaces which promote learning [Swe88]. But the theory is relevant in designing operational interfaces which should mitigate intrinsic load while reducing extraneous load [Rui11]. Flight deck tasks can have inherently high intrinsic load, requiring experience, effective communication, and an ability to deal with stressful and high workload situations [NAS10; Bla+14]. Natural, flexible, and intuitive interfaces can help reduce extraneous load when encountering more complex tasks, bypassing the requirement for higher level processing (see

Figure 2.2), which is demanding of mental resources [Emb05; Vid+10; LV15].

2.1.3. Multiple Resource Theory

Wickens's Multiple Resource Theory (MRT) defines the limited resources that a human has at his or her disposal for processing and reacting to external stimuli [Wic08]. In their seminal 1983 study, Wickens, Sandry, and Vidulich proposed the makings of a model which described the compatibility and overlap of differing mental resources (those available to WM) in information processing and task performance [WSV83]. Wickens's Multiple Resource Theory is one of the most widely used theories to support multiple task performance research [Wic02] and, according to Kramer, Wiegmann, and Kirlik, is one of the best design heuristics for human-machine interaction [KWK07, chap. 4].

The original model, depicted in Figure 2.4, is expressed in a cube which represents the three dimensions of information processing. The author has replaced the "visual" modality with "ocular" and the "verbal" code with "phonological" so as to avoid confusion when referring to them in shorthand in Chapter 4. The dimensions, as described in [Wic08], are described below.

Stages provide the temporal axis of the resource model. The left and middle stages represent the perceptual and cognitive resources, and the right stage represents the resources for selection and execution of action [Wic08].

Modalities depict the human's available perception resources, vision and hearing being the two major channels through which humans perceive sensory information [Wic08]. The modality dimension only spans the *perception* stage.

Responses depict the human's available resources to execute a response, manual or vocal being the two major channels [Wic08]. The response dimension only spans the *response* stage.

Codes separate the human's available resources for spatial and linguistic activity (the slave systems of WM). Codes span all three stages.

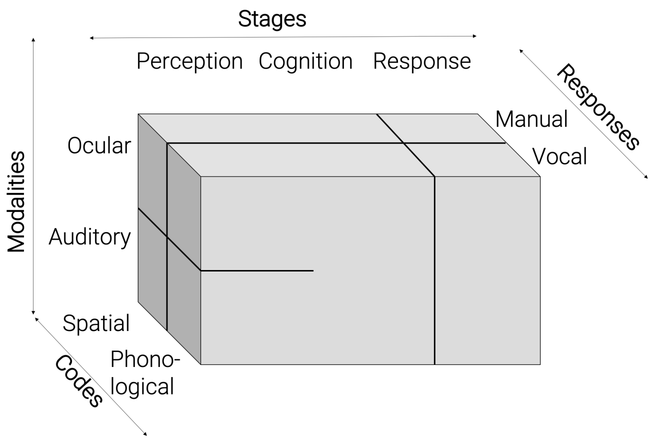


Figure 2.4.: Author rendition of the 3-D “box model” of Wickens’s Multiple Resource Theory according to [Wic08], depicting the mental resources a human has at his/her disposal for processing and reacting to external stimuli

The resources for which concurrent tasks compete are across each processing stage, as well as within each: perception (human senses), cognition (WM across the slave systems and coordination with schemas in LTM), and output modalities (manual and vocal) [WSV83]. Wickens, Sandry, and Vidulich's model hypothesizes that tasks that are code incompatible, from perception to cognition to response, will have greater resource competition. A stimulus perceived and processed spatially is best responded to manually; similarly, stimuli presented and processed verbally causes the least mental resource interference when responded to vocally [HJ74; Wic02; Cha17]. A task requiring the judgement or orientation of the three axes of translation or orientation is spatially processed [WSV83]. Speech and linguistic tasks are processed verbally [Wat72; KRC07]. Wickens, Sandry, and Vidulich's box model conforms with Sweller's CLT [Swe94] in that it can also be used to explain resource competition with increasing task complexity: the more complicated the task (either through extraneous, germane, or intrinsic load), the greater the demands on the cognition stage, leaving fewer resources for perception and response [WSV83; Cha17].

The original "box model" has proven parallels to brain anatomy [Wic08], and its ease of explanation allows human factors professionals to apply the theory in designing workflows in complex environments [KWK07]. Expansions to the model allow professionals to tailor it to their specific use cases [HW03; KWK07; Wic08]. Horrey and Wickens broke down the visual modality for perception into *focal* and *ambient* vision [HW03]. They postulated that focal vision supports granular perception, such as involved in reading this dissertation, while ambient vision supports peripheral perception, such as that used by pilots to monitor the flight deck [HW03].

Hancock, Gilad, and Szalma claim that the phonological code is just an abstraction of temporary spatial codes, and that this axis can be abstracted further into an information code, which is a combination of the individual, his/her experience, his/her training, and the environment s/he is in [KWK07, chap. 4]. Each code is represented as an abstraction of the code before it. Phonological information is a more efficient representation of visuospatial information, dictated by a social consensus on a collection of symbols [KWK07], such as the Roman alphabet or Chinese characters. The information code is a more efficient representation of phonological or visuospatial information, such as air traffic control (ATC) chatter or aeronautical charts.

Similar to the expansion of the code axis, Hancock, Gilad, and Szalma also expanded the modality axis, depicted in Figure 2.5, to include other human senses, and adjusted the boxes to reflect the dominance of each [KWK07, chap. 4]. The kinesthetic sensory system was added by the author to include all sensory systems relevant in a flight deck setting.

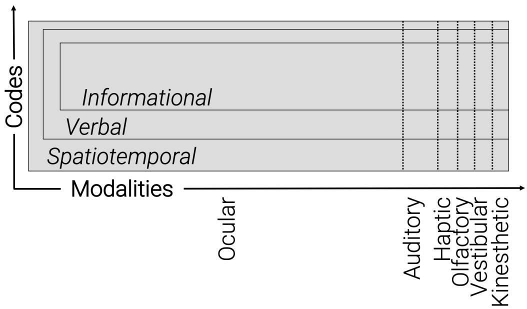


Figure 2.5.: Author rendition of the modality axis and code abstraction of Wickens's MRT, proposed by [KWK07], with minor author adaptations

2.1.4. Attention and Mental Effort

Mental resource and information processing theories are being revisited in the fields of human factors and ergonomics in part due to advances in technology that have provided objective metrics for measuring cognitive workload [Paa+03]. Four main methods are summarized below.

Indirect performance measures, through accuracy and efficiency of a primary task, hypothesize that as an operator's workload increases, performance of the primary task will decrease [HM79]. Objective measures of the primary task can include error rate, time on task, response time, comparison of the target versus actual state, and anomaly detection rate [Paa+03]. Performance measures must be defined uniquely in terms of the primary task, and multiple measures are often combined or normalized to obtain an aggregate measure of workload [Swe88; Kar06; AK07].

Spare mental capacity assumes that as mental resources are expended, an operator's workload will increase until the point of overload [HM79]. Secondary tasks are the most common method to create competition in mental resources [HM79], and are often used in testing MRT to understand how mental resources interact with one another [WH02; Wic02]. Either the primary or secondary task is prioritized, and the performance of the other task can be used to quantify mental effort [DT86; WHX02; HM79].

Subjective self-report measures require the subject, at various intervals during or after an experiment, to assess the mental effort required to perform a task [HM79]. The most commonly implemented method for subjective, self-reported measures is the multidimensional National Aeronautics and Space Administration (NASA) task load index (TLX) questionnaire [Mil01]. The TLX score is a weighted average of five, twenty-one point scales of workload characteristics (mental demand, physical demand, temporal demand, effort, and frustration) and one, twenty-one point scale on performance success [HS88]. The interpretation of subjective, self-report measures is highly dependent on the user and therefore inconsistent [Dub12].

Direct physiological measures, such as heart rate variability [PV94], speech features [KRC07], electroencephalography (EEG) [Ant+10], volume and concentration of carbon dioxide in respiration [Mil01], hydration levels [Dav+08a], and pupillary response [Van+04], have all been used to measure workload. The data collection is easily contaminated by movement and variables other than mental erffort [Ovi06].

No agreed upon method for measuring mental effort exists [HM79; KWK07; Rui11; Lep+15].

2.2. A History of Human-Machine Interaction

WIMP (windows, icons, menus, and pointers) graphical user interfaces (GUIs) were introduced in 1970 by Xerox Parc, commercialized by Apple over the decades, and are the standard interaction framework for desktop computers [Sol12]. Modern aircraft flight deck interaction, such as Boeing's 787 or Airbus's A350, is generally a combination of rotary knobs, push buttons, cursor control devices (CCDs), and WIMP GUIs on multiple configurable displays [Cha17; Hol+17]. WIMP has been optimized for the pairing of a mouse (proxy of user's attention) and keyboard (data entry) [Hin02; Ber08]. Direct touch interfaces do away with the pointer abstraction, and have expanded WIMP interfaces with i.a. buttons, sliders, and checkboxes [Hin02].

WIMP interfaces are learned interfaces, however, and will not be adequate for human-machine interaction in the future [SS01; Rui11; Gar13; MB14]. WIMP interfaces become cumbersome with increasing application complexity [TR00] and task information processing requirements [Vid+10]. Too much time is spent manipulating the interface rather than performing the task objectives [Sol12]. A drive to develop more flexible and natural human-machine interaction and improvements in computing and processing power allows a redefinition of interaction philosophy [Sol12; Hol+17]. This redefinition requires new interfaces and interaction techniques [Hin02]. Multimodal interaction is lauded as one of the most promising post-WIMP methods to process natural human communication [OC00; Ber08; DLO09].

2.3. Multimodal Interaction

In human-to-human interaction, people speak, gesture, shift eye gaze, and communicate in a way that bears little or no resemblance to the serial keyboard and mouse clicks with a WIMP GUI [OC00; Hol+17]. Multimodal interaction can provide a contextually suitable combination of control modalities that allow the human to interact with a system more like they would with another human [OC00; Ber08; DLO09]. Multimodal interaction is ambiguous, so this section will describe what exactly is meant by the term for the purposes of this dissertation.

2.3.1. Defining (Multi)modal Interaction

Interaction can occur between humans, imperceptible to a machine, and between machines, imperceptible to a human [Ber08; Sol12; Qva+17]. Bernsen argues that successful human-machine interaction is really the exchange of information between the two entities, and consists of three parts [Ber08], which are described below.

A physical medium is required to carry information, such as sound waves, light, or mechanical sensors [Ber08].

A human sense, such as sight, hearing, touch, smell, taste, or the kinesthetic and vestibular system, is needed to perceive the information, and each sensory system has its own limitations (e.g. a human cannot see infrared or ultraviolet light, or hear a sound above 20.000 Hz) [Ber08].

An information representation is derived from the chosen medium and human sense [Ber08]. Text, images, facial expression, or a gesture all employ light as the medium and vision as the human sense [Ber08].

Bernsen's definition of interaction describes both input to a machine, and output from it [Ber08]. *Input* is any information about a physical environment that a machine can sense (e.g. movement of a mouse, stroke of a key on a keyboard, changes of light across a light sensor, sound waves into a microphone); *output* is any modification to that physical environment that a machine can control (e.g. light from a display, audio from speakers, pulses from vibration actuators) [HJW14]. An *interaction technique* is any exemplification of interaction, consisting of a combination of all input and output hardware and software which allows a user to accomplish a low-level task [Ber08; HJW14]. Every interaction technique is best for something and worst for something else [Bux90].

Bernsen defines a modality as the third criteria of interaction: a way of representing information in a particular medium and employed human sense [Ber08]. The benefits of multimodality come from complementing modality synergy and equivalence [Bol80; Sal+95; Sol12; Tur14]. To realize the benefits of multimodality, an interface should use multiple input modalities to communicate the intended command [Qva+17; Sol12]. Synergistic interaction provides those modalities in parallel, such that input is integrated and processed continuously [Ber08; Tur14]. Equivalent interfaces provide multiple modalities to communicate the same intended command [Sal+95; Sol12]. Such interfaces provide more natural interaction because they are nearer to multimodal human-human communication [Qva+17] and capitalize on the strengths of each modality to overcome weaknesses of others, while adapting to user preference and choice [Sol12; Qva+17].

If interaction is a single (unimodal) interaction technique consisting of both input and output modalities, and multimodality combines modalities, *multimodal interaction* can therefore be defined as human-machine communication systems that employ two or more interaction techniques, and whose input and output modalities are synergistic, complementary, provide equivalence, and consider the limitations of both the machine and the user.

2.3.2. Engineering and Designing for Multimodal Interaction

While the term “multimodal” is used to describe vastly different types of interaction, the benefits are almost always cited as reduced cognitive workload, from naturalness and flexibility, and increased operator efficiency, from robustness and error reduction [OC00; KA12; Cha17]. These advantages are summarized below.

Naturalness Communication between humans is naturally multimodal, involving speech, gesture, facial expression, and body posture [Sol12]. Multimodal interaction provides better symbiosis with the human operator’s sensory system and limitations [Ber08]. Advances in micro-processors and storage capacity open up potential of the machine to interpret how and what the human is communicating naturally [KST98]; sometimes even without needing a defined command input set [Que+02].

Flexibility Multimodal interfaces allow users to choose what modalities they want to use according to task demands and user preference [NMF16], increasing the level of input expressivity [RCO10]. Users tend to mix unimodal and multimodal interaction as they see fit [Ovi99].

Robustness When a human communicates similar or related information through different modalities, it increases the likelihood that the computer will recognize the human's intent [KST98]. Redundancy in multimodal input results in better quality human-machine interface (HMI), especially when modalities are paired such that one overcomes the deficiencies of the other(s) [Qva+17].

Minimizing errors Multimodal interfaces (MMIs) have been shown to increase performance by lowering the number of errors and providing faster error-correction [RCO10]. They can also reduce the potential for errors of omission (e.g. due to distraction) [DK15], which is one of the most common mistakes in high workload situations [LDB09].

Multimodal interaction isn't always as efficient as unimodal interfaces [DLO09]. Dostál and Kolčárek argue that the flexibility of a user to choose a certain (multi)modal method may lead to rigid use of one interaction technique, including situations where it may not be appropriate [DK15]. Complex operations are often highly standardized, and, especially in the flight deck, the lexicon is specific [DK15; Dee16]. More natural forms of communication, particularly non-standardized speech, may be seen as a disadvantage, as it could stray from the tested and approved operation [Dee16]. Other disadvantages associated with multimodal interaction have to do with the technologies themselves, such as the fusion and accurate interpretation of multiple input streams [DLO09] or computational processing power required for algorithms [Ber08].

All modalities, both input and output, differ in expressiveness, and each is suited for exchanging different kinds of information[3] [Ber08]. As Wickens, Sandry, and Vidulich state, interface requirements cannot be arbitrarily assigned input or output modalities to produce an effective system [WSV83]. To reduce extraneous cognitive load and increase the effectiveness of a chosen input modality, one must also consider suitable output modalities for each [WSV83; Ber08; RCO10].

[3] A blind person cannot use modalities requiring vision; an image is better at expressing how a shape looks than a textual description [Ber08].

As tasks increase in complexity (either through extraneous, germane, or intrinsic load), demand on the cognition stage of mental resources leaves fewer resources for perception and response [WSV83; Swe94; Cha17]. To design a multimodal system, one needs to understand the types of tasks an operator might encounter in that system, the mental resources the tasks will demand, and the competition and interference amongst resources that will result [WSV83; OC00; Ber08; RCO10]. Only by considering both human and machine information processing limitations can engineers design effective multimodal interaction [Tur14; STB09].

2.3.3. Common Interaction Modality Combinations

The following section lists of some of the most prominent interaction modality combinations in the literature and the benefits associated with each.

Speech and gesture Bolt's "Put-That-There" system, in which users can move objects on a wall display using hand gestures and a voice command, found that multimodal interaction provided a powerful and natural user interface [Bol80]. In 1989, researchers found that users preferred to perform graphics manipulation tasks with combined gesture and speech modalities over either modality used separately [Hau89]. Schapira and Sharma found that combining speech with gesture provides a natural interface that increases the ability of a user to perform complex operations [SS01]. Other studies investigate unintended gesture to provide complementary information to speech [Gol+01; Que+02].

Speech and gaze Hatfield and Jenkins found the use of gaze to be a powerful deixis to establish context or spatial coordinates, and to provide feedback on user verbal commands and selection on a mission planning interface [HJ96]. Merchant and Schnell combined speech and gaze on a Boeing 777-300 simulator to provide a more effective solution for the activation of controls and data entry in high workload situations (e.g. takeoffs and approaches) than voice alone [MS00]. Kang et al. built a query processing system that combined gaze with geo-location information to increased the number of answered queries by 65%, which would have otherwise gone unanswered or required looking up information on a separate display [Kan+15].

Speech with touch and other conventional modalities Dostál and Kolčárek found that speech, in combination with touch screens, CCDs, and a keyboard improved efficiency and performance in flight planning tasks on a navigational display [DK15]. Cohen et al. combined speech with a pen-based input device and found the interface to be $3.2-8.7$ times faster than the traditional GUIs for map-based tasks (e.g. object placement, highlighting, creating shapes) [Coh+98]. Castronovo et al. combined speech with a turn-and-push dial for secondary tasks while driving a vehicle, which resulted in significantly safer driving [Cas+10].

Speech, gesture, and gaze Koons, Sparrell, and Thorisson present a system for integrating gesture, speech, and gaze to manipulate spatial objects on a map [KST98]. Nesselrath, Moniri, and Feld proposed a multimodal interaction concept that allows vehicle drivers to make free choices on how to activate a function (e.g. activating turn signals, opening windows, folding the side-view mirrors) depending on the demands of the situation and user preferences to reduce driver distraction. A-PiMod (Applying Pilot Models for Safer Aircraft), a 7^{th} Framework Programme research project, investigated the use of speech, eye tracking, and gesture in a conventional flight deck to support interactions such as pointing, object selection, spatial manipulation, data entry, option setting, command input, command execution, communication, and navigation [Sev16a].

Multimodality is an expansive research field that still has a lot of unexplored novelty [Ber08]. Multimodal interfaces can create more intelligent user interfaces than their WIMP predecessors [OC00; RCO10], but only if designed according to the specific needs of the human-machine system [STB09; Tur14]. Individual interaction techniques, and their specific benefits, will now be presented.

2.4. Gaze as an Interaction Technique

The following subchapter will describe gaze as an interaction technique, including common application domains for gaze, and the advantages and disadvantages of the input modality. While the technological implementation is irrelevant for the goal of this study, a review of state of the art methods to implement gaze as an interaction technique is provided in Section A.3.1 for completeness.

2.4.1. Gaze Application Domains

Eye tracking, otherwise known as gaze tracking, technologies track and measure individual eye movements to determine what a subject is looking at and how their gaze wanders over time [PB05]. Eye movements for this purpose are defined as fixations (a moment where the eye is relatively stationary and perceiving information) and saccades (the quick eye movement in between fixations) [CY13]. These, and other metrics derived from these measurements, can be used to provide insight into a user's operation of a system, including user efficiency, regions of interest or frustration, and fatigue [Dee16]. Gaze recognition systems exist along a continuum of increasing information provision and control, from passive eye monitoring used primarily for post-hoc diagnostics, to explicit eye input for direct command and control of a system [Hol+17]. The first three categories along this continuum, represented in Figure 2.6, do not require the user to change their natural gaze behavior [MB14; Hol+17]. Attentive and explicit interfaces use gaze as an overt input modality [HJ96; MS00] and are the interaction techniques of relevance for this study. Most studies combine gaze tracking with other modalities to help differentiate a gaze used for perception of information and a gaze intended as command [HJ96; MS00; TBL15].

Multiple commercial off the shelf (COTS) software and hardware packages for gaze monitoring exist [Duc08; MB14]. In the automotive industry, the technology is being used to monitor driver activity, namely safety, fatigue, and missed alerts [CH11; MB14]. Caterpillar sells a monitoring eye tracking package that can be retrofitted into its vehicles to detect driver fatigue [Kel13]. Gaze tracking is being integrated into computers and monitors to become a standard computer interface [Tob11; Fuj12]. Lufthansa Systems uses gaze recognition

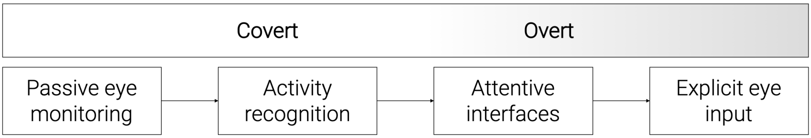


Figure 2.6.: Continuum of gaze input, in order of increasing information provision, adapted from [MB14]

together with its operations control software to determine whether operators in the airline operations center (AOC) have noticed important messages, or if they require additional alerting, such as sound [Luf16].

Specific to the flight deck, Calhoun, Arbak, and Boff used explicit gaze to activate switches and found it to be a feasible alternative to manual switches when handsfree operation was desired [CAB84]. Hatfield and Jenkins designed a system that allows pilots to gaze at a display and use voice commands and queries to operate that display handsfree [HJ96]. Merchant and Schnell developed a part task flight simulator to test the activation of various controls combined with automatic speech recognition (ASR) [MS00]. Thomas, Biswas, and Langdon used eye tracking as a CCD and found that it had significant potential with regards to performance, cognitive load, and interference with flight tasks in a wholly heads-up flight deck [TBL15].

2.4.2. Advantages and Disadvantages of Gaze

Gaze has been found to be much faster than CCDs at pointing and identifying a target if the target objects are large enough [SJ00]. Attentive and explicit gaze interfaces are cited as a natural and intuitive technique to interact with a system, as they show where a user's current attention is, which is usually a precursor to user action [MB14; TBL15; Hol+17]. Researchers assert that eye tracking, especially when combined with voice input, can reduce the pilot's peak workload and cut down the number of dedicated switches and control panels in the flight deck [HJ96; MS00].

Using gaze as an input modality presents a challenge because is used for both perception and response [Dee16; Gie18]. "Midas touch," or inadvertent activation, occurs when a system interprets unintended fixation on a region of interest as command activation [Hol+17]. The most common method to mitigate this risk is by combining it with other modalities, such as a button, hand gesture, or eye wink [HJ96; MS00; TBL15; Hol+17]. Inaccuracy of eye tracking is inherent to the modality because of the natural physiological limitations of eye movements [SJ00], but this is mitigated with the introduction of multiple input modalities and re-design of the information output modality required for the task [Hol+17].

2.5. Voice as an Interaction Technique

Voice as an interaction technique, including common application domains for voice, and the advantages and disadvantages of the input modality will be provided in the following subchapter. While the technological implementation is irrelevant for the goal of this study, a review of state of the art methods to implement voice control is provided in Section A.3.2 for completeness.

2.5.1. Voice Application Domains

Using language and speech to create a more natural interface between man and machine is a large research body [Eul06]. In 1952, Davis, Biddulph, and Balashek developed a system for recognizing telephone quality digits spoken at a normal rate by converting sound waves into electrical impulses [DBB52]. In the decades following, automatic speech recognition techniques went from laboratory to operational in a variety of industries, including telecommunications, office environments, disabled user applications, and transportation [NS96]. Speech input is integrated into everyday life, from Siri[4] to military aviation single-pilot cockpits [Bux+15].

In 1980, Mountford and North found voice input for data entry tasks while time sharing with a tracking task to improve operator performance over manual

[4] Siri is a voice-based virtual assistant for Apple phones, tablets, and computers [App18].

keyboard entry [MN80]. In 1996, Williamson, Barry, and Liggett implemented a vocabulary of 54 words in an OV-10A to design tasks that could be accomplished in a military aircraft, such as setting range on a display, going to a specific display page/layer, and changing display orientation [WBL96]. Voice control is implemented in the U.S. joint strike fighter aircraft, allowing pilots to change radio frequencies and adjust volume, and the Eurofighter Typhoon, to control displays, radar, radios, target designation, and navigation aids [Dav+08b]. Barón and Green found that most studies investigating voice interfaces resulted in better driving and task performance, and workload reduction, but the literature lacks firm conclusions and it is difficult to compare results across studies [BG06].

Wesson and Pearson tested voice control for direct aircraft system queries, flight management system (FMS) autopilot, communication system data entry, memo creation, and checklist assistance, and found a significant reduction in workload and high usability [WP06]. Serban and Houston found that voice activated technologies improve efficiency and minimize human factors contribution to critical decision making [SH12]. Buxbaum et al. began equipping their tower and ATC training simulators with voice recognition to reinforce a standard phraseology, and the system was found to significantly reduce workload for controllers [Bux+15].

Rockwell Collins's Advanced Technology Center features a commercial aircraft simulator that can be flown by voice [Pat15]. Honeywell's Flight Deck of the Future Lab is developing voice control and other interaction modalities for next generation cockpits [Cro14]. VoiceFlight, a speech recognition system for programming a Garmin global positioning system, was the first voice control system certified by the Federal Aviation Administration (FAA), but was discontinued in 2014 after an improved manual interface impacted its sales [Voi14]. Certification is the greatest hurdle for voice control in flight decks, so Wesson and Pearson suggest only using ASR in non-safety critical applications, such that any system would have, at most, a minor effect on flight safety should it fail[5] [WP06]. Despite evidence demonstrating the benefits of voice recognition in complex work environments, and tested applications in military aviation, the technology has yet to be successfully implemented in civil aviation [CB14].

[5] For definitions of failure conditions, the reader is referred to FAA Advisory Circular 25.1309-1A.

2.5.2. Advantages and Disadvantages of Voice

Natural language is widely regarded as a natural input and control method between man and machine, given that it is fast and intuitive [HJ96; WP06; SH12; DK15]. Bittner compares the advantages of voice recognition technology to those of automation, namely reduced operator workload and increased efficiency [Bit06]. Voice has been proven to benefit safety and efficiency [BG06; WP06; CB14; Bux+15], but lack of robustness is the primary reason for the the lack of uptake in modern day flight decks [WP06]. Wesson and Pearson claim voice interfaces reduce complexity and increase efficiency by simplifying menu structures in avionics displays, reducing frequent manual data entry, and providing direct access to functions across various flight deck interfaces, such as radio frequencies, route modifications, electronic logbook, and aircraft system status queries [WP06]. These applications have been found to reduce head down and command entry time, and allows the operators' hands to be engaged elsewhere [HJ96; WP06; CB14]. Wesson and Pearson also mention an increased rate of incorrect data entry via manual keyboards during high workload situations, which ASR technologies would prevent [WP06].

Speech commands are not well-suited for tasks that involve continuous input or require fine-grained tuning, e.g. "turn the radio up *a little bit* louder" [Cas+10]. Voice recognition systems are inherently dependent on the interpretation of language, which can be influenced by accents, background noise, speaker vocabulary, and grammar [Bit06; CB14]. Learning and memorizing a system's vocabulary requires training, and recall of appropriate commands may be impaired in situations of high workload [MS00]. Semantic language processing systems attempt to address the additional user workload by using natural language processing technologies that no longer adhere to strict dialogues [Que+02; NMF16]. Another method to increase robustness, accuracy, and usability of an ASR system is to synthesize it with additional contextual information [MS00]. Blackstun et al. require an accuracy of 99% or greater, even for non-safety critical functions, to ensure pilot acceptance [Bla+14]. The Deutsches Zentrum für Luft- und Raumfahrt, eng. German Aerospace Center (DLR) suggests a 95% recognition rate for ATC applications [Bux+15].

2.6. Gesture as an Interaction Technique

Gesture as an interaction technique, including what is meant by the term “gesture”, common application domains for gesture, and the advantages and disadvantages of the input modality will be provided in the following subchapter. While the technological implementation is irrelevant for the goal of this study, a review of state of the art methods to implement gesture control is provided in Section A.3.3 for completeness.

2.6.1. Gesture Application Domains

For the purposes of this research, a gesture is defined as any deliberate movement (or sequence of movements) of the arm(s) or hand(s) that conveys information, which begins and ends from a rest position, and includes static positions in between [Sch84; KH90]. By this definition, waving a hand to indicate moving forward or backward in a document or video player is a gesture, but the same gesture to shoo away a fly is not, as it lacks communicative intent. Only “freehand gestures” without the use of a screen or other physical medium whereupon the gestures are conducted are considered.

Gestures are a continuum of increasing information provision, as depicted in Figure 2.7 [McN06]. Moving from left to right along the continuum, speech becomes less of a necessity, and the correspondence to language increases [McN06]. Gesticulation unintentionally accompanies speech whereas sign language has complete lexical and grammatical specification [Que+02]. To increase intuitiveness and adhere to the definition above, gesture in this study is limited to language-like and pantomime gesture, henceforth referred to as manipulative gesture, and emblems, henceforth referred to as semaphoric gesture, per Quek et al.’s higher-level classification. Most COTS gesture recognition technologies are limited to these three types of gestures [Gar13; Nor18a; Lea18b; Mic18] and these gestures also represent a significant portion of natural human hand use [Que+02].

In desktop applications, gestures are typically used as an alternative to the mouse and keyboard [KS05]. Manipulative and semaphoric gesture has been proven to be effective for navigation and 3D camera manipulation in graphic

Figure 2.7.: Author rendition of Kendon's continuum of gesture from [McN06] and Quek et al.'s higher-level classification from [Que+02], in order of increasing information provision from left to right

applications, including translation, orbiting around a point, zooming in and out, and navigation around an object [ZF99]. Lenman, Bretzner, and Thuresson developed a hierarchal pie menu for manipulative and semaphoric gesture, as it is easier to move the hand without feedback in a given direction than to a menu item at a defined position [LBT02]. The media controls used with Leap Motion implement a similar command set in which a user points at the screen with a semaphore gesture to reveal a pie menu of common media controls (play/pause, previous, next, and volume), and the manipulated finger trajectory serves as the control selection [Lea18b].

Manipulative gesture interaction in computer supported cooperative work across large displays has also been proven to be highly effective [Kar06]. Wu and Balakrishnan developed a multi-user tabletop display for designing furniture layouts that leveraged the type of multi-finger and hand actions people perform when interacting with physical entities on a table [WB03]. Ou et al. developed a system in which a remote helper uses manipulative gesture to annotate over video feed from another user to enhance task performance over traditional video-only systems [Ou+03]. Chatty and Lecoanet's gesture system for air traffic controllers was found to be easy and intuitive because the fifteen unique gesture inputs were based off of a shorthand used amongst the controllers [CL96].

Ubiquitous computing was first coined by Weiser in 1991 [Wei91], but when discussing the physical devices themselves, the more common term is the internet of things (IoT) [FR11]. Gesture is a common mode of natural interaction with ubiquitous computing [Kar06]. Vogel and Balakrishnan demonstrated the benefits of combining manipulation and semaphores for tasks which are best performed from a distance, e.g. sorting through photos and presenting to a group while annotating [VB05]. Streitz et al. developed a collaborative workspace which used gestures to move and manipulate digital objects on a distant display wall [Str+99]. WiSee uses Wi-Fi signals in a home to interpret a set of nine semaphoric gestures that can be used to interact with smart devices [KTG14]. AllSee offers a similar gesture set, and can run on battery-free devices, such as power-harvesting sensors and radio frequency identification (RFID) tags, lending itself to interaction in the IoT, e.g. turning on smart appliances or opening a door [Pu+13]. Fails and Olsen developed a manipulative gesture system that can turn any surface interactive, e.g. controlling stereo volume along the side of a desk or turning on a television from the corner of a nightstand [FO02]. In ubiquitous computing examples, the same manipulation and semaphore gesture sets could be used to control or interact with multiple objects; *where the user performs the gesture* provides context for what command is carried out or what object is acted upon [FO02; Kar06; Cha17].

Telematics is sometimes viewed as a subset of ubiquitous or pervasive computing [FR11], but is here evaluated separately due to its parallel to a flight deck. With the increase of computing technology in in the automotive industry [KS05], gesture interaction is being investigating for smart vehicles, but the literature lacks a framework for understanding appropriate gesture applications [PBR07]. Alpern and Minardo found that the use of a limited set of eight semaphoric gestures provided a viable alternative to a haptic interface for common secondary tasks in a vehicle cockpit [AM03]. Loehmann et al. developed a culturally independent and easy-to-learn semaphore gesture set, and confirmed its potential as an alternative input modality for secondary tasks in a car [Loe+13]. Pickering, Burnham, and Richardson is investigating applications beyond infotainment systems, but specifically calls out safety critical functions as inappropriate for freehand gesture control [PBR07]. Other telematic applications focus on external controls that offer the user ease of use or convenience, e.g. waving the foot underneath the tailgate to open the tailgate handsfree [Pop16].

