Communications Between Air Traffic Controllers and Pilots During Simulated Arrivals:

Relation of Closed Loop Communication Deviations to Loss of Separation

by

Christopher Lieber

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved October 2020 by the
Graduate Supervisory Committee:

Nancy J. Cooke, Chair
Robert S. Gutzwiller
Mary Niemczyk

ARIZONA STATE UNIVERSITY

December 2020

ABSTRACT

Communications between air traffic controllers and pilots are critical to national airspace traffic management. Measuring communications in real time made by pilots and air traffic controllers has the potential to predict human error. In this thesis a measure for Deviations from Closed Loop Communications is defined and tested to predict a human error event, Loss of Separation (LOS). Six retired air traffic controllers were recruited and tested in three conditions of varying workload in an Terminal Radar Approach Control Facility (TRACON) arrival radar simulation. Communication transcripts from simulated trials were transcribed and coding schemes for Closed Loop Communication Deviations (CLCD) were applied. Results of the study demonstrated a positive correlation between CLCD and LOS, indicating that CLCD could be a variable used to predict LOS. However, more research is required to determine if CLCD can be used to predict LOS independent of other predictor variables, and if CLCD can be used in a model that considers many different predictor variables to predict LOS.

ACKNOWELDGEMENTS

This research was partially supported by NASA University Leadership Initiative (ULI), Award NNx17AJ86A. I acknowledge the large effort from the team of undergraduate and graduate students that have conducted data collection and recruitment for this experiment, Arizona State University Aviation and Air Traffic Management programs who provided the lab space and the Metacraft simulation environment, the task groups involved on the NASA ULI project, and the NASA ULI industry advisors.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

List of Figures

CHAPTER 1

INTRODUCTION

Safe operations within the National Airspace (NAS) require clear and effective communication between air traffic controllers (ATC) and pilots. ATCs are responsible for the separation of aircraft, issuance of radar safety alerts, and other required services (U.S Department of Transportation, 2017). Pilots are responsible for the final authority as to the safe operation of their respective aircraft (U.S Department of Transportation, 2017). Although there are individual responsibilities relative to either the ATC or pilot, there are also many roles and responsibilities that are shared between them (U.S Department of Transportation, 2017). In the words of the FAA: “The responsibilities of the pilot and the controller intentionally overlap in many areas providing a degree of redundancy. Should one or the other fail in any manner, this overlapping responsibility is expected to compensate, in many cases, for failures that may affect safety.” (“aeronautical information manual”, 2018)

These roles and responsibilities are primarily centered around sharing information through communication between pilots and controllers where a pilot, for example, may primarily communicate to a controller where they are going, what speed and altitude they are at, and what their intentions are in the airspace. While a controller may communicate to a pilot, for example, clearances, altitude assignments, and speed changes. Assuming all the required communications are made and understood correctly by controllers and pilots, traffic within the NAS can be safely managed. However, when communication errors are made, and communication fails there is an increased chance that the shared roles and responsibilities between controller and pilot are not met. When these shared roles and responsibilities are not met, it can result in catastrophic consequences (Hargestram, Lindkvist, Brulin, Jacobsson, & Hultin, 2013; Helmreich, Meritt, Wilhelm, 1999).

To ensure the roles and responsibilities between pilots and controllers are preserved, many of the communications between them are standardized and follow an established pattern of Closed Loop Communication (CLC) (Helmreich, Meritt, Wilhelm, 1999). Adhering to CLC provides many benefits in preserving clear and effective communication between pilots and controllers. These benefits include minimizing communication errors, miscommunications and team breakdowns (Helmreich, Meritt, Wilhelm, 1999). When CLC is not adhered to the possibility of a communication error increases which in turn can result in an unsafe outcome in the NAS (Helmreich, Meritt, Wilhelm, 1999). Based on the relationship between CLC and communication error, and communication error and operational error, deviation or nonadherence to CLC between controllers and pilots could theoretically be used to predict operational errors in the NAS. The purpose of this thesis is to first evaluate the relationship between Closed Loop Communication Deviations (CLCD) and an operational error, Loss of Separation (LOS), and second determine whether the variable CLCD established in this work can be used to predict LOS. The goal of this work is to evaluate the potential for using CLCD as a real time measurement that could be used to predict LOS, and also propose CLCD as a variable that could be used as a feature in more complex models of real time LOS prediction.

CHAPTER 2

BACKGROUND

Air Traffic Controllers (ATCs) are personnel stationed at an air traffic control center and are responsible for efficiently and safely managing air traffic within a zone of control known as an airspace. ATCs manage airspace by using radar systems to monitor traffic within their zone of control and use radio communications to communicate with pilots operating aircraft within their airspace. These communications are critical as they ensure separation exists between all aircraft within the airspace and that aircraft adequately reduce their speed and altitude in order to safely land at the respective airport of that airspace. Typical orders that an ATC may give are speed, altitude, and vector (direction) changes. Additionally, ATCs are responsible for contributing to pilot situation awareness (SA) by communicating changes in the weather, relevant nearby traffic, and other necessary information. National Separation Standards for aircraft are a nationally accepted standard established by the provisions of the International Civil Aviation Organization (ICAO) doc 4444 (Procedures of Air Traffic Management) which dictate that aircraft in class B airspace must maintain a minimal vertical distance of 1000 ft., and a longitudinal distance of 15 minutes apart (ICAO Document 4444, 2007; “aeronautical information manual”, 2018). When surveillance systems are used the minimum separation is 5 nm (nautical miles). At any time when two aircraft are a distance less than these separation standards they are in Loss of Separation (LOS) which is considered a serious Federal safety violation.

Although communications are utilized to administer orders and communicate critical information, they also can be a source of LOS events, and even more serious aviation disasters, such as collisions or crashes. Previous research has found communication errors to be a main contributor to many aviation mishaps, and errors in general even across different communication dependent domains (Hargestram, Lindkvist, Brulin, Jacobsson, & Hultin, 2013). In recognition of the importance of minimizing communication errors, miscommunications, and to avoid team breakdowns, standardized formats for communications were established as outlined by a review on Crew Resource Management (CRM) (Helmreich, Meritt, Wilhelm, 1999). CRM is a series of training procedures used to improve aviation safety which emphasizes communications, decision making and error reduction in the National Airspace (NAS) (Hargestram, Lindkvist, Brulin, Jacobsson, & Hultin, 2013; Helmreich, Meritt, Wilhelm, 1999).

Closed Loop Communications

In air traffic control, an ATC communicates with pilots using closed loop communications (CLC) which is a core component to CRM (Hargestram, Lindkvist, Brulin, Jacobsson, & Hultin, 2013). CLC is a strategy for communication that is used a standardized format for communications, emphasizing three stages of communication: (1) A sender transmits a message, labelled a callout (2) the intended receiver accepts and verbally acknowledges the reception of the message from the sender by reading back the received message (3) and then the sender verifies that the message was received as intended, thus closing the loop (Salas, Wilson, & Murphy, 2008). This model can be seen by referencing Figure 1. An example of a closed loop communication in air traffic operations is as follows:

Delta 123: *"Phoenix Approach this is Delta 123 we are arriving on the SUNNS 8 with information Tango at altitude 12,000, looking for approach to Sky Harbor"*

Phoenix Approach: *"Delta 123, Phoenix Approach, Roger descend and maintain 10,000, expect ILS runway two-five left."*

Delta 123: *"Roger Expect ILS runway two-five left, descend and maintain 10,000 Delta 123"*

In this sample communication *Delta 123* is an aircraft arriving into the Phoenix Sky Harbor airspace managed by *Phoenix Approach ATC*. *Delta 123* fulfills step 1 of CLC by fulfilling the call out, *Phoenix Approach ATC* completes step 2 by completing the readback, while simultaneously administering an order, and finally *Delta 123* verifies that the message has been properly received and assumes he is clear to land at Sky Harbor Airport by reading back the instructions to do so, thus fulfilling step 3 of CLC, closing the loop.