2.6.2. Advantages and Disadvantages of Gesture

Research on gesture-based interaction postulates that gesture provides a more natural, intuitive, and simple way to interact with a system [McN96; Wex97; CV05]. The naturalness of gesture interaction stems from well-established psycholinguistic evidence that gesture and speech are integral parts of a single process of communication and are processed in the same part of the brain [McN96]. Manipulative hand gestures are primarily found in association with spoken language [McN06]. Semaphoric gestures provide a greater range of meaning and control, but when poorly mapped to their command input, result in extraneous cognitive load from the user having to memorize a gesture set and retrieve it from LTM [CV05]. Manipulative gesturing can facilitate lexical access [RKC96], and optimize cognitive resources [Gol+01] when accompanied with speech. Alibali, Kita, and Young found that prohibiting gesture leads to decreased speech rate and increased disfluencies in communicating spatial information [AKY15].

A large part of the literature on gesture interaction is focused the addition of other input sources to augment the meaningfulness of manipulation gestures [Bol80; SS01; NS03]. Unimodal applications often use semaphores to create a limited gesture set for shortcuts [CL96; Gar13; Nor18a] whereas multimodal applications are most often manipulative gestures, due to their propensity to be combined with speech [Bol80; SS01; Gol+01; Que+02].

To capitalize on the advantages of gesture interaction, gestures in multimodal interaction should limit the number of pre-defined semaphoric gesture which require prior training, and focus on gestures that come naturally to humans [McN96; Wex97]. Pickering, Burnham, and Richardson found benefits in creating shortcut gestures for frequently used controls, such as dialing home or setting a preset navigation destination in telematics applications [PBR07]. AllSee and WiSee systems program a library of eight to nine gestures to control frequent functions in ubiquitous environments, such as opening a door or turning on an appliance [Pu+13; KTG14]. The Myo armband is a COTS device that has a predefined template of five semaphoric gestures that can be programmed to various shortcuts, while manipulative gesture can be used for scrolling up/down and left/right [Nor18b].

The most common disadvantage of gesture interaction is commonly called "Gorilla-Arm Syndrome", and refers to the fatigue-inducing arm positions which are oftentimes required [Mor15]. David Hincapié-Ramos, Guo, and Irani developed the consumed endurance (CE) metric to quantify and characterize the fatigue (primarily in the arms) a user experiences when interacting with freehand gesture [DGI14]. Manipulation gesture is not effective at quantitative control, where the end result of an interaction is an exact target [LBT02; Mor15]. Gestures often rely on feedback, either through visual confirmation of the desired manipulation or via haptic feedback from the handling of the object itself [Que+02; KS05]. When manipulating content (e.g. buttons, sliders, shapes) with visual feedback, it is easier when the regions of interest are comparatively large and do not require precise selection [Bol80; VB05; Mor15].

2.7. Interaction Taxonomies

Traditionally, interaction taxonomies were based on mechanical and electrical properties of the input or output devices themselves [Bux90; Lin08]. The most common device-based classifications, according to [Bux90; Lin08], are depicted in Figure 2.8.

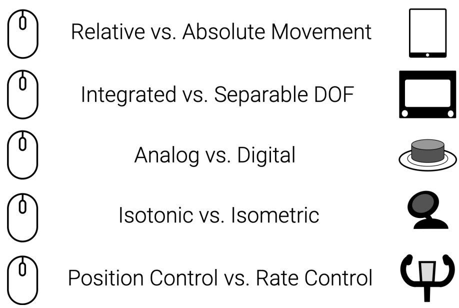


Figure 2.8.: Author interpretation of device-based taxonomies of input from [Bux90] and [Lin08]

Because machines, both in their computing and sensing ability, and device design, are rapidly changing, several attempts have been made to abstract interaction into a taxonomy, in order to standardize human-machine interaction, regardless of what technique a user may choose [HK93; Int94; Ber09]. The Graphical Kernel System (GKS), an International Organization for Standardization (ISO) standard, defines virtual input in terms of the intended action, or task primitives, returned to the computer [Ter85; Int94]. Functional classification schemas, such as GKS, allow system designers to experiment with interaction without needing to know the input devices that individual users will employ [Bux90]. The GKS task primitives, as defined in [EKP12] and ISO/IEC 7942-1:1994, are described below.

Locator returns the real values of a user-selected point in system coordinates.

Choice returns a selection from a set of alternatives.

Pick returns the user-defined selection of displayed objects.

Valuator returns a single, real number.

Stroke returns a sequence of locator coordinates.

String returns a string of characters.

He and Kaufman quantified input devices according to their ability to perform some of these GKS task primitives and the degrees of freedom they provide [HK93]. Their classification is depicted in Table 2.1. The "command" task primitive was replaced with "pick", so as to better match the GKS, as the definition in [HK93] is the same. Hinckley, Jacob, and Ware argue that such a fundamental task approach is not complete, and that advances in technology will invent new interaction task primitives (e.g. a finger scanner or microphone) [HJW14].

Bernsen presents a model of the dominant modalities perceivable by a human via the media of graphics, acoustics, and haptics [Ber08]. The taxonomy is also independent of interaction device, and focuses on all possible ways in which human and computer can exchange information and the properties of each [Ber08; Ber09]. Although considered too granular for a reduced crew flight deck setting, the taxonomy highlights the fact that the design, development, and evaluation of interactive systems is highly complex [Ber95; Ber08].

2.8. Concurrent Task Management

A pilot is responsible for planning and performing multiple tasks whose status must be monitored concurrently, which is referred to as concurrent task management [LDB09]. CTM does not assume that multiple tasks are performed simultaneously, rather that multiple tasks must be monitored concurrently [LDB09; STB09]. A useful way of representing CTM is in terms of the time spent on one task before switching to another [STB09]. This continuum, depicted in

	Degrees of Freedom					Functionality			
	0D	1D	2D	3D	Special	Locator	Choice	Pick	Valuator
Keyboard	*					3	2	1	3
Dial/Turnknob		*				4	4	4	1
Mouse			*			1	1	3	4
Isotrak				*		1	4	4	4
Flying mouse			*	*		1	1	3	4
Spaceball				*		1	2	3	4
Dataglove				*		1	3	1	4
Eye Tracker					*	2	2	4	4
Voice					*	4	2	1	3
Pressure Sensitive Tablet			*			1	2	4	4

1 very suitable 2 suitable 3 can be used 4 not at all suited

Table 2.1.: Author rendition of the original He and Kaufman [HK93] interaction-based input modality classification

Figure 2.9, can range from a second or less to long spans of time between task switches [Sal05; STB09].

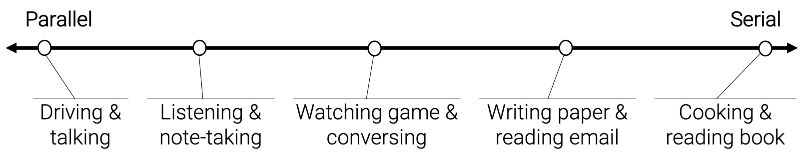


Figure 2.9.: Author rendition of the CTM continuum according to [STB09]

Parallel CTM research, also called dual-task in the literature [Bol+07], dates back as early as the 1930s [Tel31] and has been explored in diverse laboratory ([HRR06; Bol+07]) and real-world situations, most prominently in driving [WH02; HW03] and piloting [MN80; Cha17]. Performing two truly simultaneous tasks is only possible when the human can automate them through experience and practice [LDB09], so real-world driving and piloting studies focus primarily on serial CTM [CCH00; AB04; IH07]. Serial CTM research focuses on measuring the task-switching costs through performance of each task, interruption lag, and resumption lag [Mon03; STB09].

As Loukopoulos, Dismukes, and Barshi discuss, CTM is achieved four ways: simultaneous execution, interleaving steps of each task together, reducing one task's demands by lowering the quality of other tasks, task deferment, or task omission [LDB09]. Each method requires some type of task planning, which itself becomes a task to be managed [LDB09]. Due to the nature of flight deck operations, pilots typically attend to situations of CTM by either deferring or interleaving tasks [WHX02; LDB09; Vid+10].

Loukopoulos, Dismukes, and Barshi cites four common prototypical situations that induce CTM in flight deck operations, which are described below [LDB09].

(1) Tasks that cannot be executed as practiced

(2) Tasks in which new demands arise

(3) Tasks which are interrupted or distracted by another task

(4) Multiple tasks, performed truly simultaneously

Regardless of where a task falls on the CTM continuum, the anatomy of switching between concurrent tasks, depicted in Figure 2.10, is the same. Any task interleaved with or interrupted by another has a beginning state and an end state before the second task is initiated, even if the first is not completed [Sal05]. Interruption lag and resumption lag are the manifestation of the cost in switching cognitive resources between two concurrent tasks [Sal05]. Interruption lag is the time and effort required to bring the first task to a state such that attention can be transferred to a second task [STB09]. Resumption lag is the time and effort required to recall the end state of the first task and resume it [STB09]. Monsell also notes that performance of either task may be affected by the difficulty of the preceding task, causing it to interfere with or require rehearsal of the current task [Mon03].

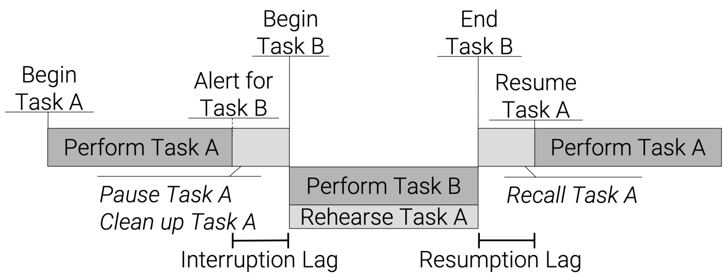


Figure 2.10.: Anatomy of any task along the CTM continuum, adapted from [STB09]

2.9. Research Objective and Gap

The objective of this study is to understand how gaze, voice, and gesture affect an operator's ability to manage multiple concurrent tasks so as to create natural and flexible communication on a reduced crew flight deck. Recent studies have proven that current operations and interfaces are insufficient for reduced crew operations (RCO) [Bla+14; Bai+17]. Replacing input modalities without considering the interaction paradigm can lead to worse performance than the contemporary modalities [WSV83; OC00; Ber08; Mor15; Dee16; Cha17]. On a reduced crew flight deck, the pilot will be responsible for knowledge-based, rather than skill- and rule-based tasks [Emb05; Bai+17; SNS18], requiring human-machine interaction that is flexible, robust, and natural [Sta+13; Bla+14; Sev16a; Bai+17]. Gaze, voice, and gesture interaction techniques, and their multimodal combinations, shift interaction closer to what comes naturally to the human [OC00; RCO10; Tur14; Sev16a; NMF16]. By creating a device-independent interaction framework, flight deck designers can develop pilot workflows suitable for the interaction technique, rather than designing for the limitations of input devices [HK93; Int94; Ber09].

This study is the first to create a device-independent, (multi)modal interaction framework towards the implementation of reduced crew operations. It addresses the need for a new interaction paradigm on RCO flight decks by abstracting interaction away from the input device itself. It specifically considers interaction techniques that are natural forms of human communication. The framework is validated in situations of concurrent task management, which are assumed to be present in any RCO implementation. The five research bodies are depicted in Figure 2.11, with the current study combining them to fill this gap.

Research / Topic	He & Kaufman [HK93]	Bernsen [Ber95; Ber08; Ber09]	Loukopoulos, Dismukes, & Barshi [LDB09]	Salvucci, Taatgen, & Borst [STB09]	Ruiz, Chen, & Oviatt [RCO10]	A-PiMod [Sev16b]	Bailey et al. [Bai+17]	Sterling [Ste17]
Device Independent Interaction	X	X						X
Concurrent Task Management			X	X				X
Multimodality		X			X	X		X
Gaze, Gesture, & Voice	X				X	X		X
Reduced Crew Operations							X	X

Figure 2.11.: Research gap filled by the current study

3. A Device-Independent, (Multi)modal Interaction Framework

The device-independent, (multi)modal interaction framework consists of three main building blocks. The first is a list of assumptions, provided in Section 3.2, describing the reduced crew operations (RCO) flight deck to which the framework can be applied. The second and third building blocks, provided in Section 3.3 and Section 3.4, respectively, are an expanded model of human information processing and a device-independent interaction taxonomy. All three building blocks were tailored for a reduced crew flight deck. The main theory of this dissertation is that interaction technique, specifically gaze, voice, gesture, and combinations thereof, affect an operator's ability to manage concurrent task demands differently. If this is the case, the developed interaction framework can be used to help minimize the costs of concurrent task management (CTM). Assuming gaze, gesture, and voice interaction techniques (and combinations thereof), enable flexible, natural, and robust human-machine interaction, the interaction framework presented herein will help flight deck designers develop suitable interaction techniques for an RCO flight deck.

3.1. The Role of the Pilot on a Contemporary Flight Deck

A common prioritization of a modern pilot's high-level workflow is the aviate, navigate, communicate, manage systems (ANCS) strategy introduced in Jonsson and Ricks's technical report [JR95] and described below.

Aviate Control the aircraft by directing heading, speed, altitude, vertical speed, pitch, roll, yaw, etc. [JR95; Ver97].

Navigate Know where the aircraft is and where it should be heading by referencing external objects, charts, procedures, navigation systems, etc. [Bil96; Ver97].

Communicate Send and receive information using voice or datalink communication systems with air traffic control (ATC), the airline operations center (AOC), crew, and passengers [JR95; Bil96].

Manage systems Monitor the state of aircraft systems (control, navigation, hydraulics, electrical, fuel, communication, environmental, safety, etc.) to identify when actions may be necessary to carry out mission objectives or return the aircraft to a nominal state [JR95].

Cummings, Stimpson, and Clamann found that Boeing pilots spend about 7 min per flight performing aviate tasks, and Airbus pilots spend half of that [CSC16]. The aviate function is performed by the autopilot a significant proportion of the time [Fab13]. A National Aeronautics and Space Administration (NASA) study on task management in non-normal situations found that pilots spent 13% of their time focusing on aviation tasks, 14% on navigation tasks, 11% on communication tasks, and the remaining 62% of the their time was used for system management tasks [ST96]. Neis, Klingauf, and Schiefele argue that pilots today are no longer "aviators" but "system managers" [NKS18]. Boeing and the Technische Universität Darmstadt (TUDA)'s concept for RCO uses the term "mission manager" to describe this paradigm shift in pilot responsibility.

To aid pilots in carrying out today's flight deck operation, required information is organized into the following main sources of information:

Flight operation manuals provided by the aircraft manufacturer and derived from aircraft design [JR95]

Procedures and checklists provided by the airline to adhere to operating and safety standards, and implement company policies [JR95]

Operational information provided by ATC and the AOC to address the specific demands of the day of operation [JR95]

On almost every civil transport aircraft, a two pilot crew divides tasks between the pilot flying and pilot monitoring [LDB09]. As a pilot gains experience, s/he goes from skill-based, to rule-based, to knowledge-based reasoning in order to achieve safe and efficient flight deck operation [Ras83; Emb05; Vid+10; Cum14]. Piloting expertise is proportional to flight hours, and comes from a pilot being exposed to as many different system conditions as possible [Cum14].

3.2. Assumptions: The Role of the Pilot in Reduced Crew Operations

In an RCO world, the aircraft provides vigilance and resistance to fatigue in performing tasks that are deterministic, time constrained, tedious, repetitive, or require great precision, while the pilot contributes creative thinking and intelligence in performing tasks strategic to the overall mission [Sch+07; Bla+14; SNS18]. Human creativity, intelligence, and expertise are required when situations become less deterministic and more uncertain [Cum14].

Removing a pilot would be a simple step technically [Bil96], but requires changes in the operational procedures, crew coordination, use of automation, displays, and roles and responsibilities of the entities acting in an RCO world to maintain the same level of safety seen in today's operations [KPM10; Bla+14; WG15; Bai+17]. Due to advances in information availability, data processing and visualization, ubiquitous sensing, and computing power, the new challenge of a flight deck designer is to decide what information is needed when, and how an operator can best interact with it [AR92; KPM10; Rui11; Bla+14; Tur14]. Boeing and TUDA's assumptions for RCO apply to a larger RCO research thrust [Ins18], and ensure that the results of each study can be applied to the same concept of operations. The assumptions are provided below.

(1) **The RCO pilot's task management strategy will shift from the traditional ANCS paradigm to one of strategic mission management.** Aviation, navigation, and communication tasks will be taken over by increased automation, except in emergency situations. The RCO pilot will monitor and manage the systems that carry out those operations with the help of

a virtual assistant and a ground station [Bla+14; NKS18; SNS18]. The pilot in the proposed RCO world will henceforth be referred to as “mission manager (MM)” [Bla+14; SNS18].

(2) **Traditional co-pilot tasks will be taken over by the MM, ground support, or on-board automation.** The MM is also expected to interact with ground support or on-board automation to manage the mission, when needed [Bla+14; NKS18; SNS18].

(3) **The MM retains the right to refuse or accept automated decisions.** Adhering to Boeing’s flight deck design philosophy, the final decision remains with the MM, even if it requires overruling a decision made by an autonomous system [Abb01; Bla+14].

(4) **The RCO flight deck will have near real-time data exchange.** Global connectivity and a stable datalink affects the information management and representation possibilities on an RCO flight deck. A stable, secure, and near real-time access to any data available to a ground station as well as any data recorded by the aircraft itself is required to realize RCO[Bla+14].

(5) **The redefined role of the MM affects the design of the RCO flight deck.** The MM’s new role will affect the future flight deck’s display layout, interaction paradigm, information management, workflow, and ergonomics, and these aspects need to be designed accordingly [Bla+14].

(6) **Situations of concurrent task management will exist in the RCO world.** A MM will encounter situations of concurrent task demands, in which multiple tasks, or statuses of tasks, must be monitored and managed simultaneously [LDB09; Vid+10; Bla+14].

The information required for the RCO operation described above, and demonstrated in the RCO prototype, which is described in Chapter 5, have not been defined to the granularity of flight operation manuals, standard operating procedures, and checklists. This dissertation therefore seeks to provide guidelines with which to develop such information with a new and natural interaction paradigm, particularly in situations of CTM.

3.3. A Model of Information Processing for (Multi)modal Interaction

In order to design an interaction taxonomy for RCO, a complete model of human information processing was created to provide a thorough representation of all of the relevant resources available to the operator of a contemporary or reduced crew flight deck. To create this complete model of information processing, the author started with Wickens's original box model of Multiple Resource Theory (MRT), and made four major elaborations, depicted in Figure 3.1, and described below.

(1) The haptic, olfactory, and vestibular modalities, and a distinction between focal and ambient vision, were added, as suggested in [KWK07, chap. 4]. The author added the kinesthetic modality, as it is also relevant to flight deck operation.

(2) The modalities were adjusted to qualitatively represent their respective dominance, as suggested in [KWK07, chap. 4].

(3) The author replaced the term "spatial" with "visuospatial" and the term "verbal" with "phonological" on the code axis. This elaboration demonstrates the compatibility of the theory of working memory (WM) with MRT.

(4) The information code was added, as suggested in [KWK07, chap. 4]. The author chose to represent it through the perception and cognition stages to demonstrate the compatibility of cognitive load theory (CLT) with MRT.

The first elaboration provides a complete picture of the perception resources available to a flight deck operator to perceive stimuli. The evolution of flight deck technology has shifted human information processing demands from the vestibular and kinesthetic systems [Vid+10], to ocular and auditory sensory systems [WSV83; Wic08; Vid+10]. The haptic, olfactory, vestibular, and kinesthetic modalities were not tested in this study, but are provided in this expanded information processing model for completeness.

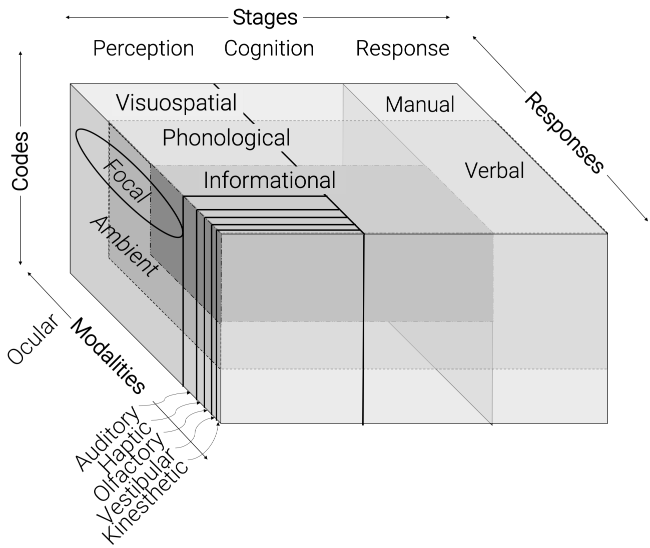


Figure 3.1.: Model of information processing for reduced crew operations, combining Wickens's MRT with abstractions from [KWK07]

The second elaboration provides a qualitative representation of the dominence of each perception resource available to a flight deck operator, or mission manager.

The third elaboration reflects the relationship of the first two stages of MRT with the theory of WM. The author has interpreted perception and cognition stages of MRT to be analogous with WM.

The fourth and final elaboration represents schema for accessing long-term memory (LTM) from WM. The information code is unique to each operator and highly dependent on his/her experience and training, but designing tasks that pull on the information code can expedite the processing and response to a task's stimuli. The information code is a more efficient representation of phonological or visuospatial information, but one which requires LTM schemas for human comprehension. Much of the information on a flight deck requires some level of the information code, such as terminal aerodrome forecasts (TAFs) or ATC chatter, which would be indiscernible to an untrained pilot or mission manager. It's availability as a cognitive resource in information processing is therefore important to highlight. The information code is assumed to end after the cognition stage and not affect response resources available to the human. The response axis in this expanded model, as in the original box model, is limited to manual and verbal resources. Gesture interaction may rely on vestibular and kinesthetic systems as confirmation of a gesture being performed, but their contribution is negligible and therefore not reflected on the model.

This is the first time that the theory of working memory, cognitive load theory, and Multiple Resource Theory have been combined to create a complete model of human information processing for flight deck operators (described individually in Section 2.1). This is also the first time that the elaborations to the MRT modality and code axes from [KWK07, chap. 4] are represented on a full box model. This single model represents all resources a human has at his/her disposal to perceive a stimuli, understand the situation, and execute a response on a flight deck. Every interaction technique a mission manager executes can be mapped to a perception code and modality, cognition code, and response code.

3.4. A (Multi)modal Interaction Taxonomy

To create a (multi)modal interaction taxonomy, each device-*independent* task primitive from the taxonomies presented in Section 2.7 was first assigned a perception/cognition and code mapping, per the model of information processing presented in the previous section. Each interaction technique (both device-dependent and independent) were assigned a code and response mapping. For an interaction technique to be suitable for a task primitive, the mappings of each should be the same or similar. For example, gaze control represents a manual response in the visuospatial code. The "locator" task primitive creates stimuli that are percieved ocularly and processed in the visuospatial code. Gaze can therefore be well suited to a location task primitive. Gaze combined with voice represents a manual response in the visuospatial code with a vocal response in the phonological code. Using the multimodal combination increases the number of task primitives the interaction technique is able to accomplish, but pulls on more information processing resources than a unimodal interaction technique would. An early version of the taxonomy is provided in Table 3.1. Scores in the taxonomy represent each interaction technique's propensity for fulfilling a particular task primitive.

	Degrees of Freedom				Locator	Pick	Choice	Valuator	String	Stroke
	0D	1D	2D	3D						
Keyboard	1	1	2	3	3	3	2	3	1	3
Dial/Turnknob	4	1	4	4	4	4	3	1	4	4
Mouse	4	2	1	2	4	4	2	4	4	4
Button/Switch/Lever	1	4	4	4	3	2	2	3	4	2
Trackpad	4	2	1	2	3	2	2	2	3	2
Touchscreen	4	1	1	2	2	1	1	2	3	1
Gesture	4	2	2	1	3	2	2	3	3	1
Gaze	4	1	1	1	2	2	3	4	4	4
Voice	1	2	2	2	4	3	4	3	1	4
Gaze-Gesture	1	1	1	1	1	2	2	3	3	1
Gesture-Voice	1	2	2	1	3	2	2	1	1	1
Voice-Gaze	1	1	1	1	2	2	3	3	1	4

1	Natural method
2	Natural but requires memorization or study, may induce physical strain
3	Unnatural, may induce physical strain, limited to specific use cases
4	Not suited

Table 3.1.: Early version of a (multi)modal interaction taxonomy, based on [HK93], and expanded through the author's research and experiments

The top section of the taxonomy in Table 3.1 provides guidance for all device-*dependent* interaction techniques relevant for flight deck operation. Device-dependent interaction techniques are not investigated in detail in the current study. From the original taxonomy (replicated in Table 2.1), the "pressure-sensitive tablet" was replaced with "trackpad". Isotrak, flying mouse, and spaceball interaction techniques were considered by the author to be irrelevant for flight deck interaction and removed. Button/switch/lever and touchscreen devices were added. The scores for each interaction technique in this top section are based primarily on He and Kaufman's original taxonomy and analysis of the relevant literature when their taxonomy did not address the specific interaction technique, specifically button/switch/lever and touchscreen. Mappings between device and task primitive (as determined according to the model of human information processing presented in the previous section) were used to update scores for all interaction techniques in the top section.

The bottom section of the taxonomy provides guidance for all device-*independent* interaction techniques relevant for flight deck operation. It was created using the original He and Kaufman taxonomy as a starting point. Device-dependent technical implementations of an interaction technique in He and Kaufman's taxonomy were replaced with device-independent techniques in Table 3.1. More specifically, "Dataglove" was replaced by "gesture" and "eye-tracker" was replaced by "gaze". The three multimodal combinations of gesture, gaze, and voice were then added to the taxonomy. The scores for each interaction technique in this bottom section of Table 3.1 are based off of analysis of the relevant literature and mappings between device and task primitive (as determined according to the model of human information processing presented in the previous section).

But mappings to the information processing model was not enough to determine interaction technique compatibility with task primitive. For example, gesture is also a manual response in the visuospatial code. But it is not suitable for some location task primitives, while it is suitable for others [Bol80; CV05]. Such nuances were determined empirically in three preliminary experiments conducted under supervision of the author [Mor15; Dee16; Cha17]. The results were used to adjust the taxonomy accordingly. The three preliminary experiments investigated the six, device-independent interaction techniques individually. The types of tasks designed in each experiment were determined according to the compatible mappings as determined according to this early version of the taxonomy, presented in Table 3.1. They will now be described.

Moritz [Mor15] investigated the feasibility of gesture using two different technical implementations. The task primitives tested for gesture control were locator, pick, choice, valuation, and stroke. Moritz's results revealed a necessity to split the locator primitive into precise location, as when marking objects, and approximate location, as when activating hierarchical pie menus. This differentiation is supported by literature, as Wickens, Sandry, and Vidulich' spilt the ocular modality into focal and ambient. This differentiation is also reflected in the model of human information processing in the previous section. Moritz's results also showed that while gesture is suitable for qualitative valuation using analog, continuous motion, valuation with discrete target values was ineffective. Gesture, regardless of technological implementation, generates extraneous workload when tuning to precise, predefined values (e.g. a radio frequency) or positions (e.g. an icon on a tablet). Gesture was found to be very natural in setting values that do not require precise end states, however, (e.g. turning up volume or light intensity). These findings suggested that the valuator primitive should also be divided into two separate primitives: continuous and discrete valuation.

Deeg [Dee16] investigated the feasibility of unimodal voice and gaze using two independent technical implementations. The tasks primitives tested for voice were choice, valuation, and string, and for gaze, a series of location and pick track primitives as per their compatibility mappings according Table 3.1. Independently confirming Moritz's results, Deeg revealed the need to differentiate the locator and valuator task primitives, this time for voice and gaze interaction techniques. His results solidified the redefinition of a discrete and approximate locator task primitive and an analog and discrete valuator task primitive.

Charrier's [Cha17] experiments investigated the feasibility of multimodal combinations of gesture, voice, and gaze interaction techniques. The task primitives tested for the multimodal combinations were locator, choice, valuator, and string. Tasks were designed per the compatibility mappings in early versions of the taxonomy, shown in Table 3.1. Charrier's experiments reinforced the need to differentiate between analog and discrete valuation.

All three preliminary experiments made use of the interaction techniques in ways that could not be categorized by even the expanded list of task primitives (expansion of locator and valuator task primitives). The author therefore created a new task primitive, "shortcuts". Shortcuts allow a user to circumvent a series of sub-tasks to achieve a higher-level goal (e.g. balling the hand into a fist or saying the word “play” to perform a task that would otherwise be comprised of a series of locator, stroke, and choice task primitives). The code mapping of the shortcut task primitives drove two separate types of shortcuts: ubiquitous manipulation, e.g. semantic gesture sets processed visuospatially, and ubiquitous articulation, e.g. voice queries, processed phonologically. Both shortcut task primitives require information representation in the information code, as some memorization of the shortcut is required for use. Flexible interaction techniques are better suited for the execution of shortcuts.

Empirically validating each interaction technique against the original task primitives of the graphical kernel system revealed the need to expand the original task primitives from six to 10. This final, validated (multi)modal interaction taxonomy, provided in Table 3.2, is a culmination of He and Kaufman’s original taxonomy, an analysis of the relevant literature presented in Chapter 2,and empirical evidence gathered from three preliminary experiments [Mor15; Dee16; Cha17]. It recommends suitable interaction techniques for each task primitive. To interpret the taxonomy, a task designer must know the task primitive which a task requires to complete. By using the numeric rating scale (or corresponding color code), a suitable interaction technique (device dependent or independent) can be chosen for the identified task primitive. Alternatively, if a task designer is limited to specific interaction techniques, the taxonomy can be used to choose task primitives that will be the most suited.

The taxonomy is limited to interaction techniques relevant to a contemporary or reduced crew flight deck. It includes device-dependent interaction techniques for completeness, but the main focus of the evaluation in the next chapter is the device-independent portion of the taxonomy. The next chapter describes a series of 10 human factors experiments that were conducted to understand how the six interaction techniques described in the bottom half of the taxonomy perform in situations of CTM. Together with the results of the human factors experiments, flight deck designers can develop interactions for an RCO flight deck that are natural, flexible, and intuitive, and apply them efficiently in CTM workflows.

	Degrees of Freedom				Locator		Pick	Choice	Valuator		String	Stroke	Shortcut	
	0D	1D	2D	3D	Precise	Approximate	Identification	Selection	Continuous	Discrete	Articulation	Direct Manipulation	Ubiquitous Manipulation	Ubiquitous Articulation
Keyboard	1	1	2	3	3	3	3	2	3	2	1	3	3	3
Dial/Turnknob	4	1	4	4	4	4	4	3	1	1	4	4	4	4
Mouse	4	2	1	2	4	4	4	2	4	3	4	4	4	4
Button/Switch/Lever	1	4	4	4	2	3	2	2	3	3	4	2	4	4
Trackpad	4	2	1	2	2	3	2	2	2	2	3	2	3	4
Touchscreen	4	1	1	2	2	2	1	1	1	2	3	1	3	4
Gesture	4	2	2	1	3	1	2	2	1	3	3	1	1	4
Gaze	4	1	1	1	1	2	2	3	4	4	4	4	4	4
Voice	1	2	2	2	4	4	3	4	3	1	1	4	3	1
Gaze-Gesture	1	1	1	1	1	1	2	2	1	3	3	1	1	4
Gesture-Voice	1	2	2	1	3	1	2	2	1	1	1	1	1	1
Voice-Gaze	1	1	1	1	1	2	2	3	3	1	1	4	3	1

1 Natural method
2 Natural but requires memorization or study, may induce physical strain
3 Unnatural, may induce physical strain, limited to specific use cases
4 Not suited

Table 3.2.: A (multi)modal interaction taxonomy, based on [HK93], and expanded through the author's research and experiments

4. A Human Factors Evaluation of Interaction Techniques

The following chapter describes a series of 10 experiments which were conducted to determine the effects of six novel interaction techniques on the management of multiple concurrent tasks. The methods for conducting the experiments are provided. Thirty-five participants were tested on a total of 12 possible concurrent task chains. The effects of concurrent task management were measured via interruption time, resumption time, performance of the first and second task, and subjective workload. Interaction technique was found to have a significant effect on the resumption time of an interrupted task, though it does not increase the intrinsic workload of concurrent task management. The stimulus/code/response mapping of a first task was found to significantly affect the performance of a second task performed with certain interaction techniques. Multimodal interaction, while providing flexibility, resulted in less effective concurrent task management when compared to unimodal interaction. Voice interaction techniques interfere more than other interaction techniques in both phonological/vocal type tasks *and* spatial/manual type tasks. It was also found that neither gender nor previous pilot experience had an effect on the operator's ability to manage concurrent task demands. The implications of these results are discussed at the end of the chapter.