Using CLC as a format for communication yields many benefits and have has been supported in research to have beneficial team outcomes, in the form of clear and accurate communications which improve task clarity and accurate instruction (Kohn, Corrigon, & Donaldson, 2000; Wilson, Salas, Priest, Andrews 2007). The clear and effective communications that come from the use of CLC benefits shared cognition between ATC and pilots (Wilson, Salas, Priest, & Andrews, 2007). In addition, shared cognition can also lead to clear and effective communications (Cooke, 2004; Wilson, Salas, Priest & Andrews, 2007). Shared cognition is a critical component to preventing and mitigating risk in the NAS due to the benefits that come from achieving shared cognition. When shared cognition is achieved teams benefit from better stress management, adaptability and flexibility, better decision making, and productivity (Cooke, 2004). CLC also reduce errors and miscommunications which are essential components to prevent team breakdowns (Burke, Salas, Wilson & Andrews, 2004; Lingard, Espin, Rubin et al., 2005; Lingard, Rehger, Orser et al., 2008; Salas, Wilson & Murphy, 2008). The many benefits discussed here are demonstrated in a study conducted on the performance of air crews during an air crew simulation task measuring for communication sequence which observed high performing teams using more frequent feedback CLC and repeated commands than low performing teams (Bowers, Jentsch, Salas, et al., 1998).

When CLC is not used or when CLC is improperly used the efficiency gains in communications are lost (Hargestam, Lindkvistt, Brulin, Jacobsson & Hultin, 2016). Research on medical teams concluded that when CLC is not followed tasks can become unclear and understanding of instruction becomes less accurate, which in turn can lead to errors in the form of unintentionally causing patient injury (Kohn, Corrigan, & Donaldson, 2000). Similarly, a research study on air crew teams found that deviations from CLC could lead to error (Bowers, Jentsch, Salas, et al., 1998;). When CLCs are non-directed, delayed, or left open, communications are vulnerable to loss of information (Parush, Kramer, Foster-Hunt, et al., 2011). This loss of information is a considerable risk as any loss of information could lead to an infinite range of potential human error.

ATC operations are time sensitive operations and become even more time sensitive under circumstances of heavy traffic. When followed appropriately CLCs are good at facilitating time efficient communications and help to prevent ATC from experiencing communication overload (Hargestram, Lindkvist, Brulin, Jacobsson, & Hultin, 2013). Communication overload can occur due to an overwhelming amount, rate, and or frequency of communications that need to be made to safely manage traffic. When communication overload occurs, communications become delayed and this can be a source of stress, distraction, and or interruption that can influence the completion of any CLC (Gllespie, Chaboyer & Fairweather, 2012: Sevdalis, Undre, McDermott, et al., 2014). In turn the outcomes of communication overload can also be a source which contributes to and or causes communication overload (Sevdalis, Undre, McDermott, et al., 2014; Wheelock, Suliman, Wharton, et al., 2016). Many of the negative outcomes that come from deviations, lack of, or inappropriate from CLC, and the lack of or inappropriate use can cause high workload, and inversely, high workload can cause error prone communications (Gllespie, Chaboyer & Fairweather, 2012: Sevdalis, Undre, McDermott, et al., 2014). To summarize, poor communications leads to high workload, which in turn can create a vicious cycle of increasing workload and error filled communications.

The Current Study

The purpose of the current study is to evaluate how CLC patterns and expected patterns exist in communications between ATCs and pilots and if deviations from regular expected patterns can be used to predict risk, in the form of LOS events, in the NAS. The current theory is based on the previous literature which is that the ideal CLC pattern is having either a pilot: ATC: pilot communication or an ATC: pilot: ATC communication pattern. Any deviation from this pattern of communication would indicate risky, error prone communications which increase the risk for mismanagement of aircraft in the form of LOS. This theory and research question will be evaluated by using transcripts collected from a study being conducted at Arizona State University, which is currently testing and measuring ATC performance in scenarios varying in traffic density and off-nominal events intended to increase ATC workload. ATCs in this experiment are being measured on three scenarios, each with different levels of workload. Data in this experiment collected on subjective workload rating, objective workload rating, number of loss of separation events, communication transcripts, heart rate and facial movement for emotion/physiological state analysis. The current work focused on communication transcripts, LOS events, and subjective workload ratings.

CHAPTER 3

PROJECT OVERVIEW

Research Questions

This study is motivated by three questions. First, what is the relationship between LOS and CLCD? Second, can CLCD be used as a variable to predict LOS. Third, do CLCD become more prevalent in trials of high workload and off-nominal events than they do in low workload trials?

Hypotheses

Hypotheses are presented below. A summary of variables and hypotheses are presented below in Table. 1 and Table. 2.

Hypothesis 1

LOS events and CLCD will be found to have a positive relationship. Trials that have high LOS events will have high corresponding CLCD when compared to trials where LOS events are low where CLCD should also be low. This relationship will then lend itself to be a candidate for real time detection of CLCD and LOS events.

Rationale. CLC serves many benefits as a communication structure that supports error reduction, however when not adhered to communications break down and errors become more prevalent (Burke et al., 2004; Salas et al., 2008; Kohn et al., 2000; Lingard et al., 2005; Lingard et al., 2008). This is largely due to the structure needed to fluently sustain effective and clear communications for complex operations between multiple team members of different roles. In a study specifically on communications between aircrews it was found that deviations from CLC could be tied to errors (Bowers et al., 1998; Siassakos et al., 2011). In a study on medical teams CLC was found to improve task clarity and accurate instruction, failure to follow CLC resulted in patient injuries (Kohn et al., 2000). Standardized communications, like those used in air traffic management, are an important prerequisite to team structure and collaboration, and when these standardized structures are broken efficiency is lost, and teams are more prone to communications that cause errors (Bowers et al., 1998; Burke et al., 2004; Härgestam et al., 2013; Lingard et al., 2005; Lingard et al., 2008; Salas et al., 2008; Siassakos et al., 2011). Because CLC is critical as an error prevention measure it is expected that the previous findings will be supported in the current experiment. LOS events are considered errors and should become more prevalent when patterns of CLC deviate from expected patterns of communication.

Hypothesis 2

The number of CLCD will be more abundant as trial complexity increases, high workload off-nominal trials will be found to have the most CLCD, high workload nominal trials will be found to have the second highest CLCD and baseline trials will be found to have the lowest CLCD.