4.1. Research Question

The theory of this dissertation is that interaction techniques affect an operator's ability to manage concurrent task demands differently. The derived research question to be answered is therefore:

> How does an interaction technique affect the cost of switching between multiple, concurrent tasks?

To answer this question, a series of 10 human factors experiments were conducted to measure the cost of interaction technique on the 12 concurrent task chains. The third prototypical situation of concurrent task management (CTM), presented in Section 2.8, served as the baseline of the experiments. The third prototypical situation represents the serial side of the continuum presented in Figure 2.9. The fourth prototypical situation, which represents the parallel side of the same continuum, is also a valid candidate for a controlled experiment. A parallel CTM experiment was designed and tested in [Gie18], but did not prove feasible. Additionally, some hypothesize that truly simultaneous task execution can only occur when the tasks are simple enough to be sufficiently automated [Emb05; LDB09]. In the reduced crew operations (RCO) flight deck, automation will be taking over these types of simple and repetitive tasks, leaving the mission manager (MM) responsible for knowledge-based tasks and higher-level thinking [Bla+14; SNS18]. The fourth situation therefore does not match the RCO scenario of this dissertation, and was not investigated further.

To ensure experiment controllability, the first CTM scenario was avoided, so as not to expect subjects to execute tasks contrary to how they had practiced. Adding workload through new, unexpected task demands will increase extraneous load of managing concurrent tasks, resulting in effects associated with task design rather than the interaction techniques employed. The second situation was therefore also not investigated further.

4.2. Experiment Design and Hypotheses

The design of the 10 CTM experiments followed design of experiments (DOE) methodology. Subjects were tested on their ability to perform a serial task chain. The first task (Task A) was interrupted by a second task (Task B), and then the subject had to resume Task A upon completion of Task B. Performance of Task A, interruption lag, performance of Task B, resumption lag, and subjective workload of the entire task chain were measured to determine the effects of interaction technique on the cost of switching between two concurrent tasks, as reflected in the "dependent" variable column in Table 4.1. An overview of all experiment variables is provided in Table 4.1 and are discussed below.

Independent/Controlled	Dependent/Response	Concomitant
Task switch alerts	$perf_A$	Subject fatigue
Time on Task B	$t_{interruption}$	Subject experience
Input device*	$perf_B$	Subject learning style
Interaction technique*	$t_{resumption}$	Time of day
Task A & B difficulty	TLX of task chain	Room temperature

**varied between experiments*

Table 4.1.: Experiment variables in assessing effects of interaction technique on management of concurrent tasks

Task Switch Alerts

Salvucci, Taatgen, and Borst showed that an alert for Task B (artificially creating interruption lag) can make Task A easier to recall and resume, while an immediately interrupted Task A requires more thought upon resumption [STB09]. Because resumption lag was the most important metric to determine task-switching cost due to interaction technique, interruption lag and resumption lag due to experiment design (e.g. by alerting the subject of an upcoming task switch) were minimized. This also ensures a resumption lag derived from interaction technique interference rather than task design. "Begin Task B" is therefore defined as the point when Task A is interrupted by Task B, and "End Task B" is the point when Task B is interrupted by Task A. No previous alerts were given to the subjects to alert them that the tasks were switching.

Time on Task B

The start of Task B was controlled to ensure that the interruption lag due to experiment design was as minimal as possible. To accurately measure resumption time (the metric for resumption lag), the end of Task B was also controlled by experiment design. Because the start and end of Task B were defined, the time on Task B was controlled (set at 1 min). This meant that Task A and B needed to have clearly defined, but user-triggered, starts so as to accurately measure interruption and resumption time.

Input Device

In an ideal experiment, input modality and derived interaction technique should be independent of input device or technology so as to avoid extraneous load derived from the weaknesses of technology itself. But the author was limited to commercial off the shelf (COTS) devices or commercially available software. Section A.3.3, Section A.3.2, and Section A.3.1 offer an overview of technologies available at the time this dissertation was written. For this experiment, the gaze modality was implemented with an Eyetribe eye tracker [The16], the voice modality with the Unity phrase recognition application programming interface (API) [Uni18] or through "Wizard of Oz" techniques[1], and the gesture modality using a Myo armband [Nor18a]. Both Eyetribe and Myo have ceased production.

Interaction Technique

Interaction techniques for Task A and Task B were designed using the device-independent (multi)modal interaction framework presented in Chapter 3. All tasks for both Task A and Task B drew parallels to the type of interaction one could reasonably imagine on an RCO flight deck. Task A represented prototypical tasks that an operator may encounter on a flight deck using traditional interaction techniques. Task B represented prototypical tasks for the gesture, gaze, and voice interaction techniques investigated in this study.

Task A consisted of optimally mapped stimulus/code/response (SCR) tasks using traditional interaction techniques. Task A had two conditions: a visuospatial/manual (S/M) task and a phonological/verbal (P/V) task, as depicted in Figure 4.1. Multiple Resource Theory (MRT) and the Theory of Working Memory (WM), postulate that mental resources are used most effectively when

[1] A human not involved in experiment acts out voice commands using traditional input devices.

stimulus and response are mapped to compatible processing codes [VW81; WSV83; KWK07] (i.e. visuospatial stimuli map best to a manual response and phonological stimuli map best to a vocal response). Non-optimal mappings, i.e. phonological/manual and visuospatial/verbal for Task A were not investigated. This helped to isolate the cost of task-switching due to interaction technique, rather than due to extraneous workload through poor task design and non-compatible mapping. It is reasonable to assume that an RCO flight deck would have tasks with optimal and non-optimal mappings, however. The two-letter code "S/M" will henceforth be used to refer to tasks in the visuospatial code with a manual response. This is to avoid confusion with the verbal response of P/V tasks.

Task B was designed using the most suitable task primitives for each of the six natural, flexible, and intuitive interaction techniques according to the device-independent taxonomy from Section 3.4. Care was taken to adhere to the appropriate SCR mapping whenever possible, as dictated by the model for information processing in Section 3.3. Multimodal gesture-voice, and multimodal voice-eye interaction techniques inherently span across the visuospatial and phonological code. Task B has six possible conditions, according to each interaction technique: unimodal eye tracking/gaze (E), unimodal voice (V), unimodal gesture (G), multimodal voice and eye tracking/gaze (VE), multimodal eye tracking and gesture (EG), and multimodal gesture and voice (GV), as again depicted in Figure 4.1.

Task A & B Difficulty

Resumption lag was the most important dependent variable to determine task-switching cost. To generate resumption lag, both Task A and B were designed to have adequate intrinsic load: Task A to require prospective memory (i.e. generate necessity for Task A recall) and Task B to distract from Task A (i.e. be engaging). Each task was designed to reduce extraneous workload according to cognitive load theory (CLT). Germane load was also kept to a minimum as it was not necessary for participants to build long-term memory (LTM) schemas. All workload of a task is therefore assumed to be intrinsic, thereby isolating the cost of task-switching due to interaction technique rather than task design. A researcher sacrifices some degree of experimental control and introduces variability with tasks of increasing fidelity [WSV83]. Some of this variability was controlled by choosing tasks whose difficulty did not change over time, and training subjects on all tasks before the experiment began.

Response Variables

Resumption lag and interruption lag are the main metrics to determine the interference of one task with another in serial CTM experiments [STB09]. Because of this particular experiment design, interruption lag measured the effect of Task A's SCR mapping on an interaction technique, whereas resumption lag was a direct measure of interaction technique effects on CTM. Resumption lag was therefore the most important metric to determine task-switching costs due to interaction technique. Parts of the experiment (e.g. time on Task B) were controlled to accurately observe the true resumption time. Subjects were told to prioritize each task the same, so the experiment design used a loading task paradigm, where the current task is always the priority, at the expense of the other task. Resumption lag was measured as the time between the end of Task B and the first correct action upon resumption of Task A. Interruption lag was also measured as the time to first correct action (and not just any action).

Performance of the two Task As and six Task Bs were unique according to task design. The performance metrics for each are described in Section 4.3. Subjective workload was measured after the end of each task chain using a modified National Aeronautics and Space Administration (NASA) task load index (TLX) questionnaire. Physical demand was removed from the six workload sub-scales. Subjective workload applies to the 12 possible CTM scenarios rather than to an individual Task A or B.

Concomitant Variables

The experiment was conducted in a climate-controlled room, which was kept between 20 °C and 25 °C. User fatigue, experience with the three input modalities, and learning style were not controlled, but were noted in a demographic questionnaire provided at the beginning of each experiment. Time of day (which may affect user fatigue) was also not controlled, but noted each time an experiment was conducted.

Data was collected for all 10 experiments in the same two-hour session. The overall experiment procedure is depicted in Figure 4.1. A CTM scenario, henceforth referred to as a task treatment, consisted of a Task A condition (traditional, prototypical SCR tasks), which was then subjected to an interrupting Task B condition (one of six levels of natural, flexible, and intuitive interaction techniques, either unimodal or multimodal). After 1 min on Task B, Task A was resumed. The sequence of task treatments were randomized within subjects. The variants of the Task A and B conditions were also randomized per treatment, to avoid any biases from the design of a unique task (i.e. if a task condition requires pictures, those pictures were new every time the task was performed).

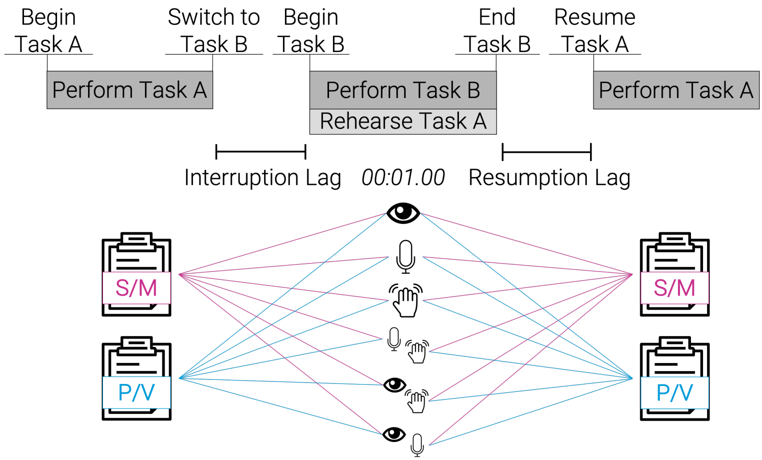


Figure 4.1.: The full experiment design to collect data for 10 separate CTM experiments

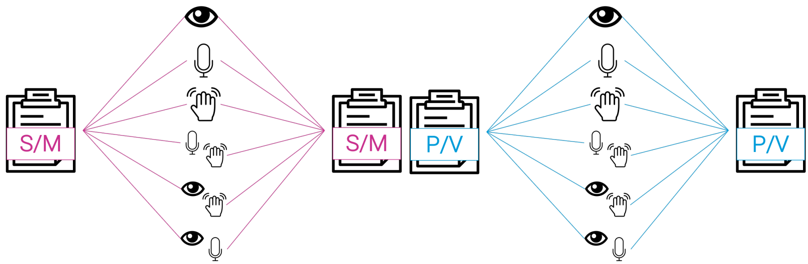
The 10 experiments supported eight null hypotheses, which are summarized in Table 4.2 and described below.

The first null hypothesis, H_1, states that no interaction technique has significantly different effects on the resumption time of an interrupted task. For

Null Hypothesis	Response	Supporting Experiment	Claim
H_1	$t_{resumption}$	X1, X2	There is no significant effect of interaction technique on the resumption time of an interrupted task.
H_2	$perf_A$	X1, X2	There is no significant effect of interaction technique on the performance of an interrupted task.
H_3	TLX	X1, X2	There is no significant effect of interaction technique on the subjective workload of an interrupted task.
H_4	$t_{inerruption}$	X3, X4, X5, X6, X7, X8	There is no significant effect of stimulus/code/response mapping of an interrupted task on the interruption time of a task with a novel interaction technique.
H_5	$perf_B$	X3, X4, X5, X6, X7, X8	There is no significant effect of stimulus/code/response mapping of an interrupted task on the performance of a task with a novel interaction technique.
H_6	$t_{resumption}$	X9, X10	There is no significant effect of unimodal or multimodal interaction techniques on the resumption time of an interrupted task.
H_7	$perf_A$	X9, X10	There is no significant effect of unimodal or multimodal interaction techniques on the performance of an interrupted task.
H_8	TLX	X9, X10	There is no significant effect of unimodal or multimodal interaction techniques on the subjective workload of an interrupted task.

Table 4.2.: The eight experiment hypotheses, their corresponding metrics, and the experiments conducted to test each

example, H_1 claims that the resumption time of a P/V task is not significantly different when interrupted by a voice interaction technique versus a gaze interaction technique. Similarly, H_2 and H_3 claim that the interaction technique does not significantly affect the performance of an interrupted task or the subjective workload of a task treatment. Performance of Task A was measured absolutely after Task B interruption and compared to the performance before Task B interruption. The specific response variables for each Task A will be clarified further in Section 4.3. Significantly degraded Task A performance infers less effective CTM, attributed to the interaction technique of the interrupting task. $H_1 - H_3$ are supported by the first and second experiment, X1 and X2, respectively, which are both single-factor experiments with six levels. X1 tests the three null hypotheses on an S/M task, and X2 tests the same three null hypotheses on a P/V task. The first and second experiments, and the corresponding response variables, are depicted in Figure 4.2.

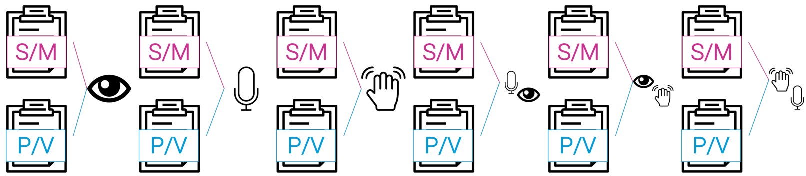


Response Variables	
X1	X2
$t_{resumption}$ $perf_A$ Subjective workload (TLX)	$t_{resumption}$ $perf_A$ Subjective workload (TLX)

Figure 4.2.: Visualization the first and second experiments (X1 and X2) testing null hypotheses $H_1 - H_3$

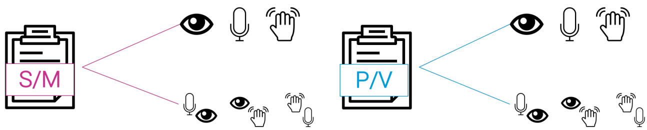
H_4 and H_5 investigate the effect that a preceding task has on a task using one of the six interaction techniques under investigation. In H_4, the task-switching cost was measured by the interruption time from either an S/M or P/V task. H_5 measured the task-switching cost via the performance of Task B. Each Task B is a unique and optimized task for the interaction technique according to Section 3.4. Performance was therefore dependent upon the task design, but can be categorized into three main performance metrics: mean time per action, rate of correct actions, and total completed actions. H_5 suggests, for example, that the task-switching cost of a prototypical eye tracking task will be the same, regardless of whether or not it interrupts an S/M or a P/V task. H_4 and H_5 are supported by experiments three through eight (X3-X8), which are all one-factor with two levels. In each of these experiments, the factor was a specific condition of Task B, compared with the two conditions for Task A (S/M and P/V). Performance of a Task B factor (e.g. an E task) will never be compared to other Task B factors (e.g. a VE task), but rather within the levels of Task A. Experiments three through eight, and the corresponding response variables, are depicted in Figure 4.3. For clarity, the individual performance metrics for each unique Task B are also provided in Figure 4.3, and will be clarified further in Section 4.3.

Finally, effects of unimodal versus multimodal interaction techniques were examined by $H_6 - H_8$. Similar to $H_1 - H_3$, the last three hypotheses measure task-switching cost through resumption time, Task A performance (after Task B and any change before and after the interruption), and subjective workload. They assume the costs are the same, regardless of whether unimodal or multimodal interaction techniques interrupted it. The three unimodal levels were collapsed into a single level, as were the multimodal levels, such that experiment nine and 10 only have two levels, as depicted in Figure 4.4. X9 tests the last three null hypotheses on an S/M task, and X10 tests the same three null hypotheses on a P/V task. The specific response variables for each Task A will be clarified further in Section 4.3.



Response Variables					
X3: E	X4: V	X5: G	X6: VE	X7: EG	X8: GV
$\mu_{time/action}$	$\mu_{time/action}$		$\mu_{time/action}$	$\mu_{time/hit}$	$\mu_{time/action}$
$N_{actions}$	$\frac{N_{correct}}{N_{total}}$	$\frac{N_{correct}}{N_{total}}$	$N_{actions}$	$\frac{N_{correct}}{N_{total}}$	$\frac{N_{correct}}{N_{total}}$
$t_{interruption}$	$t_{interruption}$	$t_{interruption}$	$t_{interruption}$	$t_{interruption}$	$t_{interruption}$

Figure 4.3.: Visualization the third through eighth experiments (X3-X8) testing null hypotheses H_4 and H_5



Response Variables	
X9	X10
$t_{resumption}$ $perf_A$ Subjective workload (TLX)	$t_{resumption}$ $perf_A$ Subjective workload (TLX)

Figure 4.4.: Visualization the ninth and tenth experiments (X9 and X10) testing null hypotheses $H_6 - H_8$

4.3. Task Selection

The following section will describe how the framework from Chapter 3 was applied to design the tasks for each Task A and B condition. From the design of the experiments presented in Section 4.2, requirements for the design of each task condition can be derived. Tasks A and B have four common requirements, and each has two individual requirements, listed below. *Shall* denotes an absolute requirement to control the experiment, while *should* denotes a requirement that involves inherent variability.

REQ_{1AB} The performance of each task shall be measurable.

REQ_{2AB} Each task shall be designed to be feasible for a non-pilot.

REQ_{3AB} Each task (from the measurement of that task's performance) should be designed such that an experienced user is not advantaged over a novice user.

REQ_{4AB} Each task, and therefore interaction technique, should have a realistic analogy to RCO flight deck operation.

REQ_{5A} Task A shall have a clearly-defined, user-triggered start point.

REQ_{6A} Task A should have adequate intrinsic load by being a prospective memory task (necessary to recall something to continue), in order to generate resumption lag.

REQ_{7B} Task B shall reduce extraneous load as much as possible (i.e. be an optimal application of each interaction technique).

REQ_{8B} Task B should have adequate intrinsic load by being engaging enough to sufficiently distract from Task A and generate interruption lag.

REQ_{1AB} was necessary, as performance was a response variable for measuring task-switching costs in all experiments. Because the role of a MM in an RCO world is dramatically different than that of a pilot in today's operation, and to obtain a sample size large enough to show statistical significance, the tasks were designed to be generic and feasible for non-pilots, as stated in REQ_{2AB}. REQ_{3AB}, in the case of Task A, ensures that a test subject's previous experience with one of the tasks would not give the subject an unfair advantage over others (e.g. an expert at solitaire would have better performance, if performance were measured by the game's score). In the case of Task B, REQ_{3AB} meant that a test subject's experience with an interaction technique, or a specific input device, would not provide an unfair advantage over others. The task and its performance should quantify the ability of a user to interact with a chosen technique, and not their ability to use one of the input technologies. REQ_{2AB} and REQ_{3AB} limited task design to generic and achievable tasks, but REQ_{4AB} ensured that the task draws parallels to the type of interaction one could reasonably imagine on an RCO flight deck. For example, a "high-five" gesture is most likely not foreseen in an RCO setting, but a pointing gesture could be. This afforded flexibility in the task design while focusing the application to realistic examples.

For hypotheses H_1 and H_6, resumption time was the main metric for determining task-switching costs. Because Task B was always $1\,\mathrm{min}$ long, REQ_{5A} states that Task A needed to have a clearly-defined, but user-triggered, begin. This ensures that the subject's resumption time could be accurately measured, and contained the same bias for every condition of Task B. REQ_{6A} required the inherent difficulty of Task A to be *reasonably* high, such that a longer resumption time was created, giving more opportunity for variation. A varied resumption lag makes it easier to observe differences in the task-switching costs of specific task treatments. REQ_{7B} ensured that the task chosen for a specific input modality, and derived interaction technique, was optimized according to the advantages of each, which are represented in the taxonomy in Section 3.4. As discussed in Section 2.1.2, it is good practice of a system designer to reduce extraneous load. This was especially important for Task B, as the goal was to understand the effects of interaction technique, not that of the input device or task design, on CTM. REQ_{8B} created a larger and more varied interruption and resumption lag, so as to better observe effects on the corresponding response variables. REQ_{6A} and REQ_{8B} demanded tasks of higher fidelity than the artificial laboratory tasks that are most often seen in CTM literature. The actual task design for Tasks A and B is presented in the next two sections.

4.3.1. Task A: Prototypical Stimulus/Code/Response Combinations

Task A had two conditions, S/M and P/V. These tasks were designed for the optimal SCR mappings according to Section 3.3, and made use of examples in the literature for prototypical task examples, which were then adapted to analogous flight deck tasks and interaction. The interaction taxonomy in Section 3.4 was also used to choose input modality and derive appropriate interaction techniques. A table synopsis of all input modalities, the interaction techniques, task primitives, and analogy to RCO interaction is provided at the end of this section in Table 4.3.

Ocular Stimuli, Visuospatial Code, Manual Response

In the S/M task, a subject was presented with a checklist of items, and the subject must then locate and choose the items from the playing field. The playing field consisted of 100 equally sized, randomly selected images in a 10×10 grid. A checklist of ten categories (albeit phonologically) was presented on the right side of the playing field. The subject was required to work his/her way through the checklist, starting with the first item. From the active checklist item, which was highlighted in yellow, the subject needed to find and select four separate images from the field that belonged to that category using a mouse. A selected image was highlighted in green if the selection matched the active category and did nothing if the selection was incorrect. After the selection of the fourth and final image belonging to a checklist item, the next item in the checklist was highlighted, which was the trigger for the user to begin with the next category. A random number generator between three and seven was used to determine how many checklist items a subject was allowed to complete before Task A was automatically interrupted by Task B. A subject performing the S/M task is provided in Figure 4.5.

Performance of the S/M condition was calculated as the mean time it took a subject to complete a checklist item after Task B interruption, $\mu_{\frac{time}{category}}$, and any changes before and after the Task B interruption, $\Delta\mu_{\frac{time}{category}}$. Incorrectly selected images did not generate a penalty other than affecting the mean time to complete a checklist item (but were recorded). The resumption time for the S/M task was measured as the time between the experiment-controlled Task B

Figure 4.5.: The task presented to a test subject in the S/M condition

end, and the time it took the subject to select the first image of the category which was next on their checklist. Upon resumption of the S/M task, the active category was not highlighted until the subject had successfully selected an image corresponding to the next category, forcing the subject to recall where s/he left off (prospective memory). The S/M task is analogous to a MM monitoring multiple systems, identifying relevant information, and comparing their current workflow against the ideal operation provided in a checklist.

Verbal Stimuli, Phonological Code, Vocal Response

The P/V task consisted of pre-recorded stories and six context questions relating to the story. An audio file of the story was played aloud to the test subject, followed by the first pre-recorded question. The subject was then required to answer the question before the next question was played. After answering the third question, Task A was automatically replaced by Task B. Upon resumption of Task A, the fourth question was automatically played.

Performance of the P/V condition was measured as the mean time it took a subject to answer a question, $\mu_{\frac{time}{answer}}$, and changes before and after the Task B interruption, $\Delta\mu_{\frac{time}{answer}}$. The time to answer a question was measured from the time the question began playing to the first utterance of the subject's response (as determined by an experiment proctor). Filler words (e.g. uh, um, er) do not count as the response and were given a time penalty of $1.0\,\mathrm{s}$. All content questions were controlled to be between 5.0 and $6.0\,\mathrm{s}$ long, with the same length questions in the same position for every story. The fourth question (the question posed after the Task B interruption) was always $5.0\,\mathrm{s}$. Subjects were instructed not to respond until the question had been fully stated. In the event that a subject answered a question before it had finished playing, the time of their first utterance was still used, resulting in some response times less than $5.0\,\mathrm{s}$. The number of questions answered correctly, with missing or irrelevant information, or incorrectly were also recorded. An image of the a subject performing the P/V task is provided in Figure 4.6.

Resumption time for the P/V task was measured as the time between the beginning of the fourth question and the time it took the subject to utter the first word of his/her response. This P/V task is similar to a test given to pilots at some airlines when they are interviewing for a position. They are provided information, interrupted with a secondary task, and after some time required to recall aspects of the information (prospective memory). Comprehension and retention of verbal information is analogous to a MM communicating with support personnel (e.g. flight attendants, ground station) and needing to recall parts of the exchange later on in the mission.

4.3.2. Task B: Interaction Technique

Task B had six conditions (E, V, G, VE, EG, GV). They were designed for the chosen input modality and derived interaction technique according to appropriate SCR mapping, as dictated by the model for information processing in Section 3.3, and the interaction taxonomy in Section 3.4. A table synopsis of all

Figure 4.6.: The task presented to a test subject in the P/V condition

tasks, the interaction techniques, task primitives, and analogy to RCO interaction is provided at the end of this section in Table 4.3.

Unimodal Eye

The unimodal eye condition (E) consisted of a location and selection task. The test subject was presented with a search and find image in the playing field. A legend of ten hidden objects was presented on the right hand side of the searchable image. The eye tracker was used to monitor the subject's gaze position, which was marked by a red square, so that the subject knew the system had correctly identified his/her gaze. The subject was allowed to scan the hidden object image at his/her discretion, but upon finding an object from the legend, s/he must concentrate his/her gaze on the object (predefined in pixels as a region of interest (ROI)) for 1 s. Once an ROI was successfully selected, its corresponding image in the legend was shaded blue, so that the subject knew the selection was successful. After correctly identifying eight ROIs, the test subject was presented with a new hidden object image. Preliminary tests in [Gie18] on the E condition revealed that subjects are somewhat blind to the last two to three objects, as the subject's brain deems some details as unimportant. A subject performing the E task is shown in Figure 4.7.

Figure 4.7.: The locate and select task presented to the subject in the E condition

Performance of the E condition, visualized in Figure 4.3, was measured as the mean time it takes for a subject to successfully identify an object, $\mu_{\frac{time}{action}}$, and the number of correctly identified ROIs, $N_{actions}$. Interruption time, $t_{interruption}$, was measured as the time between the experiment-controlled Task A end the first successfully identified ROI. Searching for information is typical in any workstation. Identifying and selecting information on distant screens or in the surrounding environment via a user's gaze is a realistic MM workflow, given its direct correlation to attention.

Unimodal Voice

The unimodal voice condition (V) consisted of an articulation task. Four distinct categories (e.g. animals, cities, instruments, and cooking) were presented phonologically (printed on a sheet of paper) and presented to the subject at the beginning of the treatment) until the test subject could successfully recite all four. When the V condition of Task B began, noises were played on a loop for the test subject until s/he named a category (as determined by an experiment proctor) to which s/he thought the noise belonged. The next sound was then played automatically. A subject performing the V task is provided in Figure 4.8.

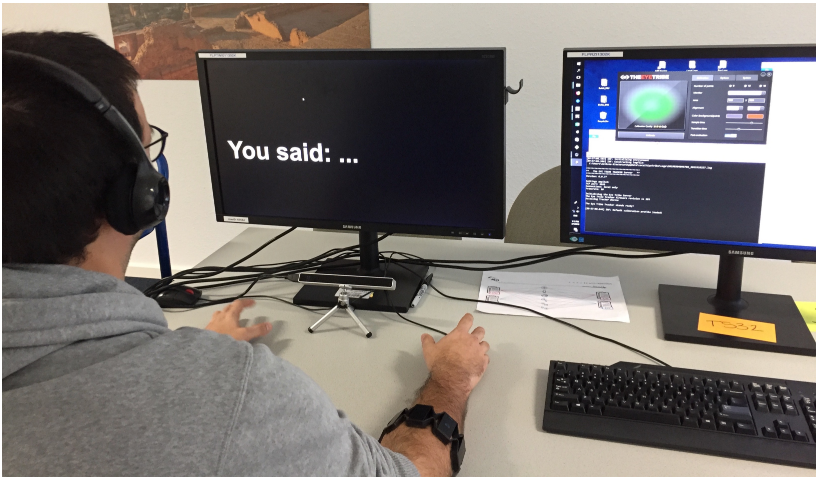


Figure 4.8.: The noise categorization task presented to the subject in the V condition

Performance of the V condition, visualized in Figure 4.3, was measured as the mean time it took for a subject to identify a category, $\mu_{\frac{time}{action}}$, and the number of correctly identified sounds versus total possible sounds, $\frac{N_{correct}}{total}$. The time per category (action) was measured from the time the audio file began playing to the first utterance of the subject's response. Filler words did not count towards the response. Interruption time, $t_{interruption}$, was measured as the time between the experiment-controlled Task A end and the first successfully categorized sound. Categorizing and articulating information is analogous to a MM perceiving information aurally, and classifying it according to its relevance for his/her mission (e.g. normal vibration versus high-pitched whining, air traffic control (ATC) chatter versus mission-pertinent information).

Unimodal Gesture

The unimodal gesture condition (G) consisted of an infinite running game with obstacles, using an analog valuation interaction technique. The playing field consisted of three lanes, with three possible height levels, resulting in a 3×3 grid of possible positions. The test subject could switch between lanes and heights by making a right/left or up/down manipulation gesture, respectively, at his/her leisure. Objects appeared at random, one at a time in the runner's path. Rock walls appear on two of three lanes, forcing the subject to switch to the far left or far right lanes. A log or an arch would span all three lanes. For the former, the subject must move up, and for the latter, the subject must move down. Once an object was cleared, regardless of whether or not the subject hit or missed it, the next object appeared. Either $2.75\,\mathrm{s}$ or $1.50\,\mathrm{s}$ elapsed between object appearance and required action. A subject performing the G task is provided in Figure 4.9.

Performance of the G condition, summarized in Figure 4.3, was measured as the number of cleared objects versus total number of generated objects, $\frac{N_{correct}}{total}$. The interruption time, $t_{interruption}$, was measured as the time between the experiment-controlled Task A end and the disappearance of the first object that the subject cleared (did not hit). The gesture interaction is analogous to qualitative control of individual functions, or approximate indication of desired direction or location when managing information across multiple displays.

Multimodal Voice and Eye

The multimodal voice and eye condition (VE) required the subject to find differences between two images, combining precise location and detection task

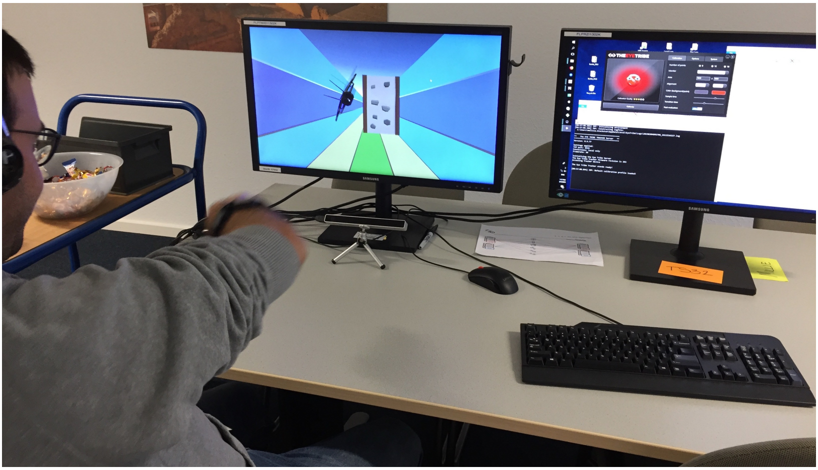


Figure 4.9.: The analog valuation task presented to the subject in the G condition

primitives for the gaze modality and articulation of the voice modality. Eight differences were hidden in two similar, side by side pictures. An eye tracker was used to monitor the subject's gaze position, which was marked by a red square, so that the subject knew the system had correctly identified his/her gaze. The subject was allowed to scan the two images at his/her discretion, but upon identifying a difference (predefined in pixels as an ROI), in either picture, and saying "hit", "da", or "there", a difference was successfully identified. Once a difference was successfully identified, the ROI on both the left and right pictures was shaded blue, so that the subject knew the selection was successful. Similar to the E condition, after six correctly identified differences, the test subject was presented with a new puzzle to avoid skewing the mean with longer search times for differences that the subject's brain had unknowingly deemed unimportant. A subject performing the VE task is shown in Figure 4.10.