Rationale. In previous studies high workload has been associated with communication overload, where communication overload has the potential to increase workload and workload has the potential to increase communication overload (Andersen 2010; Woloshynowych et al., 2007). According to researcher's, stress, distractions and interruptions can negatively impact performance in a way that influences task, and CLC completion (Sevdalis et al., 2014; Suliman et al., 2015). Because CLCD can have the potential to cause complexity, it is presumed that it also co-occurs or creates workload, and vice versa. Therefore, it is expected to see the number of CLCD increase as trial complexity increases.

Table. 1
Summary of Variables

Name of Variable	Variable Definition
Loss of Separation (LOS)	LOS is a variable counted when two aircraft in the simulation are calculated as being less than 1,000 feet vertically from each other and/or 5 nautical miles laterally.
Closed Loop Communication Deviation (CLCD)	CLCD is based on a coding scheme established from an expected CLC exchange. A deviation is counted if sequentially a pilot communication follows another pilot communication, or an air traffic controller communication follows an air traffic controller communication.
Density	A categorical variable used to categorize trials of high density and low density. Where high density is when traffic levels were between 10-12 aircraft in the air space at one time, and low is when traffic levels were between 4-5 aircraft.
Workload (Subjective Rating)	Likert Scale Rating between 1 (very low) and 7 (very high)

Note. These are the variables measured and tested in the results section.

Table. 2
Summary of Hypotheses

	Hypothesis	Expectation
1	LOS and CLCD Relationship	Positive Correlation
2	CLCD count and trial type	High workload off-nominal > High workload > Baseline

Note. LOS and CLCD are analyzed using linear regression in the results section. CLCD count is counted and variance is measured using repeated measures ANOVA in the results section.

CHAPTER 4

METHODS

Participants

Six participants were recruited through LinkedIn, recruitment emails, and flyers. Participants were required to have been retired from an air traffic control position. All participants filled out a demographic questionnaire represented in Appendix A and rated their experience working in specific environments detailed in Appendix B. Participants were asked questions about their vision and additional questions represented in Appendix C. Participants reported experience in air traffic control ranging 9 to 40 years $M_{experience}$= 30, $SD_{experience}$= 10.97). All participants reported having working experience in civilian TRACON. Working experience varied when reporting civilian tower experience, military tower experience, military TRACON, civilian center experience, and military center experience. Five out of the six participants reported normal or corrected to normal vision, and all six participants reported normal color vision.

Each participant participated in each trial condition, (baseline, high workload, and high workload off-nominal). Order of each trial condition was randomized in order of participation by each participant to achieve counterbalancing. All participants received the same training before engaging in the experiment. Participation took place at the Arizona State University TRACON Simulation Lab. Each participant was compensated $60 per hour and was debriefed and interviewed at the end of the experiment.

Materials

Metacraft. Metacraft is a TRACON radar simulation used to train prospective ATCs. It incorporates a full range of functions like radar systems used in real world air traffic control settings. Users of the Metacraft system can type commands to manipulate things such as aircraft altitude, speed and heading, among numerous other functions. An ATC in the simulation environment is limited in the commands they can use, meaning they cannot themselves directly adjust any of the previous functions mentioned for any given aircraft. The system for both pilot operators and ATC allows for the respective user to click on aircraft in order to accept handoffs. The system uses 5 separate computers, and at each computer a user signs into their own respective role to operate within the simulation. These roles are set up by arrival route, the arrival routes are ARLIN4/BLYTHE 5, SUNNS8, and HYDRR1. In this experiment, there are three researchers who assigned in and participated as pseudo pilots to their respective assigned traffic route, a researcher proctor who will sign in as a ghost controller and function as final approach, and a participant who will sign in as an ATC. Metacraft hosts its own data log which tracks for LOS events, time aircraft enter the airspace, and distance between aircraft, all while keeping track of the time in the simulation.

Audio and Screen Recordings. Audio recordings were collected by using two methods handheld recording devices and through the screen capture software Open Broadcast Studio. Open Broadcast Studio is a free to download software which can be used to record what happens on one screen on a computer. This screen recording can capture and record events as they happen in real time.

Workload Probe. The workload probe was developed using Microsoft Access and implemented on a Surface Pro Go Tablet. The workload probe works utilizing a modified SPAM method for collecting workload and situation awareness which combines methods previously used (Ligda, et al., 2010; Vu et al., 2012; Durso & Dattel, 2004). This workload probe was designed to administer questions, either situation awareness or workload, by first using a ready button and then a selection of a response, both of which would be timed (Ligda, et al., 2010; Vu et al., 2012; Durso & Dattel, 2004).

Design

The experiment utilizes Metacraft, a mid-fidelity radar simulation, that simulates TRACON airspace operations in the KPHX Quartz airspace with traffic arrival from the West and South East arrival flows. The study is designed using a within subject's design measured across three simulation trials titled Baseline, High Workload Nominal, and High Workload Off-Nominal. Each trial lasts 25 minutes and varies in level of workload by manipulating two variables, traffic density and off-nominal events. Traffic density per trial is as follows: Baseline trials have 4-5 aircraft in the airspace at any given time, while both High Workload Nominal and High Workload Off-nominal have 10-12 aircraft in the air space at any given time. Baseline trials and High Workload Nominal Trials consist of no off-nominal events. High Workload Off-Nominal consists of four off-nominal events which are: moderate turbulence, pilot deviation NORDO (No Radio Aircraft), runway switch, and minimal fuel advisory.

Procedure

Testing took place at Arizona State University in the TRACON simulation lab. Invited participants would arrive, complete relevant demographic and consent forms and then would undergo an orientation and training presentation. Once participants were orientated, they would undergo a practice trial where participants would become familiar with the simulation environment and comfortably achieve competency in using the system. Participants were considered to have achieved competency once they were able to demonstrate accepting a handoff in the Metacraft radar simulation, and verbally demonstrate radio communications for vector, altitude, and speed adjustments using proper phraseology. After achieving competency and comfort in using the simulation participants would undergo the three experimental trials. Participants were made aware during orientation of the following conditions to be maintained throughout each trial, taken from the previous publication by Ligda et al. (2019) with a different data set on the same study:

1. *You must accept all handoffs from center approach. Center will not hold.*
2. *You will only hand off to final approach/KPHX tower. No route modifications that result in aircraft leaving your control*
3. *You will not request/issue command to land at an airport other than the field destination. No alternate airports. You may only hand off to the final approach.*
4. *Keep aircraft in your airspace. No handoffs (except to 120.9 sector) and no point outs.*
5. *You must not declare emergencies.*

In each trial, pseudo pilots would interact with the participant and follow the commands administered by the test participant ATC. Each pseudo pilot was trained in pilot communications, terminology and phraseology and would interact with the ATC as such.

To increase workload, ATCs were instructed that they were unable to declare emergencies, direct aircraft out of the controlled airspace, and were only allowed to hand off to final approach. This meant that once an ATC had an aircraft enter the controlled air space the ATC had to maintain control of them until they were cleared for final approach.

ATC performance during each scenario was measured by counting the number of LOS events that occurred between aircraft in each testing condition which was tracked by the Metacraft software. During each trial ATCs were measured using workload and situation awareness probes, facial recognition software, heart rate, recorded audio transmissions, and screen recordings. Audio recordings collected were transcribed and coded using an established coding scheme to analyze for deviations from expected CLC patterns. An expected pattern for this coding scheme was either ATC: pilot: ATC or pilot: ATC: pilot. An example of a deviation would be pilot: pilot: ATC.

Measures

Several different measures were collected throughout the course of this study including verbal and non-content verbal analysis, biometric data, head position, facial recognition, heart rate, Metacraft data logs and subjective and objective workload ratings. In this thesis however, only several of these measures are examined in this thesis. The measures included in this thesis include non-content verbal analysis measuring for CLC patterns and CLCD, content analysis of transcripts, workload ratings and Metacraft data logs for measuring Density and LOS.

Closed Loop Communication and Closed Loop Communication Deviation Analysis

CLCs were coded using a binary coding system that was applied to detect CLCD from patterns of communication between pilots and ATC. An expected pattern of CLC was defined as one where any communication following a pilot would be an ATC's and vice versa. Any time a pilot's communication would follow another pilot's communication or any time an ATC communication was followed by another ATC communication was considered a CLCD. Normal CLC patterns were coded as 0 in the transcripts, while CLCD was coded as 1. This was done to count the number of CLCD's within a given transcript. Coding using this coding scheme had a potential for detecting various CLCD's categorized in Table. 1. found in the results section of this thesis.

Density

Density is a measure of the number of aircraft in the ATC's controlled airspace at one time. Each condition had its own respective average Density which was categorized as being of low or high density. Low Density was defined as the airspace having an average aircraft count between 4 and 5 aircraft in the airspace at one time. High Density was defined as the airspace having an average aircraft count between 10-12 aircraft. Using these definitions of Low and High Density, a binary categorical scheme was applied to define a trial as being of Low or High Density. Because of experimental design, baseline conditions were categorized as being Low Density and High Workload and High Workload Off-nominal conditions were categorized as being High Density.

LOS

LOS was tracked by the Metacraft system and calculated aircraft as being in LOS based on a measure of vertical and horizontal distance of each aircraft from one another. If two aircraft were less than 1000 feet vertically and/or 5 nm laterally from one another, the pair of aircraft were in LOS. The Metacraft system performs these calculations at a 5 second interval rate. These measures were adapted to the sampling rate of each communication in the transcript using a binary coding scheme which counted whether two aircraft were in LOS at the time of each communication.

Workload Ratings

Subjective workload ratings on a Likert 7-point scale were collected for each participant were collected 3 times during each condition for each participant, once at the 0-minute mark, once at the 9-minute mark and once at the 18-minute mark of each trial. The experiment administered these using a probe that would present these ratings at each time interval, and participants had 60 seconds to respond. These ratings were averaged for each participant and were used in this thesis to validate the level of workload participants were intended to experience in each condition. The experiment was designed for participants to experience low workload in baseline trials, higher workload in high workload conditions, and even higher workload in high workload off-nominal conditions.

Content Analysis of Transcripts

Transcripts were analyzed based on the content of each transmission in order to categorize the different CLCD's and report on the types of CLCD's were observed in each one of the trials. Categories of CLCD are fully defined in Table. 8 found in the results section.

CHAPTER 5

RESULTS

Descriptive Statistics for variables LOS and CLCD are first presented followed by simple linear regression between CLCD and LOS, then Multiple regression between CLCD and Density to predict LOS is presented, followed by a Pearson correlation Table that shows the relationship between each of the considered variables (LOS, CLCD and Density), then a repeated measures one way ANOVA comparing CLCD and average workload ratings with pairwise comparisons, and finally a summary of the types of CLCD observed and corresponding samples of text from the experiment.

Table. 3
Descriptive Statistics by Trial Type

Baseline (N=6)	N	Min	Max	Mean	Std. Deviation
LOS	6	11	135	41.67	48
CLCD	6	9	16	12.33	2.42
High Workload (N=6)					
LOS	6	116	223	155.83	47
CLCD	6	16	34	26.67	7.44
High Workload Off-Nominal (N=6)					
LOS	6	57	159	110.17	39.08
CLCD	6	19	36	29	6.96
All Trials (N=18)					
LOS	18	11	223	102.56	64.09
CLCD	18	9	36	22.67	9.47

Note. There were 6 participants that were exposed to each condition (trial type), descriptive statistics here were looked at the condition level and then at the all trials level that encompassed the scores of all trials and all participants N=18.

CLCD (Simple Linear Regression)

The initial analysis began with a linear regression analysis along with Pearson correlation to determine the relationship between variables CLCD and LOS. Initially the simple linear regression was run by condition type, but because each simple linear regression had a sample size of n=6, this failed to produce any significant models. Due to insufficient power the simple linear regression and correlations were conducted across the three conditions, thus bringing the sample size to *n*=18. The results of the simple linear regression and correlation for all three conditions resulted in a statistically significant model *b*=.603, *t* (16) =3.024, *p*=<.05 with strong correlation *r*(16)=.603, *p*=<.05. These initial findings supported the hypothesis that CLCD and LOS would be positively correlated.

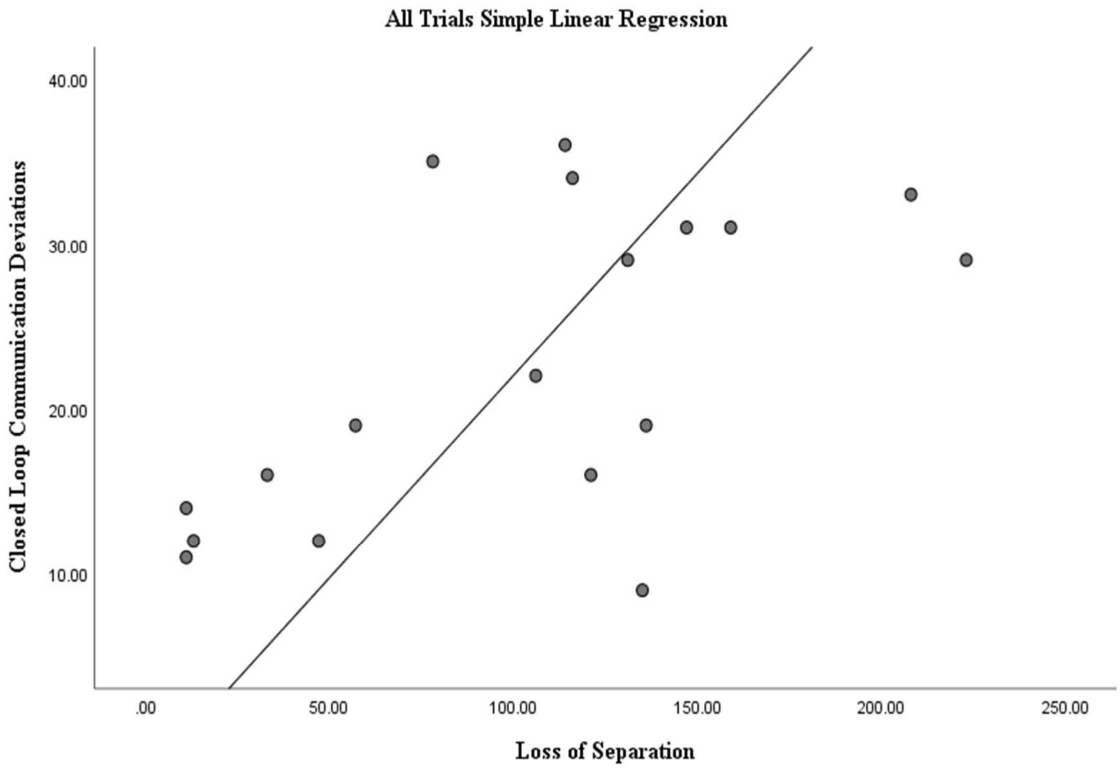
Table. 4

CLCD Simple Linear Regression Coefficients

							95% CI for B			
Trial Type	Model	USCB	SE	SC Beta	t value	Sig.	Lower Bound	Upper Bound	R^2	Adjusted R^2
Baseline	Univariate	-11.773	7.969	-0.594	-1.477	0.214	-33.899	10.354	0.353	0.191
High workload	Univariate	2.382	2.922	0.378	0.815	0.461	-5.729	10.494	0.143	-0.072
High workload Off-Nominal	Univariate	2.339	2.533	0.416	0.916	0.412	-4.75	9.428	0.173	-0.033
Across All Trials	Univariate	4.079	1.349	0.603	3.024	.008*	1.219	6.939	0.364	0.324