Performance of the VE condition, summarized in Figure 4.3, was measured as

Figure 4.10.: The locate and articulate task presented to the subject in the VE condition

the mean time for difference identification, $\mu_{\frac{time}{action}}$, and the number of correctly identified ROIs, $N_{actions}$. Interruption time, $t_{interruption}$, was measured as the time between the experiment-controlled Task A end and the first identified ROI. A MM will be responsible for monitoring multiple screens and systems across his/her workstation and recognize when a system or display is different than its nominal or allowable state (anomalies). The VE task condition is analogous to this type of RCO interaction.

Multimodal Eye and Gesture

The multimodal eye and gesture condition (EG) required a subject to find and destroy objects that appeared at random on a screen. The gaze modality was used to locate and pick the object, while a manipulation gesture was used to destroy the object. A 4×4 matrix of holes was presented to a test subject. The object (a bird in a bubble) appeared randomly from one of the sixteen holes. An eye tracker was used to monitor the subject's gaze position, which was marked by a blue square, so that the subject knew the system had correctly identified his/her gaze. When the subject identified an active object and made a fist (with the dominant hand, on which the subject was wearing the Myo armband), the object was successfully destroyed. If the subject did not hit the object within $2\,\mathrm{s}$ of it appearing, the object disappeared and another object appeared at another random location. The time it took for a new object to appear was randomized between 0–$2\,\mathrm{s}$. A subject performing the EG task is provided in Figure 4.11.

Figure 4.11.: The locate and manipulate task presented to test subject in the EG condition

Performance of the EG condition, visualized in Figure 4.3, was measured as the mean time to hit an object, $\mu_{\frac{time}{action}}$, and the success rate, $\frac{N_{correct}}{total}$. Missing an object was a time penalty in that it increased the time between successful hits. Interruption time, $t_{interruption}$, was measured as the time between the experiment-controlled Task A end and the first successfully hit object. Identifying and selecting information on distant screens or even in the surrounding environment via a user's gaze is a realistic MM workflow, given its direct correlation to attention. Semaphoric gesture is effective for quick activation of a limited set of frequent functions (e.g. opening/closing the flight deck door, or turning a system on/off).

Multimodal Gesture and Voice

The multimodal gesture and voice condition (GV) was an adaptation of one of the most common gesture and voice interactions in the literature, Bolton's "Put-That-There" system [Bol80]. The task used the discrete valuation task primitive of voice to select a shape, and the combined approximate location of gesture and articulation of voice to place a shape. The task's playing field was split into four quadrants. The rows dictated the number of sides a shape had (odd or even number) and columns dictated the color a shape was (green or magenta). A key of five shapes was displayed at the bottom of the screen. Voice was used to select the shape with a predefined value (e.g. "water" or "auto") and a manipulation gesture combined with a voice command ("hit", "da", or "there") was used to indicate approximate object placement. A subject had to use logic to place the selected object into the proper quadrant. Once an object was placed, five new shapes appeared at the bottom of the screen. A subject performing the GV task is given in Figure 4.12.

Performance of the GV condition, summarized in Figure 4.3, was measured as the mean time to place an object, $\mu_{\frac{time}{action}}$, and the success rate, $\frac{N_{correct}}{total}$. Interruption time, $t_{interruption}$, was measured as the time between the experiment-controlled Task A end and the first successfully placed object. Gesture is well suited to ascertain approximate location of a user's attention. Voice is well suited to set discrete, predefined values, e.g. volume at 80%, or to articulate predefined settings, e.g. calling up a predefined procedure.

A table synopsis of all input modalities, the interaction techniques and associated task primitives, and analogy to RCO interaction is provided in Table 4.3.

Input Modality	Task	Interaction Technique	Analogy to RCO Interaction
S/M	Checklist	Precise location (2) Selection (2)	Monitoring multiple systems, identification and comparison of relevant information
P/V	Short stories	Articulation (1)	Communication with support personnel, short term memory recall
E	Hidden object search	Precise location (1) Selection (3)	Environmental perception, identification and selection of relevant information
V	Noise categorization	Articulation (1)	Identification and articulation (categorization) of relevant information
G	Endless runner	Continuous valuation (1)	Qualitative control
VE	Find the difference	Precise location (1) Identification (2) Articulation (1)	Monitoring multiple systems, recognition of anomalies
EG	Whack-a-mole	Precise attention(1) Identification (2) Ubiquitous manipulation (1)	Environmental perception, activation of singular functions
GV	Sorting	Discrete valuation (1) Articulation (1) Approximate location (1)	Articulation (categorization) of relevant information, approximate indication of attention, setting discrete values

Table 4.3.: Summary of Task A & B conditions

Figure 4.12.: The articulation and discrete valuation task presented to test subject in the GV condition

4.4. Experiment Setup and Procedure

A total of 35 subjects participated in the experiment, nine female and 26 male. Eight had previous pilot experience (three hobby pilots, four retired commercial pilots, and one military helicopter pilot). The age of participants ranged from 15 to 56 ($M = 39$, $SD = 16$). Each test lasted an average of 115 min ($SD = 20$).

At the beginning of the experiment, the subject would fill out a pretest questionnaire (provided in Appendix B, Figure B.4) that would capture basic demographics, self-reported fatigue, pilot experience, experience with any of the input modalities, a five-question assessment of learner type, and self-assessed learner type. After the pre-test questionnaire, the Myo armband and Eyetribe eye tracker were calibrated to the individual user. The subject would then be briefed on the experimental procedure and trained on each of the eight task conditions (two Task A and six Task B). Each subject was given the opportunity to repeat any task condition, should they so choose.

The bulk of the experiment (approximately 70 min) was spent performing the 12 unique task treatments. The sequence of treatments were randomized within subjects. The variants of Task A and B conditions per treatment were also randomized, to avoid any biases from the design of specific task variants. At the end of each treatment, the user would fill out a modified TLX questionnaire (provided in Appendix B, Figure B.5) to determine subjective workload.

Upon completion of all twelve treatments, the subject was provided a post-test questionnaire (provided in Appendix B, Figure B.6). The questionnaire sought feedback from each subject about the input modalities and their efficacy in human-machine interaction, as well as any general comments, questions, or feedback. The overall procedure of the experiment is depicted in Figure 4.13.

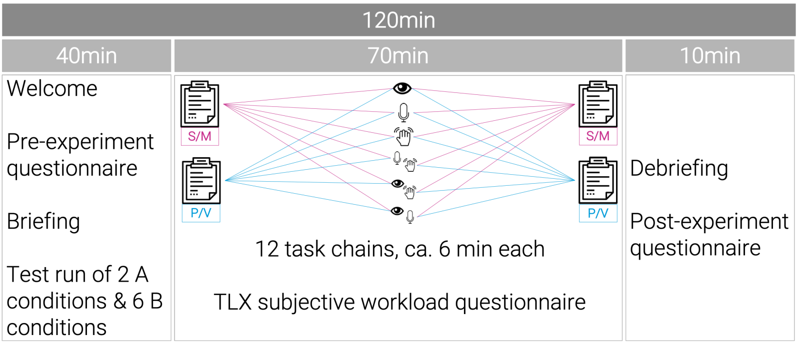


Figure 4.13.: Procedure for the experiment examining interaction technique effects on CTM

The experiment was conducted on a 28-inch display. The Eyetribe eye tracker was placed below the display. The subject was required to wear the Myo armband on his/her dominant arm and a headset with a microphone throughout the entire experiment. The microphone was used to capture voice commands of the subject, while sound came out of the headset speakers. The experimental setup, consisting of workstation and devices, is depicted in Figure 4.14. For each treatment, the playing field filled the entire display.

Figure 4.14.: Equipment setup for the experiment examining interaction technique effects on CTM

4.5. Results

The results are broken down into three main sections: X1 and X2, measuring the resumption time, performance of Task A, and subjective workload, X3-X8, measuring the interruption time and performance of Task B, and X9 and X10, duplicating X1 and X2, but grouping unimodal and multimodal interaction techniques together. Additional findings are reported at the end of the section.

4.5.1. Experiments One and Two: $H_1 - H_3$

Results from X1 and X2 testing H_1 and H_2, are provided in Table 4.4. Outliers of resumption time, defined as any value greater than or less than the mean resumption time of an individual task treatment plus or minus two standard deviations away from said mean ($\mu \pm 2\sigma$), were removed, so as to avoid skewing the distribution. Any value with outliers removed is denoted with an asterisk (*). Performance of both Task A conditions had two possible metrics. Performance of the S/M condition was represented by the mean time it took a subject to complete a checklist item after Task B interruption, $\mu_{\frac{time}{category}}$, and any changes before and after the Task B interruption, $\Delta\mu_{\frac{time}{category}}$. Performance of the P/V condition was represented by the mean time it took a subject to answer a question, $\mu_{\frac{time}{answer}}$, and changes before and after the Task B interruption, $\Delta\mu_{\frac{time}{answer}}$. See Section 4.3.1 for how each performance metric was measured. A one-way analysis of variance (ANOVA) was conducted to determine if interaction technique had an effect on a task treatment's resumption time, $t^*_{resumption}$, and each of the performance metrics.

Resumption time of a task treatment (outliers removed) was significantly affected by interaction technique for both the S/M condition, $F(5,184) = 4.72, p = 0.0004, \eta_p^2 = .06$, and the P/V condition, $F(5,185) = 2.40, p = 0.0385, \eta_p^2 = .03$. Interaction technique effects on resumption time are medium for S/M tasks and small for P/V tasks [Coh88; Fie09]. A box plot of resumption time (outliers removed) for S/M and P/V tasks is provided in 4.15a and 4.15b, respectively.

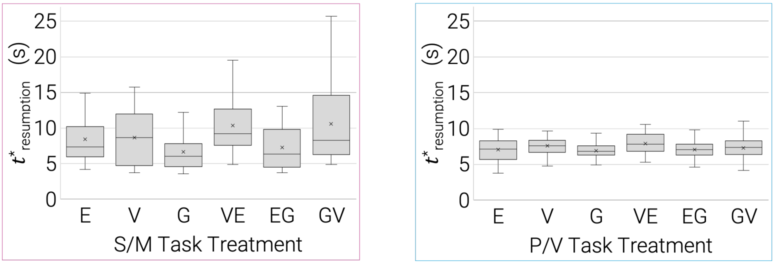
The average time a subject needed to complete a checklist category for the S/M task upon resumption, $\mu_{\frac{time}{category}}$, was not significantly affected by the interaction technique of an interrupting task, $F(5,189) = 1.39, p = 0.2311, \eta_p^2 = .02$, but had a small-sized effect. The change in the time needed per category upon resumption, $\Delta\mu_{\frac{time}{category}}$, was also not significantly affected by interaction technique of the interrupting task, $F(5,189) = 1.90, p = 0.0954, \eta_p^2 = .02$, but also had a small-sized effect. On average, the number of incorrectly selected images went up for each of the task treatments, but no further analysis was performed on the incorrect selections.

	Task Treatment	$\bar{t}^*_{resumption}(s)$	$\bar{\mu}^*_{time/category}(s)$	$\overline{\Delta\mu}^*_{time/category}(s)$
X1: H_1 & H_2	SM−E−SM	8.5 ± 3.5	22.6 ± 8.2	-3.3 ± 14.1
	SM−V−SM	8.7 ± 4.0	26.0 ± 8.1	3.5 ± 10.5
	SM−G−SM	6.7 ± 2.9	21.8 ± 6.5	-2.0 ± 9.3
	SM−VE−SM	10.3 ± 4.2	24.7 ± 7.7	3.2 ± 8.4
	SM−EG−SM	7.3 ± 3.6	24.9 ± 8.7	0.2 ± 11.9
	SM−GV−SM	10.6 ± 5.7	25.3 ± 8.2	0.5 ± 11.6
	Task Treatment	$\bar{t}^*_{resumption}(s)$	$\bar{\mu}^*_{time/answer}(s)$	$\bar{\mu}^*_{time/answer}(s)$
X2: H_1 & H_2	PV−E−PV	7.0 ± 1.6	7.1 ± 1.1	0.4 ± 0.8
	PV−V−PV	7.6 ± 1.2	7.3 ± 1.2	0.1 ± 1.1
	PV−G−PV	6.9 ± 1.2	7.1 ± 1.0	0.2 ± 0.9
	PV−VE−PV	7.9 ± 1.3	7.5 ± 1.3	0.4 ± 1.1
	PV−EG−PV	7.0 ± 1.1	7.3 ± 1.2	0.2 ± 0.9
	PV−GV−PV	7.3 ± 1.5	7.2 ± 1.3	0.3 ± 0.9

Table 4.4.: Results of the first and second experiments, X1 and X2, with regards to the null hypotheses H_1, resumption time, and H_2, Task A performance

The average time a subject needed to respond to a question for the P/V task upon resumption, $\mu_{\frac{time}{answer}}$, was not significantly affected by the interaction technique of an interrupting task, $F(5, 113) = 0.60, p = 0.7008, \eta_p^2 = .01$, but had a small-sized effect. The change in the time to respond to a question upon resumption, $\Delta\mu_{\frac{time}{answer}}$, was also not significantly affected by interaction technique of the interrupting task, $F(5, 191) = 0.67, p = 0.6503, \eta_p^2 = .01$, but also had a small-sized effect. On average, the number of questions answered incorrectly or with missing or irrelevant information stayed the same or decreased for all interaction techniques, but no further analysis was performed on questions answered incorrectly or with missing or irrelevant information.

Results from X1 and X2 testing H_3, are provided in Table 4.5. A one-way ANOVA was conducted to determine if interaction technique had an effect on a task treatment's subjective workload, TLX. Subjective workload was not



(a) (b)

Figure 4.15.: Box plot of resumption time ($t^*_{resumption}$) for (a) X1, the S/M condition, and (b) X2, the P/V condition

significantly affected by interaction technique, neither for the S/M condition, $F(5, 201) = 0.24, p = 0.9235, \eta_p^2 = .003$, nor the P/V condition, $F(5, 202) = 0.54, p = 0.7455, \eta_p^2 = .007$. Effect sizes of interaction technique on both Task A conditions were negligible.

	Task Treatment	$\overline{TLX}$
X1: H_3	SM−E−SM	37.1 ± 17.1
	SM−V−SM	37.1 ± 16.1
	SM−G−SM	36.9 ± 17.5
	SM−VE−SM	36.9 ± 19.3
	SM−EG−SM	38.4 ± 18.9
	SM−GV−SM	41.0 ± 20.2

	Task Treatment	$\overline{TLX}$
X2: H_3	PV−E−PV	35.5 ± 14.8
	PV−V−PV	34.3 ± 16.4
	PV−G−PV	30.0 ± 14.9
	PV−VE−PV	35.2 ± 16.8
	PV−EG−PV	34.4 ± 18.0
	PV−GV−PV	34.7 ± 16.3

Table 4.5.: Results of the first and second experiments, X1 and X2, with regards to the null hypothesis H_3, subjective workload

It should be noted that neither resumption time nor TLX score were normally distributed. Normal probability plots for each are given in Appendix B, Figure B.1, Figure B.2, and Figure B.3. ANOVA is robust to violations of its assumptions

[Fie09], but a Kruskal-Wallis test was conducted for non-parametric data on both resumption time and TLX score. Resumption time (outliers removed) was significantly affected by interaction technique for both the S/M condition, $H(5) = 23.92, p < 0.0002$, and the P/V condition, $H(5) = 10.88, p < 0.0538$. Subjective workload, measured via TLX, was not significantly affected by interaction technique for either Task A condition. The ANOVA results are therefore assumed to be valid.

4.5.2. Experiments Three Through Eight: H_4 and H_5

Results of the six experiments, X3-X8, represented by the null hypotheses H_4 and H_5, are provided in Table 4.6. As discussed in Section 4.3.2, multiple performance metrics for each possible individual Task B factor were measured. Outliers of each individual performance metric, defined as any value greater than or less than the mean of an individual task treatment, plus or minus two standard deviations away from said mean ($\mu \pm 2\sigma$), were thrown out, so as to avoid skewing the distribution. Values with outliers removed are denoted with an asterisk (*). See Section 4.3.1 for how each performance metric was measured for each unique task and interaction technique. The Task B performance metrics are therefore only compared within groups and not between, as represented visually in Figure 4.3. Each experiment had two metrics for Task B performance except for the G condition, which only had one performance metric. A two-sample t-test was conducted on interruption time and each of the performance metrics. The interruption time was found to only be significantly affected by the preceding task's SCR mapping for the E condition (X3) and EG condition (X7). Performance of Task B was never significantly affected, but some small effect sizes were observed. Interruption time and performance metrics not reported below were not statistically significant nor did they have significant effect sizes.

For the E condition, the interruption time, $t^*_{interruption}$, was significantly longer for the S/M task treatment ($M = 7.5, SD = 3.4$) than for the P/V task treatment ($M = 5.5, SD = 2.7$), $t(59) = -2.52, p = 0.0144, r = .31$. The effect is medium-sized. A box plot of interruption time (outliers removed) is provided in 4.16a.

For the V condition, the number of successfully identified categories versus

		Task Treatment	$\bar{t}^*_{interruption}(s)$	$\bar{\mu}^*_{time/action}(s)$	$\bar{N}^*_{actions}$	$\frac{\bar{N}^*_{correct}}{\bar{N}^*_{total}}$
H_4 & H_5	X3	SM−E−SM	7.5 ± 3.4	5.4 ± 2.2	11.8 ± 4.2	
		PV−E−PV	5.5 ± 2.7	6.0 ± 3.4	10.9 ± 5.3	
	X4	SM−V−SM	5.4 ± 2.1	3.3 ± 0.5		$\frac{15.7 \pm 4.9}{18.6 \pm 3.8}$
		PV−V−PV	4.5 ± 1.4	3.3 ± 0.6		$\frac{16.7 \pm 3.3}{19.2 \pm 3.7}$
	X5	SM−G−SM	10.7 ± 2.7			$\frac{23.2 \pm 7.9}{40.8 \pm 1.3}$
		PV−G−PV	11.8 ± 2.7			$\frac{22.4 \pm 7.7}{40.5 \pm 1.3}$
	X6	SM−VE−SM	9.2 ± 4.5	6.5 ± 2.5	8.9 ± 3.7	
		PV−VE−PV	8.4 ± 4.7	6.7 ± 2.9	8.6 ± 3.1	
	X7	SM−EG−SM	5.4 ± 2.5	1.4 ± 0.3		$\frac{11.5 \pm 3.3}{15.5 \pm 0.6}$
		PV−EG−PV	8.3 ± 6.2	1.4 ± 0.3		$\frac{11.0 \pm 3.5}{15.6 \pm 0.7}$
	X8	SM−GV−SM	13.4 ± 4.4	6.9 ± 1.9		$\frac{7.6 \pm 2.9}{9.0 \pm 2.5}$
		PV−GV−PV	13.1 ± 3.8	7.9 ± 3.6		$\frac{7.4 \pm 3.7}{8.5 \pm 3.2}$

Table 4.6.: Results of the third through eighth experiments, X3-X8, with regards to the null the hypotheses H_4, interruption time, and H_5, Task B performance

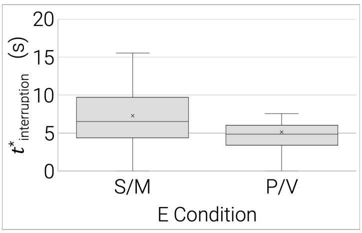
total number presented ($\frac{N_{correct}}{total}$) was greater for the P/V treatment ($M = .89, SD = .13$) than for the S/M treatment ($M = .83, SD = .16$). This difference was not significant $t(63) = 1.70, p = 0.0948$; however it did represent a small-sized effect $r = .21$. The interruption time, $t^*_{interruption}$, was longer for the S/M task treatment ($M = 5.4, SD = 2.1$) than for the P/V task treatment ($M = 4.5, SD = 1.4$). This difference was not significant

$t(50) = -1.81, p = 0.0761$, but it did represent a small-sized effect $r = .25$.

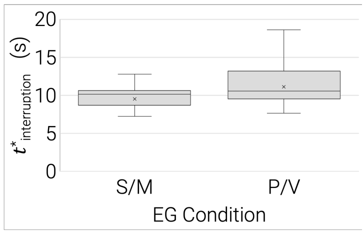
For the G condition, the interruption time, $t^*_{interruption}$, was longer for the P/V task treatment ($M = 11.8, SD = 2.7$) than for the S/M task treatment ($M = 10.7, SD = 2.7$). This difference was not significant $t(61) = 1.55, p = 0.1265$, but it did represent a small-sized effect $r = .19$.

For the VE condition, no performance metric was statistically different or had significant effect sizes.

For the EG condition, the interruption time, $t^*_{interruption}$, was significantly longer for the P/V task treatment ($M = 8.3, SD = 6.2$) than for the S/M task treatment ($M = 5.4, SD = 2.5$), $t(39) = 2.40, p = 0.0212, r = .31$. The effect was medium-sized. A box plot of interruption time (outliers removed) is provided in 4.16b. The number of hit objects versus total possible was greater for the S/M treatment ($M = .75, SD = .21$) than for the P/V treatment ($M = .71, SD = .22$). This difference was not significant $t(63) = -0.78, p = 0.4355$; however it did represent a small-sized effect $r = .10$.



(a)



(b)

Figure 4.16.: Box plot of interruption time ($t^*_{interruption}$) for each task treatment in the (a) E condition and (b) EG condition

For the GV condition, on average, it took subjects longer to place an object, $\mu_{\frac{time}{action}}$, for the S/M condition ($M = 7.9, SD = 3.7$) than for the P/V condition ($M = 6.9, SD = 1.9$). This difference was not significant $t(50) = 1, p = 0.1561$; however it did represent a small-sized effect $r = .20$.

4.5.3. Experiments Nine And Ten: $H_6 - H_7$

Results from X9 and X10 testing H_6 and H_7, are provided in Table 4.7. Outliers of resumption time, defined as any value greater than or less than the mean resumption time of an individual task treatment plus or minus two standard deviations away from said mean ($\mu \pm 2\sigma$), were removed, so as to avoid skewing the distribution. Any value with outliers removed is denoted with an asterisk (*). Performance of both Task A conditions had two possible metrics. See Section 4.3.1 for how each performance metric was measured. A two-sample t-test was conducted to determine if interaction technique had an effect on a task treatment's resumption time, $t^*_{resumption}$, and each of the performance metrics.

	Task Treatment	$\bar{t}^*_{resumption}(s)$	$\bar{\mu}^*_{time/category}(s)$	$\overline{\Delta\mu}^*_{time/category}(s)$
X9: H_6 & H_7	Unimodal	7.9 ± 3.6	24.1 ± 5.1	-1.2 ± 8.7
	Multimodal	9.4 ± 4.8	26.4 ± 7.0	12.1 ± 6.2
	Task Treatment	$\bar{t}^*_{resumption}(s)$	$\bar{\mu}^*_{time/answer}(s)$	$\bar{\mu}^*_{time/answer}(s)$
X10: H_6 & H_7	Unimodal	7.2 ± 1.4	7.1 ± 1.0	0.0 ± 0.8
	Multimodal	7.4 ± 1.3	7.3 ± 1.0	0.4 ± 0.6

Table 4.7.: Results of the ninth and tenth experiments, X9 and X10, with regards to the null hypotheses H_6, resumption time, and H_7, Task A performance

Resumption time of multimodal task treatments (outliers removed) was significantly longer than unimodal task treatments for the S/M condition in X9, $t(173) = 2.39, p = 0.0180$, which represented a small-sized effect $r = 0.18$. Resumption time of multimodal task treatments (outliers removed) was not significantly longer for the P/V condition in X10, and the effect was negligible.

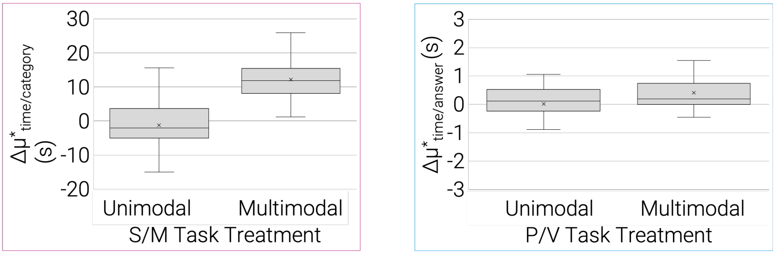
The average time a subject needed to complete a checklist category for the S/M task upon resumption, $\mu_{\frac{time}{category}}$, was greater for multimodal interaction techniques ($M = 26.4, SD = 7.0$) than for unimodal interaction techniques ($M = 24.1, SD = 5.1$). This difference was not significant $t(62) = -1.53, p = 0.1310$, but it did represent a small-sized effect $r = .19$. The change in the time needed per category upon resumption, $\Delta\mu_{\frac{time}{category}}$, increased significantly for multimodal interaction techniques ($M = 12.1, SD = 6.2$) compared to unimodal interaction techniques ($M = -1.2, SD = 8.6$), $t(60) = -7.30, p < 0.0001, r = .69$. The effect was large. A box plot of change in performance (outliers removed) is provided in 4.17a. On average, the number of incorrectly selected images went up for both unimodal and multimodal interaction techniques, but no further analysis was performed on the incorrect selections.

The average time a subject needed to respond to a question for the P/V task upon resumption, $\mu_{\frac{time}{answer}}$, was greater for multimodal interaction techniques ($M = 7.3, SD = 1.0$) than for unimodal interaction techniques ($M = 7.1, SD = 1.0$). This difference was not significant $t(66) = -0.86, p = 0.3949$, but it did represent a small-sized effect $r = .11$. The change in the time to respond to a question upon resumption, $\Delta\mu_{\frac{time}{answer}}$, increased significantly for multimodal interaction techniques ($M = 0.4, SD = 0.6$) compared to unimodal interaction techniques ($M = 0.0, SD = 0.8$), $t(61) = -2.07, p = 0.0427, r = .27$. The effect was small. A box plot of change in performance (outliers removed) is provided in 4.17b. On average, the number of questions answered incorrectly or with missing or irrelevant information decreased for all unimodal and multimodal interaction techniques, but no further analysis was performed on questions answered incorrectly or with missing or irrelevant information.

A two-sample t-test was conducted to determine if unimodal and multimodal interaction techniques had an effect on a task treatment's subject workload, as represented by the final hypothesis, H_8. Subjective workload, measured via TLX, was not significantly affected by unimodal or multimodal interaction techniques, either for the S/M condition or the P/V condition. Effect sizes of unimodal and multimodal interaction techniques on both Task A conditions were negligible.

4.5.4. Additional Results

A two-way ANOVA for H_1 was also conducted to determine whether or not gender or pilot experience had an effect on resumption time of S/M and P/V task treatments (outliers removed). Resumption time of an S/M task treatment was not significantly affected by gender, $F(1, 178) = 0.04, p = 0.8418, \eta_p^2 =$



(a) (b)

Figure 4.17.: Box plot of change in performance of Task A for (a) X9, the S/M condition, and (b) X10, the P/V condition

$.0001$, or pilot experience, $F(1,178) = 0.53, p = 0.4687, \eta_p^2 = .001$. Similarly, resumption time of an P/V task treatment was not significantly affected by gender, $F(1,179) = 1.30, p = 0.2558, \eta_p^2 = .003$, or pilot experience, $F(1,179) = 0.79, p = 0.3757, \eta_p^2 = .002$. Gender and pilot experience were not found to show significant effects for any analysis of interruption time, resumption time, Task A and B performance, or subjective workload.

4.6. Discussion

The results provided in the previous section will now be discussed critically. Each individual hypothesis is discussed, followed by a general discussion, and a summary of implications for the eight null hypotheses.

4.6.1. Experiments One and Two: $H_1 - H_3$

TLX score was not found to be significantly different for any task treatment, implying that interaction technique alone, independent of input device and task design, does not affect the inherent workload of a task. This result serves to justify other analyses between task treatments. Had the TLX score been significantly more or less for any task treatment, any observed effects on other task-switching costs may have been attributed to germane or extraneous cognitive load of a task treatment and not isolated interaction technique effects. The failure to reject H_3 therefore provides justification that Task B conditions can be compared between treatments in X1 and X2, despite the inherently different design of each Task B condition.

From the ANOVA analysis of resumption time ($t^*_{resumption}$), interaction technique does affect the human's ability to manage serial tasks concurrently, both in the visuospatial/manual (S/M) and phonological/verbal (P/V) experiment. The null hypothesis, H_1 is therefore rejected for both conditions. A Fisher comparison was used to group task treatments according to their effective resumption times. 4.8a and 4.8b indicate which interaction technique interferes most with each prototypical stimulus/code/response task.

S/M	N	$\bar{t}^*_{resumption}$ (s)	Grouping		
GV	32	10.6 ± 7	A		
VE	31	10.3 ± 4.2	A		
V	32	8.7 ± 4.0	A	B	
E	31	8.5 ± 3.5		B	C
EG	32	7.3 ± 3.6		B	C
G	32	6.7 ± 2.9			C

(a)

P/V	N	$\bar{t}^*_{resumption}$ (s)	Grouping	
VE	31	7.9 ± 1.3	A	
V	32	7.6 ± 1.2	A	B
GV	33	7.3 ± 1.5	A	B
E	32	7.0 ± 1.6		B
EG	31	7.0 ± 1.1		B
G	32	6.9 ± 1.2		B

(b)

Table 4.8.: The tables provide a summary of grouping information using the Fisher method and 95% confidence for resumption time* as represented by H_1 in both the (a) S/M experiment and (b) P/V experiment. Means that do not share a letter are significantly different.

From the information processing model presented in Section 3.4, gaze (E), gesture (G), and gaze-gesture (EG) techniques should interfere more with an S/M task than techniques that are P/V in nature, namely voice (V). Similarly, the voice interaction technique was predicted to interfere more with a P/V task than gaze, gesture, and gaze-gesture techniques. The voice-gaze (VE) and gesture-voice (GV) conditions were a mixture of SCR combinations, and no prediction was made.

As expected, the voice condition interfered more with the P/V task than the gaze, gesture, or gaze-gesture condition, but not significantly. The opposite of what was expected for the S/M treatments occurred. Gaze, gesture, and gaze-gesture conditions interfered significantly less with the S/M task than the voice condition. No prediction about how gesture-voice or voice-gaze would interfere with an S/M or P/V task was made, but they interfered significantly more than most interaction techniques for S/M tasks. Voice-gaze interfered significantly more than most interaction techniques for P/V tasks. The same three interaction techniques (gesture-voice, voice-gaze, and voice) interfered with both S/M and P/V tasks more than gaze, gesture, and gaze-gesture. This similarity in interference effects may mean that workload due to the task design of those conditions is responsible for the differences in resumption time, rather than isolated effects due to interaction technique. This result also goes against the expectation for the S/M task.

While no significant effects were found in the metrics for performance of Task A, small effect sizes for all performance metrics were observed. This leads to the conclusion that the performance of an interrupted task is affected by the interrupting task beyond the fact that it is simply interrupted. From these results, it can also be assumed that by designing concurrent tasks according to the interaction framework in Chapter 3, performance degradation of S/M and P/V tasks, regardless of the interaction technique which interrupts it, can be minimized. It is clear from the rejection of H_1 and small observed effects in H_2 that interaction technique does have an effect on the task-switching costs of both S/M and P/V tasks. Additional experiments need to be designed, however, that better isolate the effects due purely to interaction technique, rather than the intertwined effects of interaction technique and task design.

4.6.2. Experiments Three Through Eight: H_4 and H_5

From the information processing model presented in Section 3.4, gaze (E), gesture (G), and gaze-gesture (EG) tasks should be more difficult to switch to from a visuospatial/manual (S/M) task than a phonological/verbal (P/V) task. A voice (V) task should be more difficult to switch to from a P/V task than an S/M task. No prediction for voice-gaze (VE) and gesture-voice (GV) conditions were made. As expected, the interruption time for a gaze task was significantly longer after an S/M task than a P/V task. The interruption time for a gaze-gesture task was significantly longer for a P/V task than an S/M task, however, contrary to expectations.