Note: Each coefficient was run on the predictor variable CLCD

Figure. 1



Note. r=.603, *p*=.004. N=18 encompassing 6 samples from baseline, 6 from high workload, and 6 from high workload off-nominal. The simple regression was found to be statistically significant at the *p*= <.05 level.

Stepwise Regression to Compare CLCD and LOS

To further evaluate CLCD as a predictor for LOS, two stepwise regressions were run to compare the difference between the ability of CLCD and Density to predict LOS. The first stepwise regression added CLCD to Density to predict LOS, shown in Table. 5, and the second stepwise regression added Density to CLCD to predict LOS, shown in Table. 6. The full model for adding CLCD and Density, regardless of the order in which the predictor variables were added, resulted in the same full model. The full model of Density and CLCD to predict LOS was statistically significant R^2=.697, F(2,15)=7.089, *p*

<.05; adjusted R^2=.486. In the first stepwise regression adding CLCD to Density, the addition of CLCD to Density to predict LOS led to a non-statistically significant decrease in the adjusted R^2 of -.028, $F(1,15)=.233$, $p=.636$. In the second stepwise regression the addition of Density to CLCD to predict LOS led to a non-statistically significant increase in adjusted R^2= .093, $F(1,15)=3.568$, $p=.078$. These results indicate that Density appears to be a more favorable variable for predicting LOS based on the decrease in adjusted R^2 for the full model when CLCD is added to Density, and the increase in adjusted R^2 of the full model when Density is added to CLCD.

Analysis of the coefficients for each variable indicated that both CLCD and Density were nonsignificant contributors to the full model, even though the full model itself was significant. This may be explained by the strong correlation between the predictor variables, represented in Table. 10, which indicates a violation of multicollinearity for the multiple regression models that were built through stepwise regression. This effect could also be attributed to a loss in degrees of freedom due to the addition of variables in the regression model and the already low sample size. These results therefore do not support a finding that CLCD adds anything to the prediction of LOS independent of Density.

Table. 5

Hierarchical Regression Results for Adding CLCD to Density to predict LOS

Variable	□	95% CI for B		SE □	□	R^2	ΔR^2
		LL	UL				
Step 1						.478	.445
Constant	41.667*	.358	82.975	19.486			
Density	91.33**	40.741	141.925	23.865	.691**		
Step 2						.486	.417
Constant	29.406	-39.431	98.243	32.296			
Density	75.925	-9.751	161.600	40.196	.575		
CLCD	.994	-3.392	5.381	2.058	.147		

*p < .05. **p <.01 ***p< .001

Table. 6

Hierarchical Regression Results for Adding Density to CLCD to predict LOS

Variable	□	95% CI for B		SE □	□	R^2	ΔR^2
		LL	UL				
Step 1						.364	.324
Constant	10.101*	-59.862	80.065	33.003			
CLCD	4.079*	1.219	6.939	1.349	.603*		
Step 2							
Constant	29.406	-39.431	98.243	32.296			
CLCD	.994	-3.392	5.381	2.058	.147		
Density	75.925	-9.751	161.600	40.196	.575		

*p < .05. **p <.01 ***p< .001

Pearson Correlation

Pearson correlation was performed along with the stepwise multiple regression between each of the variables: CLCD, Density, and LOS across all conditions for all participants, N=18. Preliminary results showed the relationship to be linear and normally distributed for CLCD and LOS, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. However, Density did not show to be linear and normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). This is because Density is measured as a categorical dichotomous variable. Results of the Pearson Correlation can be shown in Table. 7 which shows a statistically significant relationship between all the considered variables. The results of the Pearson Correlation showed Density having a stronger relationship to LOS than CLCD, indicating that Density has better potential for predicting LOS than CLCD. Additionally, CLCD and Density were found to have a statistically strong relationship indicating that CLCD is likely to increase as Density increases, and vice versa. These results indicate that CLCD and Density are highly likely to occur together. Therefore, the results support hypothesis one which hypothesizes that as CLCD increases so does LOS, however the relationship between CLCD and Density to predict LOS cannot be disentangled.

Table. 7
Pearson correlations for main study variables

	LOS	Density	CLCD
LOS	1	.691*	.603*
Density	.691*	1	.794*
CLCD	.603*	.794*	1

Note. LOS= Loss of Separation, CLCD= Closed Loop Communication Deviation
*=statistically significant at p < .05 level

Workload and CLCD

Participants were measured at three different intervals per trial on a self-report subjective workload rating using a 7-point Likert Scale. Appendix. H shows non-adjusted responses to workload ratings and Appendix I shows adjusted responses to workload ratings. Workload ratings were adjusted to reflect missing responses, where missing values were assumed to be rated as a 7. Missing values in the adjusted workload rating were assumed to be 7 as participants exceeded the maximum amount of allotted time to respond at the time the probe was presented. These assumed values assume that the participant was under high workload and unable to attend to the workload probe. The non-adjusted average workload rating for all participants was 2.76 in baseline trials, 4.33 in high workload trials and 4.75 in high workload off-nominal trials. The Adjusted average workload rating for all participants was 3 in baseline trials, 5.05 in high workload trials, and 5.5 in high workload off-nominal trials. These workload ratings support the condition's intended workload, showing that baseline trials with low traffic density were of low workload relative to high workload and high workload off-nominal conditions.

The results of counts in CLCD increasing as per condition can be found in an appendix F and Appendix G. A one-way repeated measure ANOVA was conducted to determine if there were statistically significant differences in CLCD by condition (baseline, high workload, high workload off-nominal). There were no outliers and the data were normally distributed as assessed by boxplot and by Shapiro-Wilk test (p >.05). Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi2(2) = .145$, $p = .930$. Results of the repeated measures ANOVA indicated CLCD was statistically significantly different for each condition $F(2,10)=22.078$, $p<.0005$. CLCD increased from baseline (M= 12.33, SD=2.42) to high workload (M= 26.67, SD= 7.45) to high workload off-nominal (M=29, SD 6.96). Analysis of Pairwise Comparisons can be shown in Appendix. E. which showed that there was no statistical difference in CLCD between high workload and high workload off-nominal trials. Mean values and their confidence intervals for CLCD by condition can be found in Appendix F.