Six possible interruption times were used to test H_4 across experiments three through eight. As long as one interruption time for a given interaction technique was found to be significantly different, H_4 could be rejected, inferring that the type of interrupted task has an effect on the interruption time of a second task with one of the six novel interaction techniques. Two of six proved to be significantly significant, one according to expectations, the other against.

A total of 11 performance metrics were measured to test H_5 across experiments three through eight. No significant performance effects were found, though three small-sized effects were observed for the voice (V), gaze-gesture (EG), and gesture-voice (GV) conditions. If one of the 11 performance measures for a given interaction technique was found to be significantly different, H_5 could be rejected. This was not the case.

The minimal effects that were observed on both the interruption time and task performance lead to the conclusion that the cost of concurrent task management can be minimized when tasks are designed according to the proposed interaction framework in Chapter 3. Significant differences in interruption time and task performance would infer ineffective use of information processing resources. Because Task B performance never varied, and interruption varied significantly in only two of six cases, it can be concluded that each interaction technique was applied to the task effectively, leading to minimal interference of information processing resources.

4.6.3. Experiments Nine and Ten: $H_6 - H_8$

The last two experiments looked specifically at differences between unimodal and multimodal interaction techniques and their effects on the resumption time, performance, and subjective workload of prototypical visuospatial/manual and phonological/verbal tasks. In the S/M condition, resumption time was significantly worse for multimodal interaction techniques than for unimodal. For both Task A conditions, performance deteriorated significantly more after it was interrupted by multimodal interaction techniques than unimodal interaction techniques. Even when designed according to the interaction framework in Chapter 3, multimodal interaction techniques significantly decreased the performance of the tasks they interrupted. This leads to the conclusion that while multimodal interaction techniques can provide flexibility and intuitiveness in an interface, they can negatively affect the cost of switching between concurrent tasks.

4.7. Hypothesis Acceptance and Conclusions

Table 4.9 provides a summary of the results of the eight major hypotheses that were posed before the experiment. H_1 was rejected, which means the alternative hypothesis can be accepted, namely, that an interaction technique has significant effects on the resumption lag of an interrupted task. The alternative hypothesis was accepted for both visuospatial/manual and phonological/verbal. H_2 was not rejected, although small effects were observed for both performance metrics for both visuospatial/manual and phonological/verbal tasks. This can also mean that designing tasks according to the interaction framework in Chapter 3 can minimize performance degradation. The results failed to reject H_3. This result justifies a comparison between tasks treatments, however, rather than only comparing within task treatments.

H_4 was rejected and the alternative hypothesis accepted because the type of task can have an effect on the interruption time of at least gaze and gaze-gesture interaction techniques. H_5 was not rejected, however, again leading to the conclusion that designing tasks according to the interaction framework in Chapter 3 can minimize performance degradation. Both H_6 and H_7 were rejected. The alternative claim for both of these hypotheses is that multimodal interaction can have a higher cost when managing multiple concurrent tasks over unimodal interaction. H_8 was rejected, which again justifies a comparison between unimodal and multimodal task treatments, as all tasks were considered equally difficult.

In addition to the eight experimental hypotheses, gender and pilot experience were used as control groups to determine whether or not they affected concurrent task management or the ability to use the six investigated interaction techniques. Neither gender nor previous pilot experience had an effect on the costs of task switching. This result strengthens the assumptions around the changing role of the reduced crew operations mission manager. The absence of gender and pilot experience effects implies that as the role of a mission manager shifts to system management and knowledge-based tasks, and new interaction techniques are introduced to the flight deck to support that role, a mission manager candidate is not limited to the characteristics and skills of pilots today.

The rejection of four of the eight original null hypotheses confirms the theory of this dissertation, namely:

> Interaction techniques, specifically, gaze, voice, and gesture, and combinations thereof, affect an operator's ability to manage concurrent task demands differently.

Null Hypothesis	Response	Result	Claim
H_1	$t_{resumption}$	Rejected	Interaction technique has a significant effect on the resumption time of both visuospatial/ manual and phonological/verbal tasks.
H_2	$perf_A$	Failed to reject	Interaction technique has a small but insignificant effect on the performance of both visuospatial/manual and phonological/ verbal tasks. Designing tasks according to the proposed interaction framework can help reduce the costs of concurrent task management.
H_3	TLX	Failed to reject	Interaction technique does not inherently increase the workload of concurrent task management.
H_4	$t_{inerruption}$	Rejected	The type of interrupted task can have an effect on the interruption time of a second task with at least two of the six investigated interaction techniques.
H_5	$perf_B$	Failed to reject	Designing tasks according to the proposed interaction framework can help reduce the costs of concurrent task management.
H_6	$t_{resumption}$	Rejected	Multimodal interaction techniques can cause significantly longer resumption time than unimodal interaction techniques for visuospatial/manual tasks.
H_7	$perf_A$	Rejected	Multimodal interaction can cause significantly worse performance of an interrupted task than unimodal interaction for both visuospatial/manual and phonological/verbal tasks.
H_8	TLX	Failed to reject	Multimodal interaction does not inherently increase the workload of concurrent task management over unimodal interaction.

Table 4.9.: Summary of null hypotheses and alternative hypotheses

5. Development of a Reduced Crew Operations Demonstrator

Based on the interaction framework, presented in Chapter 3 and tested in situations of concurrent task management (CTM) in Chapter 4, a reduced crew operations (RCO) demonstrator was built into a Mercedes Viano van. This prototype serves to demonstrate how gaze, voice, gesture, and combinations thereof, can be used on an RCO flight deck. The following chapter describes the iterative design process used to build the van demonstrator and describes the six operational scenarios that a user encounters and solves with a multimodal approach.

5.1. Demonstrator Design

This study was the first attempt to take a high-level RCO concept [Bla+14] from an idea on paper to a physically functioning model. The original design for the RCO demonstrator stems from prior research conducted at Boeing in Neu-Isenburg, Germany and is depicted in Figure 5.1. The van demonstrator is a physical realization of this concept with a focus on multimodal interaction, as dictated by the multimodal interaction framework presented in Chapter 3. The main goal of the demonstrator was to apply the six interaction techniques investigated in this study to use cases on an RCO flight deck. It neither represents the entire concept in [Bla+14], nor an entire RCO concept of operations. The assumptions from Section 3.2 apply to the vision demonstrator as well. The multimodal interaction requirements for the demonstrator are provided in Table 5.1.

The prototyping process began with the proportional design drawing shown in

Input Modality	Hardware requirements	Software requirements
Touch screen	Touch display	Touch listener
Trackpad	Touch sensitive surface	
Keyboard	String input device	Button press listener
Gaze	Gaze capture device	Gaze listener
Voice	Microphone	Speech processor
Gesture	Gesture capture device	Gesture listener

Table 5.1.: Interaction requirements for future flight deck (FFD) vision demonstrator

Figure 5.1. The largest difference between the contemporary flight deck and this design is the replacement of the traditional cockpit displays (e.g. primary flight display (PFD), navigation display (ND), engine indicating and crew alerting system (EICAS), mode control panel (MCP)) by a central mission management display and immersive head-up displays (HUDs). The RCO design also depicts a clear separation of information according to the high level workflow it supports. Aviation and navigation information, with the help of augmented reality (AR), is now almost exclusively shown in the head-up position on the windscreen. The benefit of this setup is to provide the information where it is relevant, thus eliminating the need for a pilot to mentally extrapolate information from an auxiliary display into the real-world surroundings. Communication and mission management information is found primarily on the central mission management display. The design also reflects the major assumption (see Section 3.2) that the role of the operator (mission manager (MM)) of a highly automated flight deck has evolved from a flight-specific disruption manager to a fleet-wide decision maker.

Figure 5.2 depicts the various steps of the prototyping process that led from an idea on paper to the physically functioning vision demonstrator. The first step was to create the ideal ergonomic model of a flight deck optimized for multimodal human-machine interaction, regardless of any technology shortcomings. Once the ergonomic model was created, a virtual computer-aided design (CAD)

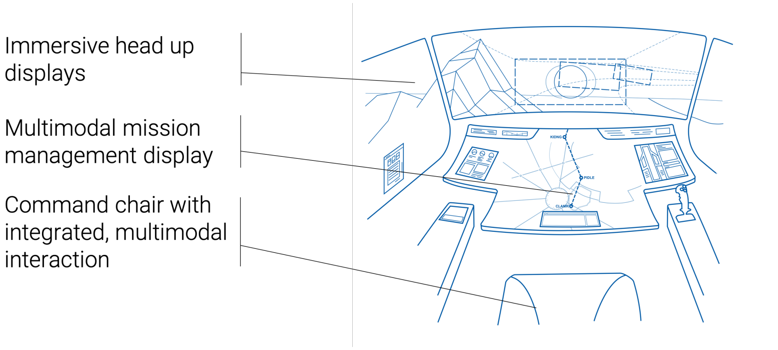


Figure 5.1.: Boeing's FFD 2040 concept design, with permission from [Bla+14]

drawing was developed to adapt the ergonomic model to a Mercedes Viano van platform. A mobile platform was used in order to showcase the RCO prototype with multimodal interaction externally, while also encouraging a participant to think outside the confines of a conventional flight deck. Once the CAD model had been adapted to the van dimensions, a desktop simulator was created to test the intersystem functionality of the incorporated technologies. The entire system was then transferred to the van platform and is currently used by Boeing in Neu-Isenburg, Germany for human factors evaluation and customer demonstration.

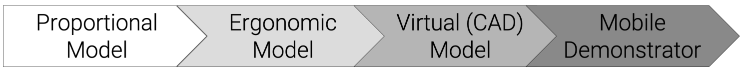


Figure 5.2.: Design process for functioning FFD demonstrator

This first RCO flight deck prototype includes all the input modalities presented in Table 5.1, from which appropriate interaction techniques were derived and applied to example RCO use cases. The van demonstrator also provides opportunity for further technology integration.

5.1.1. Ergonomic Model

The ergonomic model starts with the proportional design in Figure 5.1, and builds the ideal workstation for human interaction with the input modalities defined in Table 5.1. The immersive HUDs and the mission management display are the two main design criteria for human-centered design of the RCO workstation. It is assumed that the immersive HUDs in Figure 5.1 are not within arm's reach of the flight crew and need only be visible over the normal range of head motion. Immersive HUDs are therefore candidates for distant display interaction via gesture. The mission display, however, replaces all other instruments that are found in a traditional flight deck, and is the MM's main interaction and information management mechanism. As such, it was designed such that all information is readily visible and accessible to the pilot.

Of the input modalities listed in Table 5.1, touchscreen is the only method that is restricted to interaction in a confined area [Wag+96]. The mission display should therefore be designed such that, where necessary, the pilot can adequately reach and operate the corresponding display surface. The exact ergonomic form factor of an interactive mission display was designed in an advanced research project at the institute for Flight Systems and Automatic Control (FSR) at Technische Universität Darmstadt (TUDA) and is documented in detail in [Kon+15]. The method, using reach and viewing envelopes, is briefly summarized below.

According to the National Aeronautics and Space Administration (NASA)'s Human Design Integration Handbook [NAS10], two boundaries are required to define a reach envelope: (1) maximum functional reach from the body and (2) area too close to the body that cannot be reached because of physical restrictions. The reach envelope was constructed using 90^{th} percentile male and 10^{th} percentile female anthropometric data provided in Appendix B of [NAS10].

From the Federal Aviation Administration (FAA)'s Human Factor's Design Guide [Wag+96] and guidelines for Part 23 cockpit design [Gen00], the primary field of view (FOV) should be used for critical information, but the secondary FOV can be used for information display and interaction. A human's normal

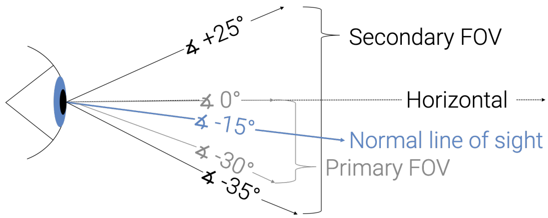


Figure 5.3.: Author rendition of vertical visibility envelope for ideal mission display interaction

line of sight is defined as $15°$ below horizontal, primary FOV as $\pm 15°$ from the normal line, and secondary FOV as $40°$ and $-20°$ from the normal line [Wag+96; Gen00]. The secondary FOV becomes the design visibility envelope, as depicted in Figure 5.3. Guidelines for Part 23 cockpits are particularly relevant given the ultimate goal of RCO to be single rather than dual pilot.

Combining the vertical reach and visibility envelopes from their respective design reference points results in overlapping envelopes. The upper edge of the display was chosen to be the intersection of the 10^{th} percentile female's reach envelope and the normal line of sight so as not to obstruct the external view, but also be able to display some information in a head-up position [Kon+15]. A touch screen should be positioned at an angle between $30° - 45°$ from the horizontal to avoid fatigue [Wag+96]. Using this criteria, the lower edge of the display was then found by the intersection of a $45°$ tangent with the lower limit of the visibility envelope. Because a head up position is preferred in a flight deck setting, the maximum angle possible was used.

An arc between the upper and lower edges (not exceeding the reach envelope) was created as the ideal vertical form factor. This method of combining reach and visibility envelopes is reconstructed in Figure 5.4. For a complete overview of the process, the reader is referred to [Kon+15].

The ideal horizontal form factor was found using the same method of combining horizontal reach and visibility envelopes, and can be found in [Kon+15].

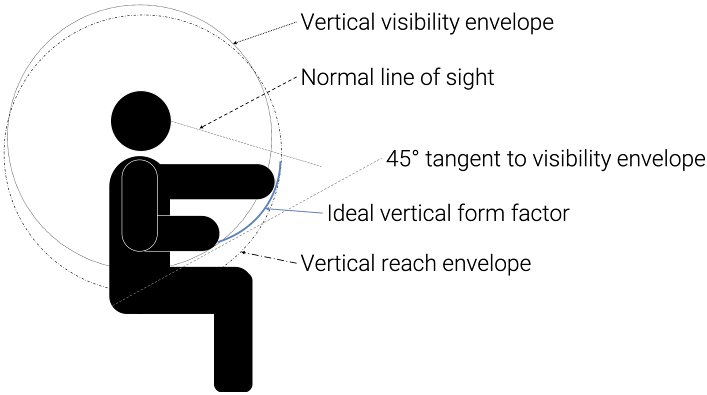


Figure 5.4.: Method for creating ideal vertical form factor for a human-machine interaction centric, mission management display, using a combination of vertical reach and visibility envelopes

The reach envelope was again constructed using 90^{th} percentile male and 10^{th} percentile female anthropometric data provided in Appendix B of [NAS10]. The secondary horizontal FOV is $\pm 60°$ from center, as per [Wag+96] and [Gen00].

Drafting the vertical curve along the horizontal curve in a CAD program provides the ideal, three-dimensional form factor for the central mission management display, and is depicted in Figure 5.5. This is the theoretical optimum shape of the mission management display, when designed for touch interaction and information visibility, and assuming technology is not a limiting factor.

Such a form factor is only technically feasible with curved or flexible displays, which are currently still in a development phase, and not readily available to a consumer market. Based on the advantages and disadvantages of various commercial off the shelf (COTS) touch screen technologies (described in detail in Section C.1, [Kon+15], and [Mhl+16]), a Displax Skin Fit, capacitive touch film was custom-created to retrofit an LG 34UC97 monitor with up to 40 simultaneous touch points and arm rejection. This allows the user to rest his/her arms on

Figure 5.5.: Ideal form factor for a mission management display optimized for human-machine interaction, with permission from [Kon+15]

the mission management display without interfering with the touch interaction, reducing arm fatigue.

5.1.2. Virtual Model

From the onset of the design process, the target platform for the demonstrator was a Mercedes Viano van. The benefit is twofold: (1) the vision demonstrator can be shown externally, and (2) by using a non-traditional setting, participants are encouraged to think outside the confines of a conventional flight deck.

With a solution for the technical implementation of the mission management display identified (see previous section and Section C.1), the rest of the multimodal interaction concept could be built around it. The remaining elements of the multimodal interaction requirements, given in Table 5.1, needed to be integrated such that a user could use the modalities from his/her position behind the mission management display.

Figure 5.6 shows a CAD model of the frame that was designed, with the help of a third party company, to hold the mission display, an "outer-view" display[1] emulating the windows of a flight deck, and the interaction technologies. Due to the limited construction space, one third of the display height is blocked from view, but this results in a wrap-around effect, and is considered advantageous.

The location of the installed devices, in reference to Figure 5.6, is described below.

[1]The design of an outer-view display solution is described in detail in [Mhl+16].

Figure 5.6.: Virtual model of a reduced crew operations demonstrator to be built in a Mercedes Viano van, specifying location of input devices

(1) Touch display and string input device, realized by touchscreen, described in Section 5.1.1, and soft keyboard

(2) Empty space for further technology integration

(3) Gesture capture device, realized by Leap Motion [Lea18a], centered to enable interaction from either hand

(4) Touch sensitive surface, realized by COTS wireless trackpad, set on right side due to majority right-hand population

(5) Gaze capture device, realized by an Eyetribe eye tracker, centered to user, below eye-level to avoid eyelash interference [Dee16]

The voice input modality was realized through a COTS wireless headset that a participant wears during a demonstration.

5.1.3. The Mobile Van Demonstrator

To build the physical model into the van, paper and wood prototypes were first created. These are depicted in 5.7a and 5.7b, respectively. The paper and wood prototypes served to finalize the dimensions of the demonstrator before constructing it out of medium-density fiberboard and finishing it with lacquer. The final construction was completed by a third party manufacturing company. The final dimensions can be found in Section C.2 for completeness. The wood components were then installed, together with the displays, computers, and interaction devices. The result of the iterative prototyping process is provided in Figure 5.8.


(a)


(b)

Figure 5.7.: (a) Paper and (b) wood prototypes that were created to verify dimensions and interaction surfaces before constructing the final model

5.2. Reduced Crew Operations Scenarios Demonstrating Multimodal Interaction

The final RCO mobile van demonstrator is used to demonstrate the multimodal interaction framework that was developed in Chapter 3 and evaluated in CTM situations in Chapter 4. The multimodal interaction techniques are highlighted in

Figure 5.8.: Final model of a reduced crew operations demonstrator built in a Mercedes Viano van, highlighting multimodal interaction techniques

Figure 5.8 and described below. This section will also describe the six operational scenarios that were chosen to demonstrate these interaction techniques on an RCO flight deck.

(1) Voice is used to articulate predefined settings, shorten response time to information queries, and provide shortcuts for quick activation of a limited set of frequent functions.

(2) Gesture is used for qualitative control of individual functions and approximate indication of desired direction or location when managing information at a distance.

(3) Gaze is used to determine approximate and precise attention to acquire context about user intention.

(4) Touch is used to dive into detailed information and select decisions.

The operational scenarios were derived from the high level concept in [Bla+14] and combined with the multimodal interaction framework to depict a flight, from departure to arrival gate, under RCO. The complete demonstration is currently presented to aviation experts and Boeing customers in Neu-Isenburg, Germany to collect qualitative feedback on the state of RCO research.

5.2.1. Hotspot Identification

The International Civil Aviation Organization (ICAO) defines a hotspot as a location on an aerodome movement area with a history or potential risk of collision or runway incursion, and where heightened attention by pilots/drivers is necessary [Int09]. Jeppesen's airport charts started to include the location of hot spots in 2001, and soon afterwards Airport Moving Map functionality in their Flight Deck Pro application incorporated hot spot notes [Ros10]. Hot spots are tabulated and a short description of each is given on a separate notes page.

This first interaction in the RCO demonstration takes the concept of hotspot identification a step further by using gaze to determine a user's intention to look up a hotspot and voice to shorten the query time. Hot spots are geo-located for the demo participant and highlighted in the outside world with the help of AR. The description of the hot spot, if desired, can be displayed next to the hot spot depiction by looking at the hotspot overlay and querying via voice for more information. Figure 5.9 depicts the hotspot identification scenario. The activated hotspot can be seen in the lower left corner of the outer view display.

Specifically, in the RCO demo, this interaction technique is implemented when the user looks at a highlighted region of interest (ROI) (via the outer view display emulating the flight deck windscreen). Image recognition, implemented by adapting open source code for real time object recognition [RF18], identifies predefined ROIs. If the user's attention is focused within the bounding box of one, and the user says "show details," the active hotspot will be supplemented to show pre-prepared hotspot notes. In Figure 5.9, the note is that a protruding propellor causes a risk of incursion, and only vehicles under three meters are allowed to pass.

Figure 5.9.: Hotspot identification via gaze and query for additional hotspot information via voice

5.2.2. Situation Awareness in Ground Operations

The next scenario demonstrates the benefit of integrating various aviation information systems to reduce the time and effort required to find and understand relevant information during ground operations. This streamlined workflow is demonstrated by allowing the user to query about other aircraft and ground vehicles (cars) at the airport (streets around the Boeing facility) without needing to contact air traffic control (ATC) or the other “pilots”. In an RCO world (and today), a MM would benefit from the integration of flight, fleet, and flow information. In times of high traffic, for example, ATC can focus on giving clearances and controlling aircraft, and pilots would still be able to query for more information about the current airport situation. Humans are far more likely to accept a situation if we understand the reason for it [Lic10]. By providing the various actors at the airport with more information, each can focus on his/her specific tasks, and MMs need not agonize about why they are stuck behind another aircraft.

The RCO demo implemented this interaction technique using computer vision to highlight ROIs, gaze to determine user attention and intent, and voice to expedite information queries. Cars (representing airplanes), bikes (representing ground vehicles), and humans (representing ground personnel) are highlighted by computer vision, and if the user's gaze is detected on one of the ROIs, it becomes actionable. Six such ROIs can be seen in Figure 5.8. Paired with a user voice query (e.g. "show details"), pre-prescribed (fake) information populates in an information pop-up next to the ROI in question. Information to be displayed next to each queried item could include intended runway, time until take off, place in queue, or destination.

5.2.3. Communication with Air Traffic Control

The next scenario demonstrates the benefit of gesture to manipulate information across multiple, distant displays. In a highly-automated RCO flight deck, a MM need not process every incoming ATC communication, but rather confirm that the aircraft is implementing the instruction.

In the RCO demo, ATC requests a runway change during taxi, due to heavy traffic on the flight's currently scheduled runway. The notification comes visually and audibly. A typed text notification appears and expands in the upper left corner of the windscreen, as can be seen in Figure 5.10, and audio of the request is played. The MM uses gesture to pull the notification from the peripheral outer view display to the mission management display, and detailed information is then displayed about the runway change and any downline impacts on the flight plan or fleet schedule, as can bee seen in as can be seen in Figure 5.11. The MM is presented with the option to accept or reject the change, along with associated financial and time costs of each option. Upon accepting the change (single finger tap on the graphical user interface (GUI)), the mission management display returns to its default, aircraft-centric view.

In the demo, automation (a virtual copilot) sends the MM decision back to ATC, and updates the flight plan accordingly, such that the aircraft continues automatically along the modified route. This story suggests that the clearance and communication with ATC is handled entirely through this virtual assistant.

Figure 5.10.: Communication with ATC about a last minute runway change in the RCO demonstration, starting with a MM notification in the periphery of the outer view display

It also implies that the runway change request was already validated for feasibility, before being presented to the MM (i.e. the assistant would not propose something that exceeded aircraft performance limits; it would instead reject the request altogether). The only pieces of information necessary to reveal to the MM is the fact that ATC is requesting a runway change and that the flight management system (FMS) was updated successfully to reflect the new taxi route.

5.2.4. Predictive Maintenance, Duty Regulations, and Enroute Optimization

Once enroute, the pilot receives a series of notifications about issues (maintenance, crew, trajectory) with which the system needs human intervention to address. The goal of these scenarios is to demonstrate a multimodal interaction

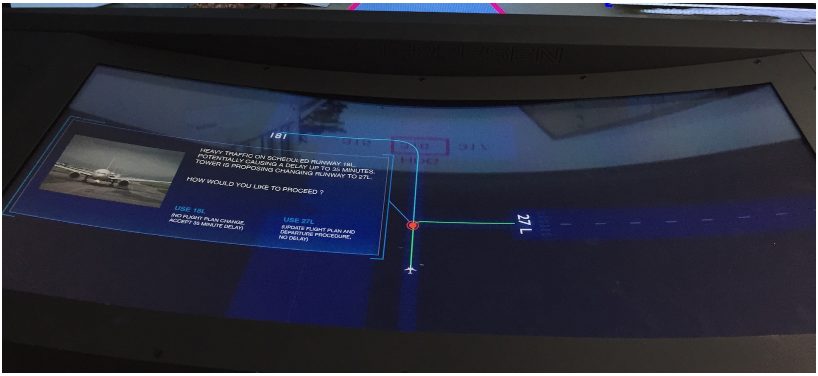


Figure 5.11.: Following the notification from ATC, the MM is presented with detailed information on the mission management display to increase situation awareness and require MM acceptance.

concept and encourage users to think about what the role of an RCO MM *could* be in the future, and not focus on the status quo. Similar to the previous scenario, the three enroute scenarios demonstrate the benefit of gesture to manipulate information across multiple, distant displays.

At predefined locations during the RCO demo, the user receives an alert announcing an issue (maintenance, crew, trajectory). The notification comes in the form of a spoken alert and simultaneous text notification that appears and expands in the upper left corner of the MM's outer view display. An example of such a notification is depicted in Figure 5.12. The typed text notification gives high level information about the alert. The MM can then use a manipulation gesture to pull or swipe the alert widget from the peripheral outer view display down to the mission management display. Detailed information is then displayed about the issue, options for resolution, and any associated downline mission or schedule impacts, measured in time or cost. Upon selection of a resolution, the aircraft's virtual assistant (similar to the previous scenario) coordinates any necessary communication or system updates.

Figure 5.12.: Example of a notification for enroute scenarios

In the maintenance event, the MM must decide how to address a hydraulic valve in the landing gear system which is reporting a steady leak. The scenario demonstrates a reallocation of responsibilities between aircraft operators, dispatchers, and maintenance controllers in an RCO world, giving the MM responsibility for tasks that are typically addressed on the ground. It assumes that the aircraft's decision support systems have access to the aircraft's maintenance plan and history, fleet schedule, airport inventories, and aircraft health-monitoring sensors to predict the impact or risk of each resolution. This event also assumes the aircraft can initiate the communication required to arrange for the necessary parts and personnel to carry out the MM's decision. The detailed information displayed on the mission management display in the maintenance scenario is depicted in Figure 5.13.

In the crew event, the MM must determine how to handle a possible duty time violation of one of the mission's flight attendants. Again, the event demonstrates a reallocation of responsibilities between aircraft and ground, giving the RCO

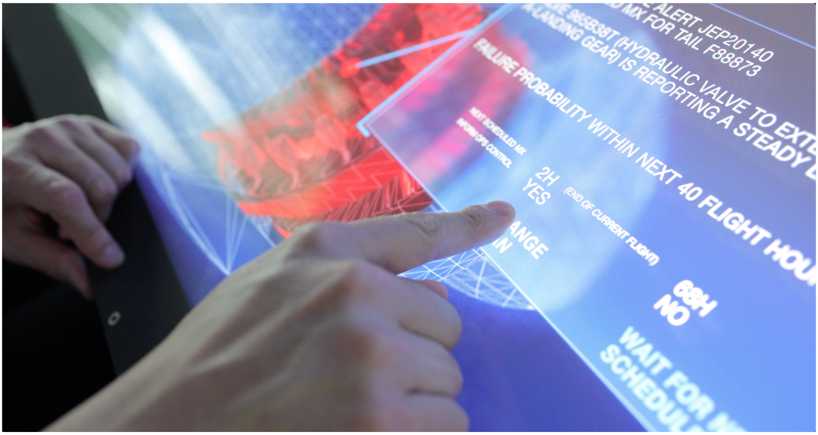


Figure 5.13.: Predictive maintenance scenario requiring mission manager support

MM increased responsibility for the entire day of operations and not just his/her current flight. It assumes the aircraft is aware of crew schedules, the fleet schedule, and any current schedule disruptions in order to predict the downline impacts of a MM decision. The detailed information displayed on the mission management display in the crew scenario is depicted in Figure 5.14.

In the enroute optimization scenario, new wind data affects the planned descent profile upon arrival. Upon acceptance, the virtual assistant updates the FMS automatically, and alerts ATC of the changes. This story suggests that the clearance and communication with ATC is handled entirely through the flight deck's automation, and that the new option has already been validated for performance and feasibility before being presented to the MM. The detailed information displayed on the mission management display in the trajectory optimization scenario is depicted in Figure 5.15.

Every scenario in the RCO demo is reminiscent of a situation that may occur in today's operations, but they are recast in the backdrop of an RCO world, with

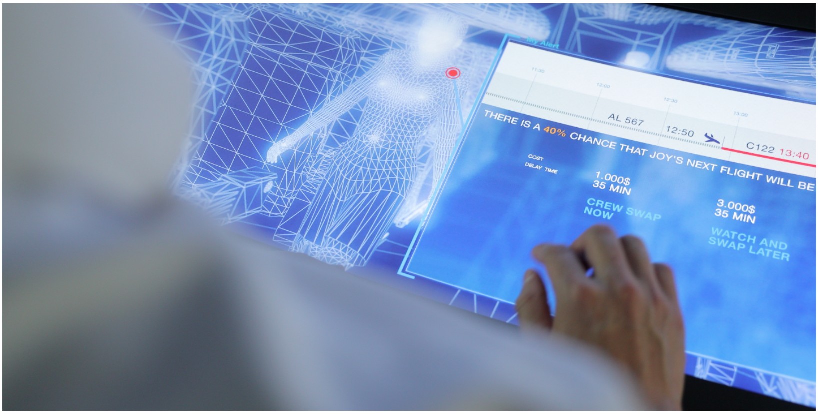


Figure 5.14.: Crew scenario requiring mission manager support to address duty time violation

Figure 5.15.: Trajectory optimization scenario requiring mission manager confirmation of updated descent profile

evolved, multimodal interaction. Each situation is dealt with proactively, and with integrated, real-time information. The MM assumes new responsibilities that a pilot of today may not have. The entire demonstration, while focusing on multimodal interaction, is a strong statement about the changes required to realize RCO, namely, that pilot tasks will shift away from the traditional aviation, navigation, and communication workflows, and towards total mission management.

6. Lessons Learned and Outlook

Gaze, voice, and gesture interaction techniques, and combinations thereof, promise natural, flexible, and intuitive human-machine interaction with increasingly complex and automated systems. In this study, the effects of these six interaction techniques were evaluated to determine their effects on concurrent task management (CTM). An interaction framework was developed, which can help flight deck designers develop suitable interaction for a reduced crew operations (RCO) flight deck. The framework was then used to test the six interaction techniques in 10 concurrent task management experiments. By understanding how interaction technique can affect an operator's ability to manage concurrent task demands, the interaction techniques can be applied effectively. An RCO prototype was was built based on the interaction framework and experiment results. The prototype demonstrates six use cases on an RCO flight deck with effective multimodal interaction. This final chapter provides suggestions for future research based on the lessons learned from the study, and an outlook for the future of RCO interaction research.

6.1. Lessons Learned from the Human Factors Evaluation and Van Demonstrator

The goal of this study was to design guidelines for flight deck designers of a new generation of aircraft; one which provides more natural, flexible, and intuitive human-machine interaction. The taxonomy and surrounding framework presented in Chapter 3 provides these guidelines. This was the first interaction framework built specifically for RCO to be device-independent and applicable in situations of CTM. The taxonomy was tested in a series of 10 human factors experiments. Three main conclusions can be drawn from the evaluation of the taxonomy.

First, this study proved that multimodal interfaces offer flexibility, but have a higher cost of task switching over unimodal interfaces. Multimodal interaction techniques suitably perform a more diverse set of task primitives, providing flexibility in interface design, but require more information processing resources when managing concurrent tasks. As such, care should be taken when introducing them onto an RCO flight deck, particularly with tasks that are time-sensitive. For tasks where time to action is the imperative, unimodal interaction should be used. When time allows, however, multimodal interaction can provide more flexible and intuitive methods for evaluating a problem. Additional human factors experiments are required to further quantify the effects of multimodal interaction on CTM.

The study also proved that the type of interaction technique does indeed have a significant effect on an operator's ability to manage concurrent tasks. Voice interaction techniques interfere more than other interaction techniques in both phonological/vocal type tasks *and* spatial/manual type tasks. Though phonological/vocal type tasks generally present higher competition between information processing resources. The developed taxonomy can be used to minimize this interference. At least one experiment resulted in interference effects contrary to the expectations of the taxonomy. This could imply that interaction technique interference during task-switching is not based on stimulus/code/response mapping alone. Continued research is required to fully understand the interactions of task primitive with interaction technique in more complex workflows.

A third and promising outcome was the observation that neither gender nor previous pilot experience had a significant effect on an operator's ability to use the six interaction techniques. This is presumably because the gesture, gaze, and voice interaction techniques were chosen due to their similarities to the way humans communicate naturally. This result implies that interaction alone will not restrict the candidate pool when identifying mission mangers for an RCO flight deck.

All eight tasks in the 10 concurrent task management experiments were designed according to the multimodal interaction framework from Chapter 3. Additional experiments would refine the interaction taxonomy presented, and expand it to include additional task primitives. Specifically, tasks should be designed with mappings contrary to the developed taxonomy, to further validate it. It is assumed that tasks designed contrary to the taxonomy will increase the interaction-technique-induced costs of concurrent task management beyond those observed in the current study.

The van demonstrator provided qualitative evidence to suggest that the six investigated interaction techniques, when applied appropriately, can facilitate intuitive and flexible interaction that more closely reproduces human-human communication. The tasks were designed according to the established framework, and the interaction techniques chosen to optimize the mental resources available to the human operator and minimize the cost of task-switching. The van demonstrator also proved, however, that while conducting laboratory studies on a diverse group of users can provide valuable system design principles, it cannot provide a replacement for using real-world scenarios to assess the feasibility of each interaction technique.

While the interaction techniques investigated herein may provide more natural communication between human and machine for complex tasks, it can be argued that they will not replace all forms of interaction. Humans will continue to interact with machines via tangible input devices, especially in emergencies and stressful situations. This is not out of necessity, but because touching, holding, and moving physical objects is central to the evolution of tool use in the human species [Hin02]. An RCO flight deck designer should look at all the task primitives necessary for a particular workflow, and compromise on individual interaction technique choices to produce a better overall design.

6.2. Outlook

A reduced crew flight deck is a highly-automated flight deck. Many proponents of RCO believe that implementation is only a question of human factors, and not automation technology [FH18]. Regardless of whether or not the technology is ready, once a highly-automated system is sufficiently capable, the human element remains for fault detection and automation systems management [Bla+14; Ask+17]. And the role of that human element requires re-conceptualizing

[Har07; Fab13]. Having a human operator in any complex system design opens up the operation to the variability, uncertainty, and inconsistency inherent of human information processing [RS16]. Embracing that variability, this dissertation provides guidelines on designing human-machine communication that is intuitive and flexible, in an effort to bring that communication closer to a natural exchange between humans.

Skill- and rule-based tasks, which are methodic and repetitive, will continue to be increasingly replaced by automated systems. As automation replaces more of these lower-level tasks, the information processing resources of a human operator can be directed towards knowledge-based tasks, which require creativity and intelligence. Natural, flexible, and intuitive communication between a reduced crew and a highly-automated aircraft is just one of many changes required to realize reduced crew operations safely and efficiently. The exact tasks have yet to be defined, but this paradigm shift requires task reallocation and a new concept of operation. Theories, models, and frameworks provide the boundary conditions for an RCO world. As the specifics of this world become more defined, the models lose their efficacy in exchange for more concrete standards, checklists, procedures, and operations. But until then, this device-independent, (multi)modal interaction framework provides a baseline for building effective reduced crew operations.

Bibliography

[11518] 115th Congress. FAA Reauthorization Act of 2018. Washington, DC, 2018.

[AB04] Piotr D Adamczyk and Brian P Bailey. "If not now when?: the effects of interruption at different moments within task execution". In: Proceedings of the SIGCHI conference on Human factors in computing systems. Vol. 6. 1. Vienna, Austria: CHI, 2004, pp. 271–278.

[Abb01] Kathy H. Abbott. "Human Factors Engineering and Flight Deck Design". In: The Avionics Handbook. CRC Press LLC, 2001. Chap. 9.

[AK07] Mark H. Ashcraft and Jeremy A. Krause. "Working memory, math performance, and math anxiety". In: Psychonomic Bulletin & Review 14.2 (2007), pp. 243–248.

[AKY15] Marth W. Alibali, Sotaro Kita, and Amanda J. Young. "Gesture and the process of speech production: We think, therefore we gesture". In: Language and Cognitive Processes 15.6 (2015), pp. 593–613.

[AM03] Micah Alpern and Katie Minardo. "Developing a Car Gesture Interface For Use as a Secondary Task". In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. Ft. Lauderdale, Florida: ACM Press, 2003.

[Ant+10] Pavlo Antonenko et al. "Using Electroencephalography to Measure Cognitive Load". In: Educational Phsychology Review 22.4 (2010), pp. 425–438.

[App] Apple Inc. Use Switch Control to navigate your iPhone, iPad, or iPod touch. URL: https://support.apple.com/en-us/HT201370 (visited on 05/01/2019).

[App18] Apple Inc. Siri does more than ever. Even before you ask. 2018. URL: https://www.apple.com/siri/ (visited on 02/25/2018).

[AR92] K. W. Alter and D. M. Regal. Definition of the 2005 Flight Deck Environment. Tech. rep. Hampton, VA: NASA Langley Research Center, 1992.

[ASA78] Richard C. Anderson, Rand J. Spiro, and Mark C. Anderson. "Schemata as scaffolding for the representation of information in connected discourse". In: American Educational Research Journal 15.3 (1978), pp. 433–440.

[Ask+17] James R. Asker et al. Pilotless Airliners-How Realistic? How Soon? 2017. URL: https://aviationweek.com/commercial-aviation/podcast-pilotless-airliner-how-realistic-how-soon (visited on 12/15/2017).

[Bad00] Alan Baddeley. "The episodic buffer: a new component of working memory?" In: Trends in Cognitive Sciences 4.11 (2000), pp. 417–423.

[Bad12] Alan Baddeley. "Working Memory: Theories, Models, and Controversies". In: Annual review of psychology 63 (2012), pp. 1–29.

[Bad83] A.D. Baddeley. "Working Memory". In: Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 302.1110 (1983), pp. 311–324.

[Bai+17] Randall E Bailey et al. "An Assessment of Reduced Crew and Single Pilot Operations in Commercial Transport Aircraft Operations". In: 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC). NASA. Hampton, VA: IEEE, 2017, pp. 1–15.

[Bar+13] K. D. Bartl-Pokorny et al. “Eye Tracking in Basic Research and Clinical Practice”. In: Klinische Neurophysiologie 44.3 (2013), pp. 193–198.

[Bar+95] Andrew K Barrows et al. “Flight Tests of a 3-D Perspective-View Glass-Cockpit Display for General Aviation Using GPS”. In: ION GPS-95. Palm Springs, CA, 1995.

[Bee01] Douglas W Beeks. “Speech Recognition and Synthesis”. In: Digital Avionics Handbook. Ed. by C.R. Spitzer. Williamsburg, VA: CRC Press LLC, 2001. Chap. 8.

[Ber08] Niels Ole Bernsen. “Multimodality Theory”. In: Multimodal User Interfaces. Ed. by Dimitrios Tzovaras. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Chap. 2, pp. 5–29.

[Ber09] Niels Ole Bernsen. NISLab-work on modalities. 2009. URL: `http://www.nislab.dk/nis%7B%5C_%7Dmodalities.php` (visited on 03/23/2018).

[Ber95] Niels Ole Bernsen. “A taxonomy of input modalities”. 1995.

[Bf18] Bundesministerium and für Wirtschaft und Energie. Förderung von Forschungs- und Technologievorhaben im Rahmen des sechsten nationalen zivilen Luftfahrtforschungsprogramms (LuFo VI). 2018. URL: `https://www.dlr.de/pt-lf/Portaldata/50/Resources/dokumente/lufo-vi/Bekanntmachung%7B%5C_%7DLuFo%7B%5C_%7DVI-1.pdf` (visited on 12/05/2018).

[BG06] Adriana Barón and Paul Green. Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving : A Brief Literature Review. Tech. rep. Ann Arbor, Michigan: University of Michigan Transportation Research Institute, 2006, p. 36.

[Bil96] Charles E Billings. Human-Centered Aviation Automation: Principles and Guidelines. Tech. rep. Moffett Field, CA: NASA Ames Research Center, 1996.

[Bit06] Dieter Bittner. “Sprachdialogsysteme: Technik, Einsatz und Perspektiven”. In: LDVZ-Nachrichten 7.1 (2006), pp. 5–9.

[Bla+14] Melissa Blackstun et al. Concept of Operations. Tech. rep. Neu Isenburg, Germany: Jeppesen GmbH, 2014.

[Bol+07] David B. Boles et al. “Predicting Dual-Task Performance With the Multiple Resources Questionnaire (MRQ)”. In: Human Factors 49.1 (2007), pp. 32–45.

[Bol80] Richard A Bolt. “"Put-That-There": Voice and Gesture at the Graphics Interface”. In: SIGGRAPH ’80 Proceedings of the 7th annual conference on Computer graphics and interactive techniques. Vol. 14. 3. Seattle, WA, USA: ACM Computer Graphics, 1980, pp. 262–270.

[Boy14] Guy André Boy. “Requirements for Single Pilot Operations in Commercial Aviation: A First High-Level Cognitive Function Analysis”. In: CEUR Workshop Proceedings. 2014, pp. 227–249.

[Bro54] D.E. Broadbent. “The role of auditory localization in attention and memory span”. In: Journal of Experimental Psychology 47 (1954), pp. 191–196.

[Bux+15] Joerg Buxbaum et al. “Spracherkennung im Flugsicherungsumfeld”. In: Innovation im Fokus 1 (2015), pp. 23–29.

[Bux90] William Buxton. “A Three-State Model of Graphical Input”. In: Human-Computer Interaction - INTERACT ’90. Ed. by Dan Diaper et al. Cambridge, UK: North-Holland, 1990, pp. 449–456.

[CAB84] Gloria L Calhoun, Christopher J Arbak, and Kenneth R Boff. “Eye-Controlled Switching for Crew Station Design”. In: Proceedings of the Human Factors Society, 28th Annual Meeting. Vol. 28. 3. Los Angeles, CA, USA: SAGE Publications Inc, 1984, pp. 258–262.

[CAE17] CAE. Airline Pilot Demand Outlook. Tech. rep. CAE Inc., 2017.

[Cap+04] Alice Caplier et al. “Comparison of 2D and 3D Analysis For Automated Cued Speech Gesture Recognition”. In: 9th Conference Speech and Computer. St. Petersburg, Russia, 2004.

[Cas+10] Sandro Castronovo et al. “Multimodal Dialog in the Car : Combining Speech and Turn-And-Push Dial to Control Comfort Functions”. In: Eleventh Annual Conference of the International Speech Communication Association. Makuhari, Chiba, Japan, 2010.

[Cas+17] Jarrod Castle et al. Flying solo – how far are we down the path towards pilotless planes? Tech. rep. August. UBS Global Research, 2017.

[CB14] Yujia Cao and Karel Benes. Preliminary speech and dialect recognition algorithms. Tech. rep. A-PiMod Applying pilot models for safer aircraft, 2014.

[CCH00] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. “Instant messaging: Effects of Relevance and Timing”. In: People and computers XIV: Proceedings of HCI. Vol. 2. British Computer Society, 2000, pp. 71–76.

[CH11] R. C. Coetzer and G. P. Hancke. “Eye detection for a real-time vehicle driver fatigue monitoring system”. In: IEEE Intelligent Vehicles Symposium. IV. Baden-Baden, Germany: IEEE, 2011, pp. 66–71.

[Cha17] Jérôme Charrier. “Design and Evaluation of a Multimodal Interaction Concept for Use in the Future Flight Deck”. Master Thesis. Technische Universitaet Darmstadt, 2017.

[Che+11] H T Cheng et al. “Contactless gesture recognition system using proximity sensors”. In: IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2011, pp. 149–150.

[CL96] Stéphane Chatty and Patrick Lecoanet. “Pen Computing for Air Traffic Control”. In: SIGCHI Conference on Human Factors in Computing Systems. Vancouver, British Columbia, Canada: ACM, 1996, pp. 87–94.

[CN11] Victor Castaneda and Nassir Navab. Time-of-Flight and Kinect Imaging. Munich, 2011.

[Coh+98] P R Cohen et al. “The efficiency of multimodal interaction: a case study”. In: Fifth International Conference on Spoken Language Processing (ICSLP). 1998.

[Coh88] Jacob Cohen. Statistical Power Analysis for the Behavioral Sciences. 2nd. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988, pp. 79–80, 285–286.

[Cro14] John Croft. Honeywell’s Flight Deck of the Future Laboratory. 2014. URL: `https://aviationweek.com/commercial-aviation/video-honeywells-flight-deck-future-laboratory` (visited on 03/05/2018).

[CSC16] M L Cummings, Alexander Stimpson, and Michael Clamann. “Commercial Pilot Views on Futuristic Single Pilot Operations”. In: AIAA Aerospace Conference. AIAA, 2016.

[Cum14] Missy Cummings. Duke Engineering TALKS: Mary "Missy" Cummings, PhD. 2014. URL: `https://www.youtube.com/watch?v=oB0JsLzxIF0` (visited on 05/28/2019).

[CV05] MM Cerney and JM Vance. Gesture recognition in virtual environments: A review and framework for future development. Tech. rep. Ames, IA, USA: Iowa State University, 2005, pp. 1–25.

[CY13] H R Chennamma and Xiaohui Yuan. “A Survey on Eye-Gaze Tracking Techniques”. In: Indian Journal of Computer Science and Engineering (IJCSE) 4.5 (2013), pp. 388–393.

[Dav+08a] “Physiology and Environment”. In: Fundamentals of Aerospace Medicine. Ed. by Jefferey R. Davis et al. 4th. Philadelphia, PA, USA: Lippincott Williams & Williams, 2008.

[Dav+08b] “Vibration & Acoustics”. In: Fundamentals of Aerospace Medicine. Ed. by Jefferey R. Davis et al. 4th. Philadelphia, PA, USA: Lippincott Williams & Williams, 2008.

[DBB52] K H Davis, R Biddulph, and S Balashek. “Automatic Recognition of Spoken Digits”. In: The Journal of the Acoustical Society of America 24.6 (1952), pp. 637–642.

[Dee16] Jan Oliver Deeg. “Entwicklung eines multimodalen Interaktionskonzeptes für zukünftige Flugdeckoperationen”. Master Thesis. Technische Universität Darmstadt, 2016.

[Def16] Defense Advanced Research Projects Agency. Aircrew Labor In-Cockpit Automation System (ALIAS). 2016. URL: `https://www.darpa.mil/program/aircrew-labor-in-cockpit-automation-system` (visited on 05/09/2018).

[DGI14] Juan David Hincapié-Ramos, Xiang Guo, and Pourang Irani. “The Consumed Endurance Workbench: A Tool to Assess Arm Fatigue during Mid-Air Interactions”. In: Companion Publication on Designing interactive systems. Vancouver, BC, Canada: ACM, 2014, pp. 109–112.

[DK15] Martin Dostál and Pavel Kolčárek. “Multimodal Navigation Display”. In: 34th Digital Avionics Systems Conference (DASC). Prague, Czech Republic: IEEE/AIAA, 2015, pp. 1–11.

[DLO09] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. “Multimodal Interfaces: A Survey of Principles, Models and Frameworks”. In: Human Machine Interaction: Research Results of the MMI Program. Ed. by Denis Lalanne and Jürg Kohlas. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 3–26.

[DT86] Robert D. O’Donnell and F Thomas Eggemeier. “Workload assessment methodology”. In: Handbook of Perception and Human Performance: Cognitive Processes and Performance. Vol. 2. Oxford, England: John Wiley & Sons, Jan. 1986, pp. 1–49.

[Dub12] Stephen J. Dubner. Lying to Ourselves (Ep. 97). 2012. URL: `http://freakonomics.com/2012/10/18/lying-to-ourselves-a-new-marketplace-podcast/` (visited on 03/16/2019).

[Duc08] Andrew T. Duchowski. “Eye Tracking Techniques”. In: Eye tracking methodology: Theory and practice. 2nd. London, UK: Springer-Verlag London Limited, 2008, pp. 1–366.

[EKP12] G Enderle, K Kansy, and G Pfaff. Computer Graphics Programming: GKS - The Graphics Standard. 2nd, Revis. Symbolic computation: Computer graphics. Springer-Verlag, 2012, p. 96.

[Emb05] David Embrey. Understanding Human Behaviour and Error. Tech. rep. Human Reliability Associates Ltd., 2005, pp. 1–10.

[Eul06] Stephen Euler. Grundkurs Spracherkennung. Wiesbaden, Germany: Friedr. Vieweg & Sohn Verlag, 2006, pp. 67–97.

[Fab13] A. Faber. “Single Pilot Commercial Operations”. Master Thesis. Delft University of Technology, 2013.

[Fel12] Klaus Fellbaum. Sprachverarbeitung und Sprachübertragung. 2nd ed. Berlin: Springer Vieweg, 2012, p. 252.

[FH18] Jamie Freed and Tim Hepher. Freight first as jetmakers study single-pilot airplanes. 2018. URL: `https://www.reuters.com/article/us-singapore-airshow-pilotless/freight-first-as-jetmakers-study-single-pilot-airplanes-idUSKBN1FS0SV` (visited on 11/30/2018).

[Fie09] Andy Field. Disovering Statistics Using SPSS 3rd ed. 2009, pp. 157, 341, 389.

[Fil] FilmmakerIQ. The Changing Shape of Cinema: The History of Aspect Ratio. URL: `http://vimeo.com/68830569` (visited on 07/20/2003).

[FO02] Jerry Alan Fails and Dan Olsen. “Light Widgets: Interacting in Every-day Spaces”. In: Proceedings of the 7th international conference on Intelligent user interfaces. San Francisco, CA, USA: ACM, 2002, pp. 63–69.

[Foe10] Gregory Foell. “Stereo Vision: Vergleich verschiedener Algorithmen zur Lösung des Korrespondenzproblems”. Master Thesis. Hamburg University of Applied Sciences, 2010.

[FP16] Mark Fihn and Robert Phares. “Touchscreen Technologies”. In: Handbook of Visual Display Technology. Bristol, UK: Springer-Verlag, 2016, pp. 1431–1489.

[FR11] Michael Friedewald and Oliver Raabe. “Ubiquitous computing: An overview of technology impacts”. In: Telematics and Informatics 28.2 (May 2011), pp. 55–65.

[Fuj12] Fujitsu Laboratories Ltd. Fujitsu Develops Eye Tracking Technology. 2012. URL: `http://www.fujitsu.com/global/about/resources/news/press-releases/2012/1002-02.html` (visited on 03/12/2018).

[Gar13] Lee Garber. “Gestural Technology: Moving Interfaces in a New Direction.” In: Computer 46.10 (2013), pp. 22–25.

[Gen00] General Aviation Manufacturers Association. Recommended Practices and Guidelines for Part 23 Cockpit / Flight Deck Design. Tech. rep. 10. Washington, D.C., USA: General Aviation Manufacturers Association, 2000.

[Gie18] Tim Giese. “Design of an evaluation method to quantify interaction modality effects on concurrent task management for future flight deck operations”. Master Thesis. Technische Universitaet Darmstadt, 2018.

[Gol+01] S Goldin-Meadow et al. “Explaining math: gesturing lightens the load”. In: Psychological Science 12.6 (2001), pp. 516–522.

[GR85] L.P. Goodstein and Jens Rasmussen. “Decision Support in Supervisory Control”. In: 2nd IFAC/IFIP/IFORS/IEA Conference on Analysis, Design and Evaluation of Man-Machine Systems. Vol. 18. 10. Risø National Laboratory. Varese, Italy: Elsevier Ltd., 1985.

[Gra+14] Jonathan Graham et al. “Design of a Single Pilot Cockpit for Airline Operations”. In: Systems and Information Engineering Design Symposium (SIEDS). Charlottesville, Virginia, USA: IEEE, 2014.

[GS12] Wiqas Ghai and Navdeep Singh. “Literature Review on Automatic Speech Recognition”. In: International Journal of Computer Applications 41.8 (2012), pp. 42–50.

[Gup+12] Sidhant Gupta et al. “SoundWave: Using the Doppler Effect to Sense Gestures”. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Austin, TX: ACM, 2012, pp. 1911–1914.

[Haa+04] Jürgen Haas et al. “Spracherkennung und Sprachdialog: Stand der Technik , Einsatzbeispiele und zukünftige Trends”. In: Design & Elektronik Entwicklerforum. Munich, Germany: Sympalog, 2004.

[Har07] Don Harris. “A human-centred design agenda for the development of single crew operated commercial aircraft”. In: Aircraft Engineering and Aerospace Technology 79.5 (2007), pp. 518–526.

[Hau89] Alexander G Hauptmann. “Speech and gestures for graphic image manipulation”. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Austin, TX, United States: ACM, 1989, pp. 241–245.

[Hin02] K Hinckley. “Input technologies and techniques”. In: The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. Ed. by Andrew Sears and Julie A. Jacko. 2nd. Lawrence Erlbaum Associates, 2002, pp. 151–168.

[HJ74] Warren H. Teichner and Marjorie J. Krebs. “Laws of visual choice reaction time”. In: Psychological Review 81.1 (1974), pp. 75–98.

[HJ96] Franz Hatfield and EA Jenkins. “Principles and guidelines for the design of eye/voice interaction dialogs”. In: Proceedings Third Annual Symposium on Human Interaction with Complex Systems (HICS). Dayton, OH, USA: IEEE, 1996, pp. 10–19.

[HJW14] Ken Hinckley, Robert J. K. Jacob, and Colin Ware. “Input/Output Devices and Interaction Techniques (2007)”. In: Computing Handbook. 3rd. Vol. 53. Chapman and Hall, 2014. Chap. 21, pp. 1–79.

[HK93] Taosong He and Aria E. Kaufman. “Virtual input devices for 3D systems”. In: Proceedings of the 4th conference on Visualization. San Jose, CA, USA: IEEE, 1993, pp. 142–148.

[HM79] Bryce O. Hartman and Richard E. McKenzie. Survey of Methods to Assess Workload. Tech. rep. San Antonio, TX, USA: Advisory Group for Aerospace Research and Development, 1979.

[Ho+17] Nhut Ho et al. “Application of human-autonomy teaming to an advanced ground station for reduced crew operations”. In: 36th Digital Avionics Systems Conference (DASC). St. Petersburg, FL, USA: IEEE/AIAA, 2017.

[Hol+17] Matthew J Hollomon et al. Current Status of Gaze Control Research and Technology Literature Review. Tech. rep. Oklahoma City, OK: FAA Civil Aerospace Medical Institute, 2017.

[HRR06] Eliot Hazeltine, Eric Ruthruv, and Roger W Remington. “The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference”. In: Cognitive Psychology 52.4 (2006), pp. 291–345.

[HS88] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Ed. by P.A. Hancock and N. Meshkati. Vol. 52. Amsterdam, Netherlands: North Holland Press, 1988, pp. 239–250.

[HSO05] P. A. Hancock, J. L. Szalma, and T. Oron-Gilad. “Time, Emotion, and the Limits to Human Information Processing”. In: Quantifying Human Information Processing. Ed. by Dr. Dennis McBride and Dr. Dylan Schmorrow. Lexington Books, 2005. Chap. 5, pp. 157–175.

[HW03] William J Horrey and Christopher D Wickens. “Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance”. In: Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Park City, UT, USA, 2003, pp. 1–5.

[IH07] Shamsi T Iqbal and Eric Horvitz. “Disruption and recovery of computing tasks: Field study, analysis and directions”. In: Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’07. 2007, p. 677.

[Ins18] Insitute for Flight Systems and Automatic Control. Future Cockpit Concepts. 2018. URL: https://www.fsr.tu-darmstadt.de/forschung%7B%5C_%7Dund%7B%5C_%7Ddienstleistung/forschungsfelder%7B%5C_%7Dneu/future%7B%5C_%7Dcockpit%7B%5C_%7Dconcepts/future%7B%5C_%7Dcockpit%7B%5C_%7Dconcepts%7B%5C_%7D1.de.jsp (visited on 04/04/2019).

[Int09] International Civil Aviation Organization. "Aeronautical Charts". In: International Standards and Recommended Practices. 11th. Montréal, Quebec, Canada, 2009. Chap. Annex 4.

[Int94] International Organization for Standardization. ISO 7942:1994: Information technology–Computer graphics and image processing–Graphical Kernel System (GKS). 1994.

[Joh+12] Walter W Johnson et al. "Task Allocation for Single Pilot Operations: A Role for the Ground". In: International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero). Brussels, Belgium, 2012, pp. 1–4.

[JR95] Jon E Jonsson and Wendell R Ricks. Cognitive Models of Pilot Categorization and Prioritization of Flight-Deck Information. Tech. rep. Hampton, VA: NASA Langley Research Center, 1995.

[JSM50] Loren F Jones, H J Schrader, and J. N. Marshall. "Pictorial Display in Aircraft Navigation and Landing". In: Proceedings of the Institute of Radio Engineers (Now IEEE) 38.4 (1950), pp. 391–400.

[KA12] Dong-Ik Ko and Gaurav Agarwal. Gesture recognition: Enabling natural interactions with electronics. Tech. rep. Texas Instruments, 2012.

[Kah73] Daniel Kahneman. Attention and Effort. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1973, pp. 178–202.

[Kan+15] Shinjae Kang et al. “Do You See What I See: Towards A Gaze-Based Surroundings Query Processing System”. In: Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. Nottingham, UK: ACM, 2015, pp. 93–100.

[Kar06] Maria Karam. “A framework for research and design of gesture-based human computer interactions”. PhD Thesis. University of Southhampton, 2006.

[Kel13] Leo Kelion. Caterpillar backs eye-tracker to combat driver fatigue. 2013. URL: https://www.bbc.com/news/technology-22640279 (visited on 03/12/2018).

[KH90] G. Kurtenbach and E. Hulteen. “Gestures in Human-Computer Communications”. In: The Art of Human Computer Interface Design. Addison-Wesley Longman Publishing Co., Inc., 1990.

[KI12] Rafiqul Zaman Khan and Noor Adnan Ibraheem. “Hand Gesture Recognition: A Literature Review”. In: International Journal of Artificial Intelligence & Applications (IJAIA) 3.4 (2012).

[Kon+15] Taisiia Kondrashova et al. Concept Design and Implementation of a Futuristic Pilot Mission Display. Tech. rep. Technische Universität Darmstadt, 2015.

[KPM10] Paul Krois, Dino Piccione, and Tom McCloy. “Commentary on NextGen and Aviation Human Factors”. In: Human Factors in Aviation (Second Edition). Ed. by Eduardo Salas and Dan Maurino. 2nd. San Diego: Academic Press, 2010. Chap. 22, pp. 701–710.

[KRC07] M Asif Khawaja, Natalie Ruiz, and Fang Chen. “Potential Speech Features for Cognitive Load Measurement”. In: Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces. ACM, 2007, pp. 57–60.

[KS05] Maria Karam and M.c. Schraefel. A taxonomy of Gestures in Human Computer Interaction. Tech. rep. University of Southampton, 2005.

[KST98] David B Koons, Carlton J Sparrell, and Kristinn Rr. Thorisson. “Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures”. In: Readings in intelligent user interfaces. Ed. by Mark T Maybury and Wolfgang Wahlster. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 53–64.

[KTG14] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. “Bringing Gesture Recognition To All Devices”. In: 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI). Seattle, WA, 2014.

[KW73] John G. Kreifeldt and Thomas Wempe. Pilot Performance During a Simulated Standard Instrument Procedure Turn With and Without a Predictor Display. Tech. rep. Moffett Field, CA: NASA Ames Research Center, 1973.

[KWK07] Arthur F. Kramer, Douglas A. Wiegmann, and Alex Kirlik. Attention: From History to Application. New York, New York: Oxford University Press, Inc., 2007.

[Lab11] Silicon Labs. AN580: Infrared Gesture Sensing. Tech. rep. Austin, TX: Silicon Labs, 2011.

[Lac+14] Joel Lachter et al. “Toward Single Pilot Operations: Developing a Ground Station”. In: Proceedings of the International Conference on Human-Computer Interaction in Aerospace. Santa Clara, CA, USA: ACM, 2014.

[LBC18] Anne-Claire Large, Cedric Bach, and Guillaume Calvet. “CONTACT: A Human Centered Approach of Multimodal Flight Deck Design and Evaluation”. In: Engineering Psychology and Cognitive Ergonomics. Ed. by Don Harris. Las Vegas, NV, USA: Springer International Publishing, 2018, pp. 593–604.

[LBT02] Sören Lenman, Lars Bretzner, and Björn Thuresson. Computer Vision Based Hand Gesture Interfaces for Human-Computer Interaction. Tech. rep. Stockholm, Sweden: Royal Institute of Technology, 2002.

[LDB09] L D Loukopoulos, R K Dismukes, and I Barshi. The Multitasking Myth: Handling Complexity in Real-World Operations. Ashgate Studies in Human Factors for Flight Operations. Ashgate Publishing Company, 2009.

[Lea18a] Leap Motion Inc. Reach into the future of virtual and augmented reality. 2018. URL: https://www.leapmotion.com (visited on 12/30/2018).

[Lea18b] Leap Motion Inc. Shortcuts-Leap Motion Gallery. 2018. URL: https://gallery.leapmotion.com/shortcuts-desktop/ (visited on 02/23/2018).

[Lep+15] J. Leppink et al. "Cognitive load theory: Researching and planning teaching to maximize learning". In: Researching Medical Education. Ed. by Jennifer Cleland and Steven J. Durning. First. John Wiley & Sons, 2015. Chap. 18, pp. 207–218.

[LHL12] Benjamin Langmann, Klaus Hartmann, and Otmar Loffeld. "Depth Camera Technology Comparison and Performance Evaluation". In: Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods. 2012, pp. 438–444.

[Lic10] Alex Lickerman. Why We Need To Know Why. 2010. URL: https://www.psychologytoday.com/us/blog/happiness-in-world/201011/why-we-need-know-why (visited on 03/30/2015).

[Lin08] Robert W Lindeman. CS-525V: Building Effective Worlds, Input Devices. Worcester, England, 2008.

[Liu+12] Yen-Ting Liu et al. SoundSense: 3D Gesture Sensing using Ultrasound on Mobile Devices. Tech. rep. Taipei, Taiwan: Department of Electrical Engineering, National Taiwan University, 2012.

[Loe+13] Sebastian Loehmann et al. "Culturally Independent Gestures for In-Car Interactions". In: Human-Computer Interaction – INTERACT, Proceedings Part III. Vol. LNCS 8119. Cape Town, South Africa: Springer, Berlin, Heidelberg, 2013, pp. 538–545.

[Luf16] Lufthansa Systems. Eye Tracking – Support for operations controllers. 2016. URL: `https://www.lhsystems.com/article/eye-tracking-support-operations-controllers` (visited on 03/12/2018).

[LV15] Jimmie Leppink and Angelique Van Den Heuvel. "The evolution of cognitive load theory and its application to medical education". In: Perspectives on medical education 4.3 (2015), pp. 119–127.

[Mas] Mashable. Glued to our screens: A look back at display technology and how far it's come. URL: `http://mashable.com/2015/01/06/screen-display-tech-ces/%7B%5C%%7D5C%7B%5C#%7DCzRB516fPuq1` (visited on 03/22/2019).

[MB14] Päivi Majaranta and Andreas Bulling. "Eye Tracking and Eye-Based Human–Computer Interaction". In: Advances in Physiological Computing. Ed. by S. H. Fairclough and K. Gilleade. London, UK: Springer-Verlag, 2014. Chap. 3, pp. 39–65.

[McN06] D. McNeill. "Gesture: A Psycholinguistic Approach". In: The Encyclopedia of Language and Linguistics. Ed. by Keith Brown. Elsevier Ltd, 2006, pp. 58–66.