Deviation Classification and Examples

Content analysis was conducted to categorize CLCD's found in the study and count them. CLCDs were categorized into five different categories: Normal, Step Over, Interruption, Repeat and Error Correction detailed in Table. 8. CLCDs categorized as Interruption, were further analyzed to determine if the transmission that was interrupted was addressed either by the sender or receiver of the message after the interruption. These results indicated that the majority of the CLCD's detected using the prescribed coding scheme were categorized as being normal, though problematic CLCDs still were observed to increase between conditions.

Table. 8
Deviation Classifications

Deviation Type	Definition
Normal	Pattern changes due to normal circumstances. This could be the result of a pilot changing radio channels as advised by the ATC or a pilot concludes a communication with an ATC and a second pilot calls in.
Step Over	A pilot calls the ATC, or an ATC calls out to a pilot during another transmission
Interruption	This happens when a pilot or an ATC start a communication before the loop of the communication finishes
Repeat	A communication that was already transmitted by the sender is transmitted again by the same sender shortly following the first transmission.
Error Correction	A transmission is sent by a sender, and the sender repeats the transmission shortly afterwards correcting wrong information conveyed in the original transmission
*Interruption-Neglect	Interruptions that were not addressed were counted as being neglected and coded under the category Interruption-Neglect.

Table. 9

CLCD Categorical Descriptive Statistics

Condition	Range	Mean	Standard Deviation
Baseline (N=6)			
Normal	9-12	10.83	1.169
Step Over	0-2	.67	.816
Interruption	0-1	.17	.408
Neglect	0-1	.17	.408
Repeat	0-2	.67	.816
Error Correction	0-1	.17	.408
CLCD Adjusted	0-5	1.83	1.941
High Workload (N=6)			
Normal	10-26	17.67	5.820
Step Over	0-3	.83	1.169
Interruption	0-4	1.83	1.472
Neglect	0-3	1.50	1.049
Repeat	2-6	3.00	1.549
Error Correction	0-4	2.00	1.414
CLCD Adjusted	4-15	9.17	4.355
High Workload Off Nominal (N=6)			
Normal	11-25	17.17	5.307
Step Over	0-2	.50	.837
Interruption	1-9	4.17	3.061
Neglect	1-6	3.33	1.966
Repeat	2-7	4.00	1.897
Error Correction	0-2	1.00	.632
CLCD Adjusted	5-18	13.00	4.980

Note. Categories are based of the descriptions made in Table. 8

Table. 10

Deviation Examples from Transcripts

Deviation Type	Sender	Receiver	Transcript	Notes
Normal Deviation	KPHX	ASA 467	"Alaska Four Sixty-Seven descend and maintain ten thousand"	ASA 467 given an instruction by the air traffic controller.
	ASA 467	KPHX	"Dropping down to ten thousand Alaska 467"	ASA 467 confirms the message and completes the communication.
	ASH 994	KPHX	"Phoenix Approach Air Shuttle Nine Ninety Four with you at one two thousand we've got information Tango"	ASH 994 Calls in following the communication exchange between KPHX and ASA 467.
Step Over	Unknown Caller	KPHX	"Phoenix Approach"	Right after a call in is initiated by an unknown sender the ATC cuts him off to begin a transmission with another aircraft.
	KPHX	SWA 9920	"Southwest ninety-nine twenty descend and maintain one zero thousand."	
Interruption/ Interruption Neglect	KPHX	EJA 53	"Execjet fifty-three turn left heading three four zero I got to pull you out of the sequence and make a hole."	ATC gives EJA 53 an order and does not wait for EJA 53 to confirm the message and

	KPHX	AAL 680	"American six eighty descend to four thousand contact approach control one two zero-point niner."	moves on immediately to contacting another aircraft AAL 680. In this instance, the controller neglected to revisit his initial contact with EJA 53 to confirm that the message had been received as intended
Repeat	KPHX	FFT 68	"Frontier Sixty ninety-four turn left heading zero four zero and expect runway seven right."	
	KPHX	FFT 68	"Frontier Sixty ninety-four turn left heading zero four zero and expect runway seven right."	The ATC repeats his transmission after no response from the intended receiver, FFT 68.
Error Correction	KPHX	AAY 417	"Allegiant four seventeen maintain one one thousand traffic alerts right below right behind ya half a mile behind climb to one one thousand three hundred."	
	KPHX	AAY 417	"Correction allegiant four seventeen descend turn right	The ATC makes a correction to

		heading one seven zero please."	his previous transmission, immediately after sending the transmission immediately.

Note. Examples were not taken from any specific trial. Examples were included from multiple participants and conditions. Categories are defined in the already presented in Table 8.

CHAPTER 6

DISCUSSION

This thesis was designed with the purpose of evaluating closed loop communication deviations (CLCD) and its relationship with Loss of Separation (LOS), evaluating CLCD as a variable that could be used to predict LOS and to determine if CLCD's were more prevalent under high workload conditions. Results of the simple linear regression found a statistically significant positive relationship between CLCD and LOS, indicating that CLCD independently has the potential to be used as a variable to predict LOS. This supported Hypothesis 1 which stated that LOS and CLCD will be found to have a positive relationship.

The initial findings of the simple linear regression encouraged further exploration of CLCD as a variable used to predict LOS. This was explored using a stepwise multiple regression which included Density as a variable. The purpose of running the stepwise multiple regression was to determine if CLCD increased prediction for LOS when added to Density. If an increase in the adjusted R squared value was observed, this would have indicated that CLCD as a variable adds to the prediction of LOS apart from Density. However, this effect was not observed. First when running the stepwise multiple regression, the multiple regression model came up as significant, however analysis of the coefficients did not show either Density or CLCD as significant contributors in the multiple regression model to predict LOS. Second, analysis of the regression model when adding CLCD to Density to predict LOS and vice versa indicated that Density had more of an impact on the adjusted R squared value than CLCD. In detail, when CLCD was in the first step, adding Density increased the adjusted R squared value for the full model, while adding CLCD to Density decreased the adjusted R squared value. This effect was interesting as Density appeared to be a more predictive variable for LOS than CLCD, but this was not supported with any statistical significance. However, this observation would have to be further evaluated with a higher sample rate.

Although, the multiple regression was unable to identify the effect CLCD had in the prediction of LOS apart from Density, the results of the Pearson correlation could be used to suggest that CLCD, although not as predictive as Density, could be used instead of Density to predict LOS when Density is unable to be measured. This could be because when CLCD is occurring Density is likely to be high and therefore LOS is also likely to occur. CLCD's high relationship with Density could also be used to detect Density and vice versa. This finding should be explored further as this could suggest that instead of using the variable CLCD to predict LOS, Density could be used as a variable to predict CLCD. This relationship would be worth exploring as Density could then also be used to predict communication errors caused by CLCDs. The analysis of CLCD in this thesis has been done to predict human error in the form of LOS, however the variable could be evaluated with a different perspective as an outcome variable that indicates potential human error related to communication, CLCD, as opposed to predicting a different outcome of human error, LOS.