[McN96] David McNeill. "Gestures of the Concrete". In: Hand and Mind: What Gestures Reveal about Thought. New ed. Chicago: University Of Chicago Press, 1996. Chap. 4, pp. 105–133.

[Mhl+16] Alois Mhlanga et al. "Equipment Design and Construction of Racks to Mount Complex Display Constellation and Geometry". Advanced Design Project Report. Technische Universität Darmstadt, 2016.

[Mic] Microchip. GestIC® Technology. URL: `https://www.microchip.com/design-centers/capacitive-touch-sensing/gestic-technology` (visited on 03/13/2019).

[Mic18] Microsoft. Kinect for Windows Sensor Components and Specifications. 2018. URL: `https://msdn.microsoft.com/en-us/library/jj131033.aspx` (visited on 01/08/2018).

[Mil01] Sarah Miller. Literature Review: Workload Measures. Tech. rep. Iowa City: University of Iowa, 2001.

[Mil56] George A. Miller. "The magical number seven, plus or minus two: some limits on our capacity for processing information". In: Psychological Review 63 (1956), pp. 81–97.

[MN80] S. Joy Mountford and R. A. North. "Voice Entry for Reducing Pilot Workload". In: Proceedings of the Human Factros Society-24th Annual Meeting. Minneapolis, MN: The Human Factors Society, Inc., 1980.

[Mon03] Stephen Monsell. "Task switching". In: Trends in Cognitive Sciences 7.3 (2003), pp. 134–140.

[Mor15] Kevin Moritz. "Machbarkeitsstudie zu Gestensteuerungstechnologien für den Einsatz in zukünftigen Flugzeugcockpits". Bachelor Thesis. Technische Universität Darmstadt, 2015.

[MS00] S. Merchant and T. Schnell. "Applying eye tracking as an alternative approach for activation of controls and functions in aircraft". In: 19th Digital Avionics Systems Conference (DASC) Proceedings, vol. 2. Philadelphia, PA, USA: IEEE, 2000.

[NAS10] NASA. Human Integration Design Handbook. Washington, D.C., USA: National Aeronautics and Space Administration (NASA), 2010.

[NC93] Laurence Nigay and Joëlle Coutaz. "A Design Space For Multimodal Systems: Concurrent Processing and Data Fusion". In: Proceedings of the INTERACT Conference on Human Factors in Computing. Amsterdam, Netherlands: ACM, 1993, pp. 172–178.

[Nei15] Stefan Manuel Neis. "Feasibility Analysis of Wearables for Use by Airline Crew". Master Thesis. Technische Universität Darmstadt, 2015.

[NKS18] Stefan Neis, Uwe Klingauf, and Jens Schiefele. “Classification and Review of Conceptual Frameworks for commercial Single Pilot Operations”. In: 37th Digital Avionics Systems Conference (DASC). London, UK: IEEE/AIAA, 2018.

[NMF16] Robert Nesselrath, Mohammad Mehdi Moniri, and Michael Feld. “Combining speech, gaze, and micro-gestures for the multimodal control of in-car functions”. In: 12th International Conference on Intelligent Environments. London, UK: IEEE, 2016, pp. 190–193.

[Nod] Nod Inc. Nod Ring. URL: `http://www.vrdb.com/hardware/nod-ring` (visited on 03/11/2019).

[Nor18a] North Inc. Getting started with your Myo armband. 2018. URL: `https://support.getmyo.com/hc/en-us/articles/203398347-Getting-started-with-your-Myo-armband` (visited on 09/24/2018).

[Nor18b] North Inc. Myo Support. 2018. URL: `support.getmyo.com` (visited on 09/24/2018).

[NS03] Kai Nickel and Rainer Stiefelhagen. “Pointing Gesture Recognition based on 3D-Tracking of Face, Hands and Head Orientation”. In: Proceedings of the 5th international conference on Multimodal interfaces. ACM, 2003, pp. 140–146.

[NS80] Donald A. Norman and Tim Shallice. Attention to Action: Willed and Automatic Control of Behavior. Tech. rep. California Univ., La Jolla. Center for Human Information Processing, 1980.

[NS96] Jan Noyes and Alison Starr. “Use of automatic speech recognition: current and potential applications”. In: Computing & Control Engineering Journal 7.5 (1996), pp. 203–208.

[Obe] Ober Consulting. Ober Consulting. URL: `http://www.ober-consulting.com/main/` (visited on 03/11/2019).

[OC00] Sharon Oviatt and Philip Cohen. “Perceptual user interfaces: multimodal interfaces that process what comes naturally”. In: Communications of the ACM 43.3 (Mar. 2000), pp. 45–53.

[OLE] OLED-info. Samsung OLED company spotlight. URL: http://www.oled-info.com/samsung-oled (visited on 03/22/2017).

[Ou+03] Jiazhi Ou et al. “Gestural Communication over Video Stream: Supporting Multimodal Interaction for Remote Collaborative Physical Tasks”. In: Proceedings of the 5th international conference on Multimodal interfaces. Vancouver, Canada: ACM, 2003, pp. 242–249.

[Ovi06] Sharon Oviatt. “Human-Centered Design Meets Cognitive Load Theory: Designing Interfaces that Help People Think”. In: Proceedings of the 14th ACM international conference on Multimedia. Santa Barbara, CA, USA: ACM, 2006, pp. 871–880.

[Ovi99] Sharon Oviatt. “Ten myths of multimodal interaction”. In: Communications of the ACM 42.11 (1999), pp. 74–81.

[Paa+03] Fred Paas et al. “Cognitive Load Measurement as a Means to Advance Cognitive Load Theory”. In: Educational Psychologist 38.1 (2003), pp. 63–71.

[Par17] Kyunghee Park. Airbus Looking Forward to a Pilotless Future. 2017. (Visited on 04/16/2018).

[Pat15] Thom Patterson. Ready for pilots to fly your airliner by voice commands? 2015. URL: http://edition.cnn.com/travel/article/future-of-travel-cockpit-technology/index.html (visited on 03/05/2018).

[PB05] Alex Poole and Linden J Ball. “Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects”. In: Encyclopedia of Human Computer Interaction. Idea Group Reference, 2005, pp. 211–219.

[PBR07] Carl a. Pickering, Keith J. Burnham, and Michael J. Richardson. “A research study of hand gesture recognition technologies and applications for human vehicle interaction”. In: 3rd Institution of Engineering and Technology Conference on Automotive Electronics. Coventry, U.K., 2007, pp. 1–15.

[Pop16] Rick Popely. Which 2016 Cars Have Hands-Free Liftgates? 2016. URL: https://www.cars.com/articles/which-2016-cars-have-hands-free-liftgates-1420692226921/ (visited on 02/21/2018).

[Pre14] Prashan Premaratne. Human Computer Interaction Using Hand Gestures. Springer Science & Business Media, 2014.

[PRS04] Fred Paas, Alexander Renkl, and John Sweller. "Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture". In: Instructional Science 32 (2004), pp. 1–8.

[Pu+13] Qifan Pu et al. "Whole-Home Gesture Recognition Using Wireless Signals". In: Proceedings of the 19th annual international conference on Mobile computing & networking. Miami, FL, USA: ACM, 2013, pp. 485–486.

[PV94] Fred G. W. C. Paas and Jeroen J .G. Van Merrienboer. "Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach". In: Journal of Educational Psychology 86.1 (1994), pp. 122–133.

[QJ82] T. J. Quinn and R.A. Jauer. Pictorial Formats. Tech. rep. AFWAL-TR-81-3156. McDonnell Douglas Corporation, 1982.

[Que+02] Francis Quek et al. "Multimodal Human Discourse: Gesture and Speech". In: ACM Transactions on Computer-Human Interaction 9.3 (2002), pp. 171–193.

[Qva+17] Pernilla Qvarfordt et al. "Theoretical Foundations of Multimodal Interfaces and Systems". In: The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1. Ed. by Sharon Oviatt et al. 1st. New York, NY, USA: Association for Computing Machinery and Morgan & Claypool Publishers, 2017. Chap. 1.

[Ras83] Jens Rasmussen. "Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models". In: IEEE Transactions on Systems, Man, and Cybernetics 13.3 (1983).

[RCO10] Natalie Ruiz, Fang Chen, and Sharon Oviatt. "Multimodal Input". In: Multimodal Signal Processing for Human Computer Interaction. Academic Press, 2010. Chap. 1, pp. 231–255.

[RF18] Joseph Redmon and Ali Farhadi. YOLOv3: An Incremental Improvement. 2018. URL: `https://pjreddie.com/darknet/yolo/` (visited on 12/30/2018).

[RKC96] Frances H. Rauscher, Robert M. Krauss, and Yihsiu Chen. "Gesture, Speech, and Lexical Access: The Role of Lexical Movements in Speech Production". In: Psychological Science 7.4 (1996), p. 226.

[Ros10] Wayne Rosenkrans. "Hot Spot Intelligence". In: AeroSafety World May (2010), pp. 20–23.

[RS16] Guy Raz and Anil Seth. What Makes Us ... Us. 2016. URL: `https://www.npr.org/programs/ted-radio-hour/485704159/what-makes-us-us` (visited on 10/24/2018).

[Rui11] Natalie Ruiz. "Cognitive Load Measurement in Multimodal Interfaces". PhD Thesis. The University of New South Wales School, 2011.

[Sal+95] Daniel Salber et al. "Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties". In: Proceedings of the Fifth International Conference on Human-Computer Interaction. Lillehammer, Norway: Chapman & Hall Publ., 1995, pp. 115–120.

[Sal05] Dario D. Salvucci. "A multitasking general executive for compound continuous tasks". In: Cognitive Science 29 (2005), pp. 457–492.

[Sch+07] Paul C Schutte et al. The Naturalistic Flight Deck System: An Integrated System Concept for Improved Single-Pilot Operations. Tech. rep. NASA, 2007.

[Sch84] Emanuel A. Schegloff. “On some gestures’ relation to talk”. In: Structures of Social Action. Ed. by Maxwell Atkinson and John Heritage. Cambridge: Cambridge University Press, 1984. Chap. 12, pp. 266–296.

[Sev14] Seventh Framework Programme. All Condition Operations and Innovative Cockpit Infrastructure. 2014. URL: https://cordis.europa.eu/project/rcn/94417/factsheet/en (visited on 02/16/2019).

[Sev16a] Seventh Framework Programme. A-PiMod | Applying Pilot Models for Safer Aircraft. 2016. URL: http://www.apimod.eu (visited on 03/21/2018).

[Sev16b] Seventh Framework Programme. ACROSS|Advanced Cockpit for Reduction of Stress and Workload. 2016. URL: http://www.across-fp7.eu (visited on 05/09/2018).

[SH12] Diane Serban and Vincent Houston. “Fly-By-Voice into NextGen Century” Voice Activated Cockpit Management Systems. 2012.

[Sha17] Hamza Shaban. Boeing will test self-flying planes. 2017. URL: https://www.washingtonpost.com/news/innovations/wp/2017/06/20/boeing-will-test-self-flying-planes/?noredirect=on%7B%5C&%7Dutm%7B%5C_%7Dterm=.a616123f0620 (visited on 05/06/2018).

[Shi+16] R Jay Shively et al. “Application of Human-Autonomy Teaming (HAT) Patterns to Reduced Crew Operations (RCO)”. In: International Conference on Engineering Psychology and Cognitive Ergonomics. Toronto, Canada: Springer, 2016.

[SJ00] Linda E Sibert and Robert J K Jacob. “Evaluation of Eye Gaze Interaction”. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. The Hague, Netherlands: ACM, 2000, pp. 281–288.

[SNS18] Sebastian Michael Sprengart, Stefan Manuel Neis, and Jens Schiefele. “Role of the human operator in future commercial Reduced Crew Operations”. In: 37th Digital Avionics Systems Conference (DASC). London, UK, 2018.

[Sol12] Maria Solorzano. “Mobile Multimodal Interaction : An Investigation and Implementation of Context-dependent Adaptation”. Master Thesis. Vrije Universiteit Brussel, 2012.

[SS01] Emilio Schapira and Rajeev Sharma. “Experimental Evaluation of Vision and Speech based Multimodal Interfaces”. In: Proceedings of the workshop on Perceptive user interfaces. Orlando, FL, USA: ACM, 2001, pp. 1–9.

[ST96] Paul C Schutte and Anna C Trujillo. “Flight Crew Task Management in Non-Normal Situations”. In: Proceedings of the human factors and ergonomics society annual meeting 40.244-248 (1996).

[Sta+13] Neville a Stanton et al. “To twist, roll, stroke or poke? A study of input devices for menu navigation in the cockpit.” In: Ergonomics February 2013 (2013), pp. 37–41.

[STB09] Dario D Salvucci, Niels A Taatgen, and Jelmer P Borst. “Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption”. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston, MA, USA: ACM, 2009, pp. 1819–1828.

[Str+99] Norbert A Streitz et al. “i-LAND: An interactive Landscape for Creativity and Innovation”. In: Proceedings of the SIGCHI conference on Human factors in computing systems. Pittsburgh, PA, USA: ACM, 1999, pp. 120–127.

[Swe88] John Sweller. “Cognitive Load During Problem Solving: Effects on Learning”. In: Cognitive Science 12 (1988), pp. 257–285.

[Swe94] John Sweller. “Cognitive Load Theory, Learning Difficulty, and Instructional Design”. In: Learning and Instruction 4 (1994), pp. 293–312.

[TBL15] Peter Thomas, Pradipta Biswas, and Patrick Langdon. “State-of-the-Art and Future Concepts for Interaction in Aircraft Cockpits”. In: International Conference on Universal Access in Human-Computer Interaction. Ed. by Margherita Antona and Constantine Stephanidis. Los Angeles, CA, USA, 2015, pp. 538–549.

[TE11] E. Theunissen and T. J. Etherington. “Reinventing the past: Avionics systems that didn’t make it”. In: AIAA/IEEE Digital Avionics Systems Conference - Proceedings. Seattle, WA, USA: AIAA/IEEE, 2011, pp. 1–18.

[Tel31] C. W. Telford. “The refractory phase of voluntary and associative responses”. In: Journal of Experimental Psychology 14.1 (1931), pp. 1–36.

[Ter85] W Terry Hewitt. “Towards a Graphics Standard for 3D (GKS-3D)”. In: Advances in computer graphics I. Springer, Berlin, Heidelberg, 1985, pp. 211–226.

[The10] The Telegraph. Ryanair boss says air stewardesses should be allowed to land planes in an emergency. 2010. URL: `https://www.telegraph.co.uk/travel/travelnews/7981643/Ryanair-boss-says-air-stewardesses-should-be-allowed-to-land-planes-in-an-emergency.html` (visited on 05/09/2018).

[The12] Erik Theunissen. “From instruments to decision support: History, trends and (missed) opportunities”. In: 31st Digital Avionics Systems Conference (DASC). Williamsburg, VA, USA: IEEE/AIAA, 2012, pp. 1–15.

[The16] The EyeTribe. The EyeTribe. 2016. URL: `http://theeyetribe.com/theeyetribe.com/about/index.html` (visited on 12/29/2018).

[The19] The Boeing Company. Pilot Outlook: 2019 - 2038. 2019. URL: `https://www.boeing.com/commercial/market/pilot-technician-outlook/` (visited on 04/16/2018).

[Tob11] Tobii Technology AB. Tobii unveils the world's first eye-controlled laptop. 2011. URL: `https://www.tobii.com/group/news-media/press-releases/tobii-unveils-the-worlds-first-eye-controlled-laptop/` (visited on 03/12/2018).

[TR00] Matthew Turk and George G Robertson. "Perceptual User Interfaces". In: Communications of the ACM 43.3 (2000), pp. 32–34.

[Tur14] Matthew Turk. "Multimodal interaction: A review". In: Pattern Recognition Letters 36 (2014), pp. 189–195.

[Uni18] Unity Technologies. PhraseRecognitionSystem. 2018. URL: `https://docs.unity3d.com/ScriptReference/Windows.Speech.PhraseRecognitionSystem.html` (visited on 12/29/2018).

[Van+04] Pascal W. M. Van Gerven et al. "Memory load and the cognitive pupillary response in aging". In: Psychophysiology 41 (2004), pp. 167–174.

[VB05] Daniel Vogel and Ravin Balakrishnan. "Distant freehand pointing and clicking on very large, high resolution displays". In: Proceedings of the 18th annual ACM Symposium on User Interface Software and Technology. ACM, 2005, pp. 33–42.

[Ver97] Patricia May Ververs. Understanding a Pilot's Tasks. Tech. rep. 1. Savoy, IL, USA: University of Illinois Aviation Research Laboratory, 1997.

[Vid+10] Michael A Vidulich et al. "Chapter 7 - Information Processing in Aviation". In: Human Factors in Aviation (Second Edition). Ed. by Eduardo Salas and Dan Maurino. Second Edi. San Diego: Academic Press, 2010, pp. 175–215.

[Voi14] Voice Flight Systems. Pilot Speech Recognition. 2014. URL: `http://www.voiceflight.com/` (visited on 03/06/2018).

[VW81] Michael Vidulich and Christopher D Wickens. Time-Sharing Manual Control and Memory Search: The Joint Effects of Input and Output Modality Competition, Priorities and Control Order. Tech. rep. Urbana-Champaign: Department of Psychology, University of Illinois, 1981.

[Wag+96] Dan Wagner et al. Human Factors Design Guide (HFDG) for Acquisition of Commercial Off-The-Shelf Subsystems, Non-Developmental Items, and Developmental Systems. Tech. rep. Springfield, VA, USA: Federal Aviation Administration, 1996.

[Wat72] Michael J. Watkins. "Locus of the modality effect in free recall". In: Journal of Verbal Learning and Verbal Behavior 11.5 (Oct. 1972), pp. 644–648.

[WB03] Mike Wu and Ravin Balakrishnan. "Multi-Finger and Whole Hand Gestural Interaction Techniques for Multi-User Tabletop Displays". In: Proceedings of the 16th annual ACM symposium on User interface software and technology. Vancouver, BC, Canada: ACM, 2003, pp. 193–202.

[WBL96] David T. Williamson, Timothy P. Barry, and Kristen K. Liggett. Flight Test Results of ITT VRS-1290 In NASA OV-10. Tech. rep. Wright Patterson AFB, OH: Air Force Research Laboratory, 1996, p. 8.

[Wei+13] Frank Weichert et al. "Analysis of the Accuracy and Robustness of the Leap Motion Controller". In: Sensors 13 (2013), pp. 6380–6393.

[Wei91] Mark Weiser. "The Computer for the 21st Century". In: Scientific American 265.3 (1991), pp. 94–104.

[Wex97] Alan Wexelblat. "Research Challenges in Gesture: Open Issues and Unsolved Problems". In: International Workshop on Gesture and Sign Language in Human-Computer Interaction. Bielefeld, Germany: Springer, Berlin, Heidelberg, 1997.

[WG15] Cynthia A. Wolter and Brian F. Gore. A Validated Task Analysis of the Single Pilot Operations Concept. Tech. rep. January. NASA Ames Research Center, 2015.

[WH02] Christopher D. Wickens and William J. Horrey. Driving and Side Task Performance: The Effects of Display Clutter, Separation, and Modality. Tech. rep. September. Savoy, Illinois: University of Illinois at Urbana-Champaign & General Motors Corporation, 2002.

[WHX02] Christopher D. Wickens, John Helleberg, and Xidong Xu. "Pilot maneuver choice and workload in free flight." In: Human Factors 44.2 (2002), pp. 171–188.

[Wic02] Christopher D. Wickens. "Multiple resources and performance prediction". In: Theoretical Issues in Ergonomics Science 3.2 (2002), pp. 159–177.

[Wic08] Christopher D. Wickens. "Multiple Resources and Mental Workload". In: Human Factors 50.3 (2008), pp. 449–455.

[Wil+13] Jennifer Wilson et al. Understanding Single Pilot Operations (SPO) Certification Requirements. Tech. rep. Moffett Field, CA, USA: NASA Ames Research Center, 2013.

[WP06] Robert B Wesson and Gary M Pearson. Voice-Activated Cockpit for General Aviation. Tech. rep. Austin, TX, USA: The Wesson Group, 2006.

[WSV83] Christopher D. Wickens, Diane L. Sandry, and Michael Vidulich. "Compatibility and resource competition between modalities of input, central processing, and output." In: Human factors 25.2 (1983), pp. 227–248.

[YS75] Laurence R. Young and David Sheena. "Survey of eye movement recording methods". In: Behavior Research Methods & Instrumentation 7.5 (1975), pp. 397–429.

[ZF99] Robert Zeleznik and Andrew Forsberg. “UniCam-2D Gestural Camera Controls for 3D Environments”. In: Proceedings of the symposium on Interactive 3D graphics. Atlanta, GA, USA: ACM Press, 1999, pp. 169–173.

[Zim+95] Thomas G. Zimmerman et al. “Applying Electric Field Sensing to Human-computer Interfaces”. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI ’95. New York, NY, USA: ACM, 1995, pp. 280–287.

List of Figures

List of Tables

A. Supplementary Material to the State of the Art

This section provides supplementary information relevant to the state of the art of the various research fields discussed in Chapter 2.

A.1. Missed Opportunities: Flight Deck Evolution

The many modes of the flight management system (FMS), the control display unit (CDU) interface for interacting with the FMS, and the primary flight display (PFD) are nothing more than a digital emulation of its electromechanical counterpart [TE11; Bla+14]. Historically, the adoption of inferior systems, despite the availability of superior concepts, was due to limitations in technology or regulation restrictions [TE11; The12]. In 1950, Jones, Schrader, and Marshall conceptualized a navigational display that combined information into a single, intuitive display that the pilot would otherwise have to integrate across the automatic direction finder, standard range receiver, omnirange receiver, distance-measuring equipment, directional gyro, maps, instrument landing system, weather charts, and collision avoidance system [JSM50]. Due to technological limitations, the display was never implemented [TE11]. Even in modern aircraft, this information is spread across the navigation display (ND), PFD, traffic collision avoidance system (TCAS), electronic flight bag (EFB), and paper charts [TE11; The12].

Predicted path displays were proven as early as 1958 to simplify the control loop for pilots, allowing them to achieve better navigation control when the predicted result of change is also provided [KW73; The12]. But the predicted path concepts were only implementable on programmable electronic displays, which weren't allowed onto the flight deck until the eighties, so the flight director, which could be implemented on existing cathode ray tube (CRT) displays, was created [TE11; The12]. Synthetic vision systems, which project the current flight path in reference to terrain and obstacles, have been proven to be useful pilot support tools, but have yet to be widely adopted by the aviation industry [QJ82; Bar+95; TE11].

In an effort to avoid such missed opportunities in the future, RCO researchers at Technische Universität Darmstadt (TUDA) are taking a clean-slate design approach, learning from the past while addressing technological and regulatory hurdles far in advance [Ins18].

A.2. (Multi)modality: An In-Depth Look

Bernsen defines a modality as consisting of three parts: the physical medium, the human sense, and the information presentation in the medium (see Section 2.3) [Ber08]. Using this definition, the requirements for a multimodal system are quite low (e.g. a screen showing pictures and text is technically multimodal) [Cha17]. The term "multimodal" can be used to describe vastly different interaction paradigms, but the benefits of multimodality, regardless of interface type, are almost always cited as flexibility and naturalness [OC00; KA12; Cha17]. These benefits lead to reduced cognitive workload, fewer errors, and increased operator efficiency [Ovi99; RCO10; Tur14]. To understand why researchers boast such enormous potential of multimodal systems [OC00; Ber08], it helps to understand how modalities can be combined to create interfaces. Two formal models exist: a fusion model, which focuses on the technical aspects of the modalities, and the CARE model (complementarity, assignment, redundancy, equivalence), which describes the type of interaction which is enabled [Ber08; Sol12].

The first model, depicted in Figure A.1, focuses on modality fusion at a technical level and has four defining characteristics [Ber08; DLO09; Tur14]. In **alternate** interfaces, modalities are used serially and separately [DLO09]. In **sequential** interfaces, individual modalities are used in serial such that only one modality is active at any given point in the interaction [Ber08; DLO09] (e.g.

clicking on a text field with a cursor control device (CCD) and then typing text input with the keyboard). In **concurrent** interfaces, modalities are available in parallel, but used separately [Ber08; Tur14] (e.g. flying an aircraft with a yoke and simultaneously lowering the landing gear with the lever). In **synergistic** interfaces, modalities are available in parallel and processed as continuous input [Ber08; Tur14] (Bolt's "Put that there" study, described in Section 2.6.1, is a synergistic interface in the fusion model). The flexibility of synergistic models is presumably the goal when most researchers boast of the benefits of multimodal systems, but each type of fusion interface has presumed benefits over purely unimodal interfaces [Tur14].

		Temporal Use of Modality	
		Serial	*Parallel*
Fusion of Modality	*Integrated*	Alternate	Synergistic
	Independent	Exclusive	Concurrent

Figure A.1.: Author adaptation of the fusion model for modality combination introduced in [NC93]

The CARE model, on the other hand, focuses on the user interaction with a multimodal interface, rather than the technical properties, and also has four defining characteristics. **Complementary** interfaces require multiple modalities to communicate the intended command [Qva+17; Sol12] (Bolt's "Put that there" study is a complementary interface in the CARE model). **Redundant** interfaces imply that multiple modalities, even if used simultaneously, will be processed individually to determine the intended command [Sol12] (e.g. summing the input of Airbus side-sticks). **Assignment** describes interfaces in which only one modality can lead to the intended command [Sal+95] (e.g. the side-sticks in an Airbus aircraft when one pilot has pushed the priority button). **Equivalent** interfaces mean multiple modalities can communicate the same intended command, but only one will be used at a time [Sal+95; Sol12] (e.g. using the camera, speech, or input field options to perform a Google search). Equivalence and assignment describe the availability of a modality, while complementarity and redundancy describe the temporal combined use of modalities [Qva+17; Sol12]. When researchers boast of the benefits of multimodal systems, they are presumably referring to the naturalness of complementary interfaces being nearer to multimodal human-human communication, and the availability of choice in equivalent interfaces, which capitalize on the strengths of each modality to overcome weaknesses of others [Qva+17].

The benefits of multimodal interfaces are primarily associated with synergistic and complementary interfaces that consider the limitations of both the machine and the user [Tur14; Qva+17]. For the purposes of this study, the term "multimodal" shall be synonymous with interaction systems using two or more modalities that are synergistic and complementary.

A.3. Interaction Technologies

The following three sections discuss the possible technological implementations of gaze, voice, and gesture at the time of the writing of this dissertation.

A.3.1. Gaze Tracking Technologies

The most common gaze tracking methods are video-based, eye-attached, and electrooculography [Mor15]. A brief overview of each technique is given below.

Video-Based

Video-based eye-tracking is the most widely used method in commercial off the shelf (COTS) systems [Mor15]. Comparable to the image-based gesture technologies, video-based eye tracking uses a camera (or multiple cameras) and image processing techniques, usually via software, to analyze image stream(s) frame by frame [CY13]. As with the gesture technologies, video-based systems can be further broken down into tower-based, head-mounted, and remote systems, according to their physical implementation [Bar+13].

Tower-based systems have largely been replaced by the less invasive head-mounted and remote solutions [CY13]. Video-based systems require an external light source, such as visible light or infrared (IR) light so that the camera can capture a scene [Mic18; Lea18a]. Appearance-based methods estimate a user's gaze by tracking changes in the recording of the eye, frame to frame [CY13], analogous to the stereo-vision technique for gesture interaction described in Section A.3.3. Feature-based methods track changes in standard characteristics of the human eye, such as pupil contours, eye corners, the limbus, and cornea reflections, to track changes from frame to frame [CY13], and are analogous to the image-based technique for gesture interaction described in Section A.3.3. The pupil is distinguished easily from the surrounding iris due to its high reflectivity [YS75; CY13]. The ratio of dark iris to light sclera is used to detect the limbus [YS75; CY13]. The corneal reflection is created by a direct light source (usually IR) [YS75; The16]. Eye features used for tracking are depicted in Figure A.2.

Eye-Attached

Eye-attached eye-tracking is the most precise method but is invasive, and generally only used in medical research [Duc08; CY13]. The eye attached technologies are further broken down into search coils and scleral lenses [Mor15]. With search coils, wires embedded in the lenses, and their movement through an artificially generated electromagnetic field, is measured [Duc08]. Scleral lenses have embedded mirrors, which reflect projected light onto light sensitive sensors to convert the incident light to x and y coordinates of the user's gaze [YS75].

Electrooculography

Electrooculography was a popular eye-tracking method 40 years ago when computing power was too scarce for the heavy image-processing algorithms

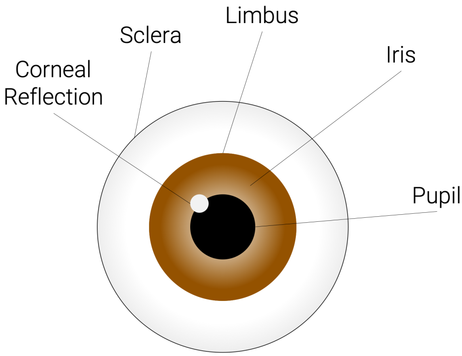


Figure A.2.: Features of an eye used in gaze tracking, adapted from [YS75]

used in the video-based methods, and is still used today in certain applications as a cheap and effective way to track eye movement [Mor15]. Electrooculography can also track the movement of the eye when it is closed, making it a popular technique for tracking during sleep [MB14]. In this method, electrodes are placed around the eyes and as the eye moves, the voltage potential changes are recorded and mapped to a gaze direction [MB14].

Other Methods

In an infrared reflection-based system, the sclera reflects projected infrared light, and the changes from light source and reflection are used to determine eye position changes [Obe]. While this method is not common, it is mentioned here for completeness as fourth category of eye tracking techniques [Mor15].

A.3.2. Voice Recognition Technologies

The process of transforming speech to command input, depicted in Figure A.3, is the same regardless of the technology used [Dee16]. First, an analogue speech signal or sound wave is converted into a digital signal by means of a microphone [Eul06]. Pre-processing of the digital signal consists of filtering out frequencies of little relevance for speech recognition and segmenting speech from non-speech [Eul06]. Feature vectors are then created by extracting speech characteristics over uniform units of time [Fel12]. A phoneme is the smallest distinct, distinguishable unit of sound [Cha17]. Every language has a standardized phoneme set, but the most notable is the International Phonetic Alphabet, which was created to describe all distinct sounds of human language [Dee16]. Feature vectors are then classified from the acoustic signal into the most likely sequence of words [Fel12]. Statistical classification, linear classification, distance classifiers, hidden Markov models, neural networks, dynamic time warping, and acoustic-phonetic classification are used to classify signals, and can be combined to obtain a more reliable result [Fel12].

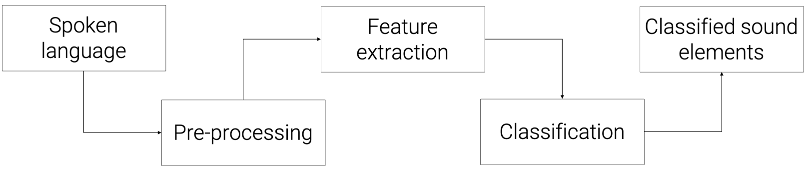


Figure A.3.: The speech recognition process, adapted from [Fel12]

Hidden Markov models are the best method for complex speech recognition systems with large vocabularies, which also forms the basis for continuous speech recognition [Haa+04; Eul06]. The classified feature vectors are then compared with a phoneme dictionary to form hypotheses about which word or words were spoken and their probabilities [Haa+04; Eul06]. The smaller the dictionary, the higher the speech recognition accuracy [Bee01; WP06]. Language models can also be applied to increase automatic speech recognition (ASR) accuracy by defining word likelihoods [Dee16]. Complex, and therefore computationally heavy, language models and other contextual information (e.g. gesture or gaze) can also be used in post-processing, to increase certainty in the classified hypotheses [Eul06; Fel12].

Voice control, in comparison to speech recognition, elicits a corresponding system response from speech input, which distinguishes it from a pure dictation system, where the end of the processing chain in Figure A.3 is already the desired end result [Dee16]. ASR systems are usually classified according to accuracy (word error rates) and speed of successful recognition [GS12]. Performance criteria to determine the complexity and quality required of a voice control system to execute a task are described below.

Speaker dependency: A speaker dependent system will have higher accuracy because it is trained to the personal language skills of a specific user, but speaker-independent systems allow general use and usually require less training [Bee01; Haa+04].