Analysis of the one repeated measures indicated that CLCD does change as an effect of condition type as it was observed to be significantly lower in baseline trials and higher in high workload trials, however there did not appear to be a significant difference in CLCD count between high workload and high workload off-nominal trials. This result may be explained by the following rationales: first CLCD is highly correlated with Density, the variable that is only different between baseline and high workload and high workload trials, but not different between high workload and high workload off-nominal trials. The density is the same between high workload and high workload off-nominal trials. Second the off-nominal events may have not been substantial enough to cause an observable increase in CLCD. These results partially supported CLCD as a variable that increases as workload increases, though more analysis would be required to evaluate this relationship. Further analysis of this observation could be used to relate workload levels to CLCD; however, this was not evaluated as a part of this thesis. This observation could be explained due to the amount of aircraft and the increased amount of communication required. As Density increases, the ATC is taxed with having to manage more aircraft, keep track of more information, and handle more communications, therefore it is expected that more deviations will occur as the ideal CLC pattern is not maintained. The observation that CLCD increases as workload increases could be used to suggest that the adherence to CLC becomes more difficult and harder to maintain. This is likely due to the increased traffic density, when there is higher density there is less time that can be afforded between each communication.

Further exploratory analysis was conducted to analyze the transcripts to define categories of CLCD and identify them in the transcripts. By categorizing and counting the number of CLCDs it appeared that a majority of the CLCD's observed were due to normal pattern changes, with the rest being mostly attributable to interruptions, some of which were never addressed by the ATC. Observations of the many of the CLCDs categorized indicated either a delay in completing a CLC or failing to complete a CLC, both which are associated with human error (Helmreich, Meritt, Wilhelm, 1999). When first measuring for CLCD based on deviations from an expected pattern of communication between pilots and an ATC, it was expected that the coding scheme would detect some normal pattern changes, however it was not expected that the majority of the CLCD's detected would be classified as being normal. This suggests that a pattern-based detection of CLCD's needs to be more robust in order to accurately identify non-normal pattern changes that indicate that a CLCD has occurred. Regardless of this observation, CLCD as a variable to measure deviations from expected patterns of communications as applied in this thesis may still be useful in LOS prediction. This is because it is possible that increased pattern changes can be used to indicate higher traffic and more communications being handled.

Implications

CLCD is a variable that is highly correlated with LOS indicating that there is potential for CLCD to be used as a variable to predict LOS and could potentially be used in addition to other variables to increase LOS prediction. However, more testing with a greater sample size would be required in order to determine whether CLCD can predict LOS independent of Density. With more research CLCD could be evaluated as a variable that could contribute to systems that monitor communications in real time that could be used to predict LOS, and potentially detect communication related errors.

The results covered in this study contribute to the ongoing conversation about how to integrate communication-based measures into real time error detection systems. As was observed in this study, measures of communication must be precise in order to accurately detect human error. Categorizing CLCD's revealed that the initial coding scheme for CLCD's was not robust enough to detect CLCD's that would be defined as being problematic. However, it is worth mention that content coding can be costly. In recognition of this, CLCD's in this thesis were measured without content coding and the results indicated that even if the "normal" deviations are removed the same pattern relating CLCD to LOS persists. The categorical scheme in this thesis used to analyze CLCD's found in this study could be used in future studies to more accurately detect CLCDs, and could even be further applied to defining a variable feature that could be used to detect CLCDs in text based systems.

Limitations

Participants in the study were selected using strict criteria that limited that number of participants that were collected for participation in the study. This is because the participants selected were required to be retired ATC's. In one regard, the data collected was of high quality as it used ATC's who had a lot of experience, however in another regard this ultimately led to a small sample size. This small sample size made it difficult for statistical processing of the data. Initially the data analysis plan was to statistically analyze each condition of each trial, however this led to low statistical power, therefore the statistical test in this study had to consider each one of the participants trials. This low sample size led to difficulty in using multiple regression models as the addition of each variable into any multiple regression model would result in a decrease in degrees of freedom making it difficult to detect which variables significantly contributed to the multiple regression model. Furthermore, due to the high correlation between each of the variables CLCD, Density and LOS, multicollinearity was violated for each of the multiple regression models.

Results of the study may not reflect what occurs in the real world as ATC's have many more options for handling air traffic than they were provided a part of the studies design. In the experiment ATC's were instructed to handle all traffic that entered their sector, and were not allowed to handoff, in addition to other restrictions. Therefore, the density and levels of workload achieve in the experiment are not reflective of real-world conditions. This was done to artificially increase workload conditions and increase the amount of LOS observed in the study. The levels of LOS observed in the experiment are not realistic to real world conditions and is a variable that can only be safely manipulated to increase workload within a simulation environment. This study utilizes one-way radio communications, therefore the observations of communication in this study may not apply to text-based communication systems.

CLCD was measured on a non-fixed interval rate, which were reported at time transmission started, where LOS was measured on a fixed interval rate. CLCD was counted at the time each communication came across in the transcript while LOS was measured on a 5 second interval rate. In order to process the data LOS was fitted to the non-fixed sampling rate of each communication. LOS was also considered as a binary categorical variable to indicate whether two aircraft were in LOS or not, however this measure did not account for which aircraft were in LOS. The variable LOS in this study was categorical for whether two aircraft were in LOS or not at the time sampled, it did not however account for number of aircraft pairs that were in LOS.

Deviations from an expected pattern of CLC communication used to define the CLCD variable was not precise enough and detected too many normal pattern changes. This was not revealed until the end of the study, however analysis of the types of CLCD observed indicated that the variable CLCD defined in this study was not robust enough to filter out normal communication pattern changes. This is because one pilot can call in and perform a complete CLC exchange with an ATC and then another Pilot can call in immediately after, this would result in a pattern being coded as pilot:ATC:pilot:pilot, therefore the coding scheme would falsely consider these back to back pilot communications from two different pilots as a CLCD, when in fact the previous pilot had completed the CLC with the ATC in the previous exchange.

Pseudo-pilot and participant interactions were difficult to control for and may have been a variable that influenced CLCD count. Pseudo-pilots were student researchers who trained in at least three training sessions, and were trained in phraseology and Metacraft software use, and although some had previous experience with pilot: controller interactions, others did not. Pseudo-pilots were trained in initial call in procedures and were instructed to verbally repeat commands given to them by the ATC. Due to high traffic volumes and competition for speaking over a way one-way channel, the ability for pseudo-pilots to communicate as instructed may have been impacted. This however does not discount the number of CLCD observed, as CLCD is expected to capture delayed communications and missing communications. Although, pseudo-pilots were instructed on procedures and communications, this did not guarantee that pseudo-pilots were always consistent with their communications and adherence to ATC instruction.

Conclusion

Previous literature on CLC's has established that CLCD's can result in catastrophic outcomes in various domains which makes for measuring for CLCD's during air traffic control operations a worthwhile pursuit (Hargestam, Lindkvistt, Brulin, Jacobsson & Hultin, 2016; Bowers, Jentsch, Salas, et al., 1998; Siassakos, Bristowe, Draycott et al., 2011; Parush, Kramer, Foster-Hunt, et al., 2011). However, applying CLCD as a variable to be used to predict LOS in real time still needs evaluation and further research. Initially, the goal of this thesis was to attempt to come up with a measure of CLCD that could be used in real time and applied during ATC TRACON Operations. Evaluation of CLCD as defined by this thesis provided mixed results about the viability of measuring CLCD for the purposes of predicting LOS in real time.