Type of speech: Discrete speech systems require pauses between spoken words, which increases recognition accuracy, but is an unnatural way of speaking [Bee01; Haa+04]. Continuous speech systems are harder to implement, and speaking faster degrades accuracy, but provide the user with a more natural form of interaction [Bee01; Haa+04].

Vocabulary: Vocabulary is the dictionary of all words available to a speech recognizer and has a direct impact on the processing requirements and accuracy of a system [Bee01; Haa+04]. A voice input system for pilots could be limited to a small to medium (100 - 1000) word vocabulary, thanks to the standardization of aviation domain specific language [Bee01; WP06].

Grammar complexity: Consistent grammar structure can increase the likelihood that words will be predicted accurately from the application context [Bee01; Haa+04]. Grammar rules have a high impact on the ASR accuracy and can be augmented to allow for varying word order and omitted words [Bee01; Haa+04].

Input medium: Background noise can be mistaken for phoneme models and analyzed as part of the speech input, negatively affecting the accuracy of an ASR system [Bee01; Haa+04]. Technical solutions, such as noise-cancelling microphones, push-to-talk input, and training the system in real-world conditions, can decrease the detriment of background noise [Bee01; Haa+04].. The placement of the microphone closer to the user can also increase recognition accuracy [Dee16].

The above criteria can be combined to compensate for individual feature deficiencies [Haa+04].

A.3.3. Gesture Recognition Technologies

For the purposes of this research, only contactless or touch-less technologies are considered. Devices that need to be picked up, or otherwise restrict hand and arm movements, are not considered as they add an extra step to the interaction workflow, such that the pilot must pick up the device before being able to interact with it. This means that motion controllers, such as Oculus Touch, are not investigated as viable solutions for gesture recognition in a flight deck setting.

The following provides an overview of common gesture recognition technologies, broken up into the three main categories [Mor15].

Image-Based

Image based gesture recognition uses a camera (or multiple cameras) and image processing techniques, usually via software, to analyze image stream(s) frame by frame [KI12]. The differences in two images captured in two separate frames are detected as gestures [KI12]. Regardless of image-based technology, which is further broken down into four categories, the software that interprets the delivered images has a decisive influence on the quality of the gesture recognition [Mor15].

Two-dimensional gesture recognition is the simplest and cheapest method to implement, as it only requires a single camera and no additional hardware [Cap+04]. It is a promising technique for modern day mobile devices as cameras are ubiquitous on smartphones, tablets, and laptops [Mor15]. In two-dimensional gesture recognition, the camera delivers a color video-stream to the interpretation software [Cap+04]. Depth cannot be determined using two-dimensional recognition, and the quality of the result depends greatly on a contrast between the color of the user's skin and the surroundings [Cap+04]. Starting with their iOS 7 release, Apple developed an accessibility feature called Switch Control which, when activated, uses the front camera to detect user head movements to the left and right [Mor15; App]. The head movements can be programmed for different functionality, but the feature allows users with limited mobility to navigate their device [App].

Stereo vision gesture recognition uses two or more cameras calibrated and slightly offset from one another, and works similarly to the human eye [Mor15]. Software analyzes the two image streams and a pixel in the first is matched with its corresponding twin in the second image [Foe10]. Once the images have been matched, the distance to the camera is determined through triangulation, and a three-dimensional scene is built [Foe10]. The stereo vision technique is computationally heavy and highly dependent on image resolution and lighting situation [KA12]. The depth accuracy is in the range of centimeters, and for consumer devices, which generally have small form factors, it can be cost effective [KA12; Mor15]. Leap motion, a popular consumer device for gesture recognition has achieved sub-millimeter accuracy with their technology [Lea18a]. The controller is equipped two IR cameras which view a scene with three IR emitters [Wei+13].

Structured light projects patterns in IR light onto a scene, which is distorted by the scene's geometry [Mor15]. The pattern can be structured (e.g. parallel lines) or pseudo-random (e.g. dots or "speckles"), but pseudo-random is better for the ensuing analysis [LHL12]. An IR sensitive camera sends the image stream to analysis software which compares the distorted pattern with the original pattern [LHL12]. Similar to stereo vision, triangulation is used to determine the 3D coordinates of an object, the accuracy of which can be in the range of micrometers, depending on the distance of the camera to the object and the patterns projected [LHL12; KA12]. Because of its high accuracy, structured light is often used for indoor applications [LHL12; Mic18]. Version one of the Microsoft's Kinect projected three different patterns to detect near (greater than 40cm), middle, and far (less than $4.5\mathrm{m}$) distances [Mic18]. Daylight is the greatest source of interference, with an irradiance as small as $6-7\,\mathrm{W/m^2}$ causing measurement error in the Kinect controller [LHL12]. Sunlight at sea level has an irradiance of around $75\,\mathrm{W/m^2}$ [LHL12].

Time of flight is analogous to a light detection and ranging (LIDAR) system, in which a receiver throws light onto a scene, object distance is calculated from the time it takes for the light to reflect back to the receiver [KA12]. Using pulse modulation, single, high-energy light impulses are emitted, the objects in a scene, reflect the light back, and the absolute time the light impulse needs to travel until absorption back at the emission source can be used to construct a 3D image [LHL12]. Pulse modulation uses continuous, high-intensity light, so background illumination of a scene does not interfere with this method [LHL12]. The objects cause a phase shift in the reflected light, which is proportional to a reflecting surface's distance from the light emission and camera source [LHL12]. Pulse modulation has a larger selection of applicable light sources, but is prone to signal noise, and is usually integrated over time to reduce said noise [CN11]. The integration is computationally heavy and can lead to motion blur when the computation cannot keep up with the frame rate of the camera [CN11; LHL12]. Microsoft's second version of their Kinect gesture control device uses the time of flight technique [Mic18]. Time of flight has a mm to cm resolution, depending on the depth of the scene, but has the largest range of the image-based techniques between $1-40\,\mathrm{m}$ [KA12; Mic18].

Regardless of employed image-based technique, the hardware delivers information about a scene's geometry, but the software interprets the images and determines if and when a gesture was performed [Mor15]. The first step in the software's analysis is the *extraction* of relevant body parts (usually the hand and arms) from the surrounding environment, for which skin color is the primary differentiating criteria [KI12]. The second step is to *extract features* from the identified body parts, such as the visible side of the hand, the number of visible fingers, and the hand's orientation [KI12]. The third and last step is to *classify gestures* [Mor15]. Gestures are identified through the differences in a series of images or frames and compared with a databank of pre-defined gestures which execute a corresponding action [KI12].

Image-based gesture techniques represent the majority of the high-end COTS for gesture recognition [Mor15]. They are accurate, and can detect complex gestures and poses, down to the individual fingers [LHL12; Wei+13; Lea18a]. But a video stream produces an abundance of data that must then be analyzed by computationally heavy algorithms, and the methods are limited to the frame rate of the camera; at a maximum they can update at 60 Hz [Zim+95; CN11; LHL12; Wei+13]. They are also highly sensitive to lighting conditions, and object texture and reflectivity [LHL12; KA12].

Reflection-based

Reflection-based gesture recognition refers to all technologies wherein a gesture is extracted from the reflection of waves in a specific medium [Mor15]. Because the gestures are calculated from scalar values, such as amplitude and frequency shift, and sometimes over multiple sensors, precise acquisition of user intent is limited [Che+11; Pu+13]. Reflection-based gesture techniques are generally not as accurate as image-based techniques, so complex gestures, where individual finger movement is decisive in the performed gesture, are not possible [Che+11; Liu+12; Pu+13], except in Zimmerman et al.'s electric field system [Zim+95]. Reflection-based gesture techniques measure temporal changes in amplitudes or frequencies, which means static poses are nearly impossible to detect [Mor15]. However, they are generally consume less power than the image-based methods, and, with the exception of IR sensors, are robust against diverse lighting conditions [Che+11; Gup+12; Pu+13]. They have longer ranges than the image-based methods and have a wider interaction space, as they don't require careful orientation of a camera [Pu+13; KTG14]. Where coarse gestures are sufficient for the intended interaction, reflection-based systems offer a cheap and computationally simple alternative to higher-end image-based systems [Che+11; Gup+12; Pu+13; Mor15].

As with image-based techniques, reflection-based techniques can be further categorized according to their implementation, each with its own advantages and disadvantages.

IR proximity sensors are combined with IR light emitting diodes (LEDs), and the object to be detected (e.g. a user's hand) reflects the IR light, the intensity of which is measured by the sensors [Lab11; Mor15]. Using the position-based method, the closer an object is to an IR sensor, the more intense the reflected light will be [Lab11]. Changes in position (reflected light intensity) can then be translated into gestures [Lab11]. In the phase-based method, temporal changes in intensity across multiple IRs sensors are used to determine the direction of a gesture [Lab11]. The detection range of an entire hand is only $7 - 10\text{cm}$, but the energy consumption is low, making it a popular technology for gesture control in mobile devices [Che+11; Mor15]. IR proximity sensor detection is limited to rudimentary, whole-hand gestures, such as swiping left and right, e.g. for scrolling or turning pages, and pushing the open hand forward [Lab11; Che+11].

Electromagnetic waves, such as those for transmitting information via mobile devices, television, radio, and wireless networks, are nearly everywhere, and the human body, albeit only slightly, influences these electromagnetic waves as it moves [Pu+13; KTG14]. WiSee uses existing radio waves, primarily from wireless networks, to detect gestures [Pu+13]. As the human body moves through an environment, such as their own home, where these waves are present, it creates minute Doppler shifts and distortions that are captured by the WiSee receiver [Pu+13]. The technique does not require line-of-sight between a user and receiver, because electromagnetic waves can travel through the walls of a home or office [Pu+13; KTG14]. WiSee offers a home-wide gesture recognition solution with a set of nine gestures, preceded by a preamble gesture, and an average accuracy of 94%, but the components require up to 13.8W of power and the calculations are computationally heavy [Pu+13]. Kellogg, Talla, and Gollakota's gesture recognition system, AllSee, consumes orders of magnitude lower power than WiSee by extracting information from existing wireless signals such as TV and radio frequency identification (RFID) transmission [KTG14]. The improved system detects gestures via changes in the wireless signal amplitude, which is computationally simpler than computing Doppler profiles [KTG14]. AllSee offers an always-on solution that can even run on battery-free devices such as power-harvesting sensors and RFID tags, and can recognize up to eight gestures with up to 97% accuracy [KTG14].

Ultrasound can also be used to detect gestures through phase or amplitude shifts, are low power, and have a range of up to six meters [Liu+12]. Liu et al. developed a method in which multiple ultrasound rangefinders are positioned a sufficient distance from one another, e.g. in all corners of a mobile device [Liu+12]. The rangefinders emit ultrasonic waves, which are reflected by an object, e.g. the human body or hand, and the position of the hand, or changes thereof, can be determined by evaluating the amplitude of the reflections at each rangefinder and translated into gestures [Liu+12]. Gupta et al. developed a method that detects the frequency shift of emitted ultrasonic waves emitted through a speaker and detected by a microphone, e.g. the ones built into almost every mobile device [Gup+12]. Gupta et al. achieved an accuracy of over 88%, and because the technique uses ultrasonic waves, interference from noisy environments was less than 5% [Gup+12].

Electric fields are a popular choice in low-powered mobile devices [Gar13; Mic]. They can be either shunted through a human body to ground or use the human body as the emitter of an electric field to stationary receivers, the distortions of which can be used to detect gestures [Zim+95]. In both cases, the sensors are low power (mW range) and pose no danger to the human object or interference with any electronic devices nearby [Zim+95]. Accuracy of these methods depends upon the distance between the electrodes, but can be as low as micrometers, and can be used for localization of people in a room to the precise tracking of a hand [Zim+95]. Microchip offers a chip for less than 5 USD [Mic].

Wearables

The third category of gesture recognition systems include devices that are partly or completely worn on the body [Nei15; Mor15]. Hand-held motion controllers are explicitly excluded because they restrict motion of the hand. Wearables are potentially uncomfortable, can restrict movement, and must be put on before use, and so have widely been replaced by contactless recognition technologies [Pre14; Nei15]. They are primarily used in the professional sector where precise hand and finger gesture detection is required, such as in drone control or virtual reality [Pre14]. Wearable technology varies greatly in terms of weight, precision, and function [Nei15]. Examples of wearable gesture recognition devices include the Nod Ring, which uses skeletal tracking to translate gestures [Nod], and the Myo wristband, which detects electromyography (EMG) muscle movements to implement six hand gestures and uses accelerometers to detect arm movement [Nor18a].

As in the previous two categories, gesture wearables can be further broken down into two main techniques, described in depth in [Pre14].

Passive wearable technologies require a user to wear markers on body parts which will interact with the system, and external devices track the markers as the user gestures [Pre14]. Recorded images of markers are compared, analogous to the image-based gesture recognition techniques described above.

Active wearable gesture technologies are those that are worn on the body and record gestures without the need for an external capture system [Nei15]. Power to the device is provided either via a battery or a cable [Pre14; Nei15]. The most common active approach is to use accelerometers to draw conclusions about a user's gesture [Pre14], but some devices measure electric impulses across muscles or analyze the movements of tendons [Pre14; Nor18b].

B. Supplementary Material to the Human Factors Evaluation

The following appendix is a collection of figures, diagrams, and plots that supplement the human factors evaluation in Chapter 4.

B.1. Images, Plots, and Tables Supplementary to the Human Factors Evaluation

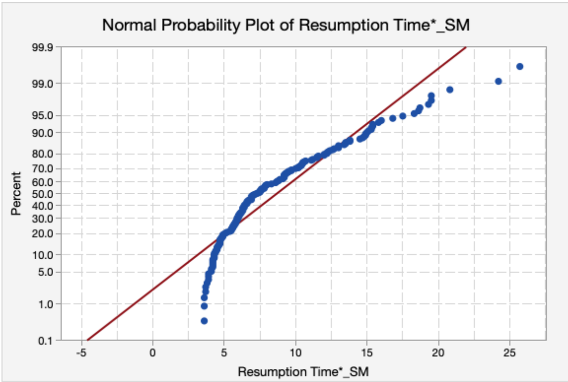


Figure B.1.: The plot shows the normality probability of resumption time* for six S/M task treatments, tested in X1 for H_1.

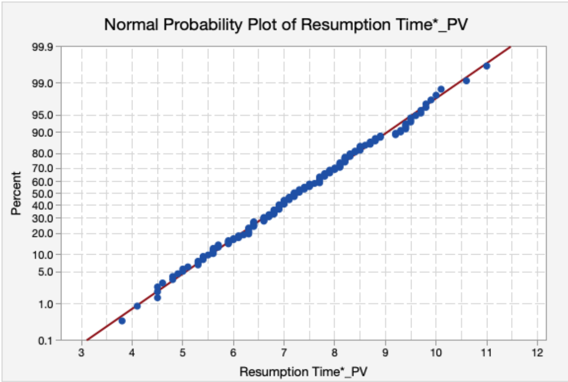


Figure B.2.: The plot shows the normality probability of resumption time* for six P/V task treatments, tested in X2 for H_1.

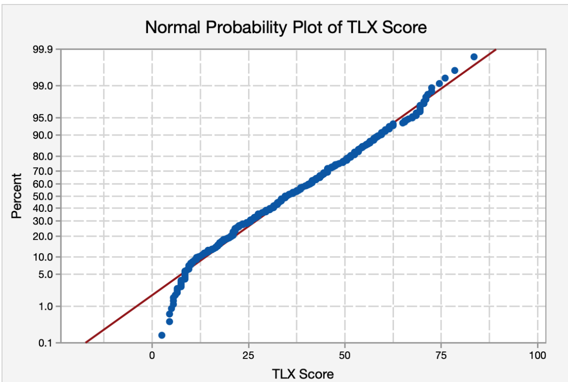


Figure B.3.: The plot shows the normality probability of task load index (TLX) scores for all twelve task treatments (six S/M and six P/V), tested in X1 and X2 for H_3.

B.2. Questionnaires Provided to Experiment Participants

Subject ID:

Pretest Questionnaire

1) Please enter today's date. ________________________________
2) Please enter the current time. ________________________________
3) How old are you? ________________________________
4) Are you… ☐ female ☐ male ☐ other ☐ prefer not to answer
5) Are you wearing vision aids today?
   ☐ Glasses ☐Contact lenses ☐Other ☐None
6) Are you… ☐ left-handed ☐ right-handed
7) Place an "X" on the scale that best represents your current fatigue state.

tired						alert			
1	2	3	4	5	6	7	8	9	10

Very tired Very alert

8) Are you a pilot? ☐ yes ☐ no
   If yes, please answer the following questions, otherwise proceed to #9:
   a) In which field of aviation are you active? (Passenger operations, Cargo, Military, Hobby, etc.)? ________________________________
   b) Do you currently have an (ATPL) Airline Transport Pilot License? ☐ yes ☐ no
      a. Do you currently have a Type Rating? ☐ yes ☐ no

         If so, what model(s)? ________________________________

      b. How many flight hours do you have? ________________________________
   c) When did you fly last? ________________________________

9) How often do you use the following interaction ***technologies***?

   Gesture control (touchless gesture):

   never very often

1	2	3	4	5

   If 2+, to what degree do you use this technology?

   ____

   ____

   Voice control

   never very often

1	2	3	4	5

   If 2+, to what degree do you use this technology?

   ____

   ____

Seite 1

Subject ID:

Eye Tracking

never				very often
1	2	3	4	5

If 2+, to what degree do you use this technology?

____

____

10) The following five questions are a test of the type of learner you are. Please check the box that represents the *best* fit.
   a. When you tackle learning a new subject, you prefer:
      - ☐ reading specialized literature and documentation on the subject
      - ☐ exchanging and interacting with other learners
      - ☐ flow charts, graphs, process diagrams, maps
      - ☐ learning by doing and tinkering
   b. You prefer instructors that design their lectures:
      - ☐ using technical teaching aids, e.g. slides with diagrams and models
      - ☐ by providing curated reading materials
      - ☐ around hands-on examples
      - ☐ by lecturing in detail and clearly structured
   c. When you receive new information, you remember it best when:
      - ☐ you discover it in conversation with other colleagues
      - ☐ you are informed in a letter or Email
      - ☐ it is demonstrated via applicable use cases in a meeting
      - ☐ details are described on paper or a whiteboard
   d. In a seminar, or at a conference, you can understand the content best if:
      - ☐ you can listen to the speaker attentively and uninterrupted
      - ☐ graphics or diagrams accompany the concepts
      - ☐ the speaker goes through examples of how it is implemented
      - ☐ you can ask questions and discuss the content
   e. If you have a busy week at work, the best way for you to relax is:
      - ☐ reading a good book
      - ☐ chatting with friends or listening to music
      - ☐ visiting art museums or creating art
      - ☐ exercising

11) In which if the four learner types would you subjectively categorize yourself?
      - ☐ aural/communicative
      - ☐ visual
      - ☐ reading/writing
      - ☐ physical/kinesthetic
      - ☐ I don't know

Figure B.4.: The questionnaire provided to participants at the beginning of the experiment.

Subject ID:

Subjective Workload of a Treatment

Please put an “X” in on the scale that reflects your gut feeling of the task chain just performed.

MENTAL DEMAND

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex?

Low High

TEMPORAL DEMAND

How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

EFFORT

How hard di you have to work (mentally and physically) to accomplish your level of performance?

Low High

PERFORMANCE

How successful do you think you were in accomplishing the goals of the tasks? How satisfied were you with your performance?

Poor Good

FRUSTRATION

How insecure, discouraged, irritated, stressed, and annoyed vs. secure, gratified, content, relaxed, and complacent did you feel during the task

Low High

Seite 1

Figure B.5.: The questionnaire provided to participants after the successful completion of each of the 12 possible task treatments during the experiment.

Subject ID:

Posttest Questionnaire

1) Please enter the current time. ______________________________

2) Were you satisfied with the environmental conditions of the experiment (volume, temperature, etc.)?

   ☐ yes ☐ no

   If no, please describe:

   ____

   ____

   ____

3) In what areas (both on the flight deck and in everyday life) do you see these interaction techniques (gesture, voice, gaze/eyetracking) as helpful?

   ____

   ____

   ____

   ____

   ____

   ____

4) Other comments, questions, concerns?

   ____

   ____

   ____

   ____

   ____

   ____

Thank you for your participation!

Seite 1

Figure B.6.: The questionnaire provided to participants at the conclusion of the experiment.

C. Supplementary Material to the Development of a Reduced Crew Operations Demonstrator

The following appendix contains supplemental information to the design and development of a mobile van simulator to demonstrate multimodal interaction on an RCO flight deck.

C.1. Ergonomic Model

The form factor described in Section 5.1.1 and depicted in Figure 5.5 is only technically feasible with curved or flexible displays, which are currently still in a development phase, and not readily available to a consumer market [Kon+15]. While several manufacturers do produce curved displays for commercial use (e.g. Samsung, LG, Alpha Displays, Kateeva) using various display technologies, they are not flexible from the user perspective [Mhl+16]. The device maker bends or curves the display into the final, company defined, form factors and the user is not able to bend it further or otherwise change its shape [OLE]. Creating flexible or curved displays requires specialized equipment, and companies often only mass-produce a limited selection of models, with rare exceptions for special research projects [Kon+15; Mhl+16]. Given this fact, and that the display is not the focus of this research, custom-making the ideal form factor described in Section 5.1.1 was well out of the research budget. The construction of the mission display was therefore limited to COTS technologies available at the time of the writing of this dissertation.

One of the main characteristics of the proposed mission display, optimized for interaction and information visibility, is the extra-wide aspect ratio of approximately $3:1$. The most common aspect ratios for modern displays are $16:9$ or $4:3$, and the reason for this is deeply rooted in cinema [Fil]. The current common aspect ratios were set to accommodate the media being produced by the film industry on actual film, and the aspect ratios eventually became standard for other media content producers [Fil]. As more and more media is created digitally, the aspect ratio is no longer restricted by a physical medium, but the result of this history is a consumer market filled with displays defined by a $16:9$ or $4:3$ aspect ratio [Mas].

In 2014, LG released its first 34-inch, ultra-wide, curved monitor, with a $21:9$ aspect ratio, marketed towards gamers and professionals who were using two or more monitors to view their content [Kon+15]. While this aspect ratio is not the same as the mission display, it is the closest of the readily available displays. Currently, COTS monitors are only curved in one direction. LG, Samsung, Dell, and other display manufacturers only offer horizontally curved displays [Mhl+16]. Given the wide aspect ratio of the proposed mission display, and the necessity of curvature to enable optimal interaction, a curved, ultra-wide monitor was identified as the best COTS solution to the design challenge. An LG 34UC97 monitor was chosen for budget and reliability reasons.

The ideal form factor described in Section 5.1.1 can be accomplished via a rear-projection solution, wherein the form is custom bent from an opaque acrylic sheet. A combination of high resolution projectors are used to achieve the necessary aspect ratio. This solution is too large to be implemented in the Viano van platform, however, so was not pursued. A detailed description of this solution can be found in [Kon+15] and [Mhl+16].

The second greatest design challenge to creating a display with the proposed form factor is the ability to make it touch-capable. None of the curved COTS displays were touch capable, so the touch capability had to be retrofitted to the purchased display. Touch screen technology can be divided into four categories, as per [FP16], and described below.

Resistive technology is one of the oldest touch screen technologies and consists of two main parts, a flexible top layer and rigid bottom layer, which are coated with a resistive material and separated by an air gap [FP16]. By touching the two layers together, the two layers close a circuit, resulting in electrical flow, and the coordinates of the touch are then determined by where the change in in voltage occurs [FP16]. Resistive technologies only allow up to two simultaneous touches [Kon+15], and so are not adequate for the current use case.

Capacitive touch screen panels are made of an insulator whose surface has a conductive coating [FP16]. When a human finger, or other conductive material, touches the panel, some of the electric charge transfers from the screen to the finger or capacitive stylus, and the decrease in capacitance is detected by sensors on the screen's edges, which can then determine the touch point [FP16]. Capacitive foils are widely used as a retrofitting solution to make COTS displays touch capable [Kon+15; FP16]. The electronics can be laid such that they do no break upon application to the curved display, but more than one radius of curvature is not recommended, and expensive to prepare [Kon+15; Mhl+16].

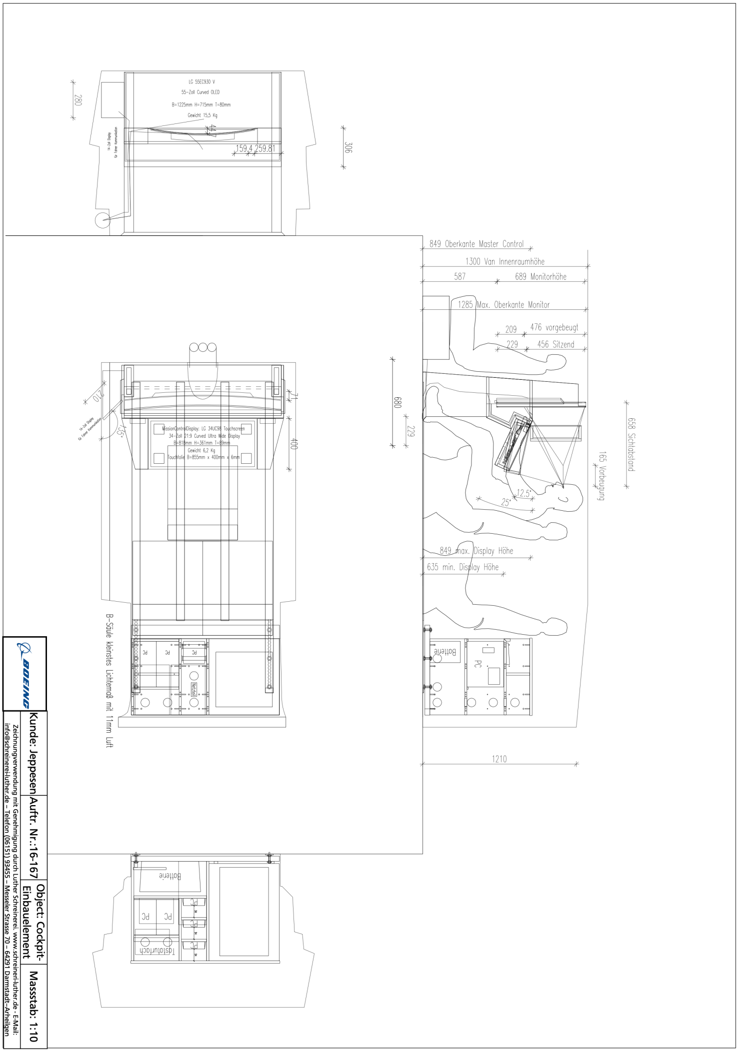
Optical touch screen technology includes scanning IR and camera-based methods [FP16]. The former calculates the position of a touch point via the interruptions of a series of IR light beams; the latter uses interruptions picked up by various cameras on the side of the touch panel [FP16]. The interruptions can be emitted or reflected light, or shadows, that at least two cameras use to triangulate the location of the touch [Kon+15; FP16]. Optical touch screen retrofitting solutions are commercially available, but are expensive to adapt to curved displays, and so is not a viable option [Kon+15; Mhl+16].

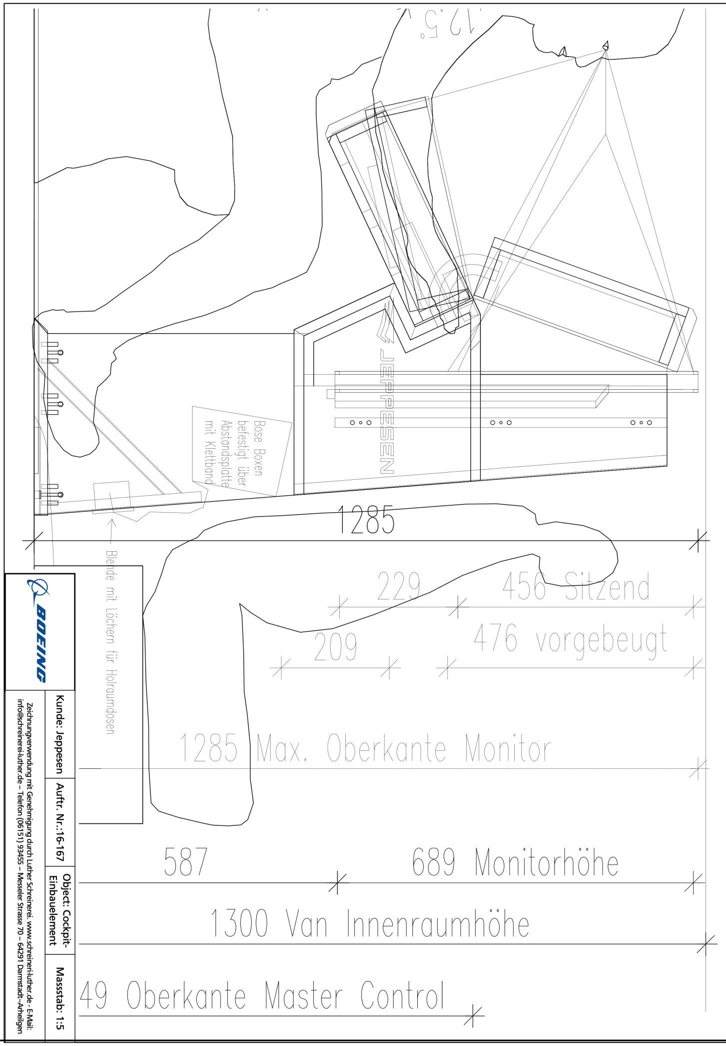
Acoustic touch screens are based on the principle of a touch creating mechanical waves which are propagated along the surface of the panel and detected by sensors along the panel's edges [FP16]. Surface Acoustic Wave, Guided Acoustic Wave, and Acoustic Pulse Recognition are all methods used to create the mechanical waves [FP16]. Acoustic touch screen technologies were not commercially successful due to manufacturing problems [Kon+15; FP16].

Based on the advantages and disadvantages of various COTS touch screen technologies (described in greater detail in [Kon+15], and [Mhl+16]), a Displax Skin Fit, capacitive touch film was custom-created to retrofit an LG 34UC97 monitor with up to 40 simultaneous touch points and arm rejection. This allows the user to rest his/her arms the mission management display without interfering with the touch interaction, reducing arm fatigue.

C.2. Van Demo Construction Parameters

This section provides the dimensions for the assembly construction that was built into the Mercedes Viano van, described in Section 5.1.2 and Section 5.2.


LG 55EC930 V
55-Zoll Curved OLED
B=1225mm H=715mm T=80mm
Gewicht 15,5 Kg
280
306
159,4
259,81
MissionControlDisplay: LG 34UC98 Touchscreen
34-Zoll 21:9 Curved Ultra Wide Display
Gewicht 6,2 Kg
400
680
229
B-Säule kleinstes Lichtemaß mit 11mm Luft
849 Oberkante Master Control
1300 Van Innenraumhöhe
587
689 Monitorhöhe
1285 Max. Oberkante Monitor
209
476 vorgebeugt
229
456 Sitzend
658 Sichtabstand
165 Vorbeugung
12,5°
25°
849 max. Display Höhe
635 min. Display Höhe
Batterie
PC
1210
Tastaturfach
BOEING
Kunde: Jeppesen
Auftr. Nr.:16-167
Object: Cockpit-Einbauelement
Massstab: 1:10
Zeichnungverwendung mit Genehmigung durch Luther Schreinerei. www.schreinerei-luther.de - E-Mail: info@schreinerei-luther.de – Telefon (06151) 93455 – Messeler Strasse 70 – 64291 Darmstadt-Arheilgen


JEPPESEN
Bose Boxen befestigt über Abstandsplatte mit Klettband
1285
Blende mit Löchern für Holraumdosen
229
456 Sitzend
209
476 vorgebeugt
1285 Max. Oberkante Monitor
587
689 Monitorhöhe
1300 Van Innenraumhöhe
49 Oberkante Master Control
Kunde: Jeppesen
Auftr. Nr.:16-167
Object: Cockpit-Einbauelement
Massstab: 1:5
Zeichnungsverwendung mit Genehmigung durch Luther Schreinerei. www.schreinerei-luther.de - E-Mail: info@schreinerei-luther.de – Telefon (06151) 93455 – Messeler Strasse 70 – 64291 Darmstadt-Arheilgen