Although CLCD as measured in this thesis is a questionable variable for predicting LOS, this thesis does provide the means by which CLCD can be redefined and applied in future studies to come up with a more precise measure of CLCD. Refinement of how CLCD is measured and defined could be applied in future research to reevaluate the relationship between CLCD, Density and LOS. Future work should focus on more precise measures of CLCD to predict LOS, as well as using CLCD to identify communication errors. Pursuing CLCD as a variable used to detect human error should be evaluated with the purpose of analyzing text-based communications between ATC's and pilots which could be applied to an alarm-based system that can detect communication errors in real time. An opportunity for fully investing CLCD could be done to consider the use of novice ATC's, doing so could provide more insight as to how the variable CLCD could be generalized across experience levels.

REFERENCES

Andersen PO, Jensen MK, Lippert A, et al. Identifying non-technical skills and barriers for improvement of teamwork in cardiac arrest teams. Resuscitation 2010;81:695–702.

Burke CS, Salas E, Wilson-Donnelly K, et al. How to turn a team of experts into an expert medical team: guidance from the aviation and military communities. Qual Saf Health Care 2004;13:I96–104.

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human factors*, *40*(4), 672-679.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. *Team cognition: Understanding the factors that drive process and performance*, 83-106.

Durso, F.T, & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), A Cognitive Approach to Situation Awareness: Theory and Application.(pp. 137-153). Ashgate: Aldershot.

FAR/AIM 2019: Federal aviation regulations, aeronautical information manual. (2018). Newcastle, WA: Aviation Supplies & Academics.

Gillespie, B. M., Chaboyer, W., & Fairweather, N. (2012). Interruptions and miscommunications in surgery: an observational study. *AORN journal*, *95*(5), 576-590.

Härgestam, M., Lindkvist, M., Brulin, C., Jacobsson, M., & Hultin, M. (2013). Communication in interdisciplinary teams: exploring closed-loop communication during in situ trauma team training. *BMJ open*, *3*(10), e003525.

Härgestam, M., Lindkvist, M., Jacobsson, M., Brulin, C., & Hultin, M. (2016). Trauma teams and time to early management during in situ trauma team training. *BMJ open*, *6*(1), e009911.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The international journal of aviation psychology*, *9*(1), 19-32.

ICAO Document 4444. (2007).

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). Errors in health care: a leading cause of death and injury. In *To err is human: Building a safer health system*. National Academies Press (US).

Ligda, S. V., Dao, A. Q. V., Vu, K. P., Strybel, T. Z., Battiste, V., & Johnson, W. W. (2010, September). Impact of conflict avoidance responsibility allocation on pilot workload in a distributed air traffic management system. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 1, pp. 55-59). Sage CA: Los Angeles, CA: Sage Publications.

Ligda, S. V., Seeds, M. L., Harris, M. J., Lieber, C. S., Demir, M., & Cooke, N. (2019). Monitoring Human Performance in Real-Time for NAS Safety Prognostics. In *AIAA Aviation 2019 Forum* (p. 3411).

Lingard, L., Espin, S., Rubin, B., Whyte, S., Colmenares, M., Baker, G. R., ... & Reznick, R. (2005). Getting teams to talk: development and pilot implementation of a checklist to promote interprofessional communication in the OR. *BMJ Quality & Safety*, *14*(5), 340-346.

Lingard, L., Regehr, G., Orser, B., Reznick, R., Baker, G. R., Doran, D., ... & Whyte, S. (2008). Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Archives of surgery*, *143*(1), 12-17.

Parush, A., Kramer, C., Foster-Hunt, T., Momtahan, K., Hunter, A., & Sohmer, B. (2011). Communication and team situation awareness in the OR: Implications for augmentative information display. *Journal of biomedical informatics*, *44*(3), 477-485.

Salas, E., Wilson, K. A., Murphy, C. E., King, H., & Salisbury, M. (2008). Communicating, coordinating, and cooperating when lives depend on it: tips for teamwork. *The Joint Commission Journal on Quality and Patient Safety*, *34*(6), 333-341.

Sevdalis, N., Undre, S., McDermott, J., Giddie, J., Diner, L., & Smith, G. (2014). Impact of intraoperative distractions on patient safety: a prospective descriptive study using validated instruments. *World journal of surgery*, *38*(4), 751-758.

Siassakos, D., Bristowe, K., Draycott, T. J., Angouri, J., Hambly, H., Winter, C., ... & Fox, R. (2011). Clinical efficiency in a simulated emergency and relationship to team behaviours: a multisite cross□sectional study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *118*(5), 596-607.

Suliman A, Wharton R, et al. The impact of operating room distractions on stress, workload, and teamwork. Ann Surg 2015;261:1079–84.

U. (2017, October 12). Aeronautical Information Manual: Official Guide to Basic Flight Information and ATC Procedures. Retrieved October 17, 2020, from https://www.faa.gov/air_traffic/publications/media/AIM_Basic_dtd_10-12-17.pdf

Vu, K. P. L., Strybel, T. Z., Battiste, V., Lachter, J., Dao, A. Q. V., Brandt, S., ... & Johnson, W. (2012). Pilot performance in trajectory-based operations under concepts of operation that vary separation responsibility across pilots, air traffic controllers, and automation. *International Journal of Human-Computer Interaction*, *28*(2), 107-118.

Wheelock, A., Suliman, A., Wharton, R., Babu, E. D., Hull, L., Vincent, C., ... & Arora, S. (2015). The impact of operating room distractions on stress, workload, and teamwork. *Annals of surgery*, *261*(6), 1079-1084.

Woloshynowych M, Davis R, Brown R, et al. Communication patterns in a UK emergency department. Ann Emerg Med 2007;50:407–13. 49.

APPENDIX A

DEMOGRAPHIC QUESTIONNAIRE

Appendix A

Demographic Questionnaire

Participant	Civilian Tower	Military Tower	Civilian TRACON	Military TRACON	Civilian Center	Military Center	Notes
SS01	Yes	No	Yes	No	No	No	
Years	10	-	25	-	-	-	
SS02	No	No	Yes	Yes	Yes	No	
Years	-	-	7	2		-	
SS03	Yes	No	Yes	No	Yes	No	
Years	6	-	20	-	4	-	
SS04	Yes	No	Yes	No	No	No	Reported that 12 years was combined between TRACON and tower
Years	12	-	18	-	-	-	
SS05	Yes	No	Yes	No	No	No	
Years	6	-	30	-	-	-	
SS06	Yes	No	Yes	No	Yes	No	
Years	7	-	18	-	15	-	
Range	0-12	0	7-25	0-2	0-15	0	
Mean	6.83	0	19.66	0.33	3.16	0	
Total Average Combined	30						
Range	9-40						

APPENDIX B

DEMOGRAPHIC QUESTIONNAIRE WORK EXPERIENCE RATINGS

Appendix B

Demographic Questionnaire Work Experience Ratings

Participant	Rate Your PHX TRACON (P50) Experience)	Rate your TRACON Experience	Rate your Experience Working and Arrival Problem in Simulation
SS01	1	7	7
SS02	1	7	7
SS03	6	7	7
SS04	7	7	7
SS05	7	7	7
SS06	7	7	7

APPENDIX C

DEMOGRAPHIC QUESTIONNAIRE WORK EXPERIENCE RESPONSES

Appendix C

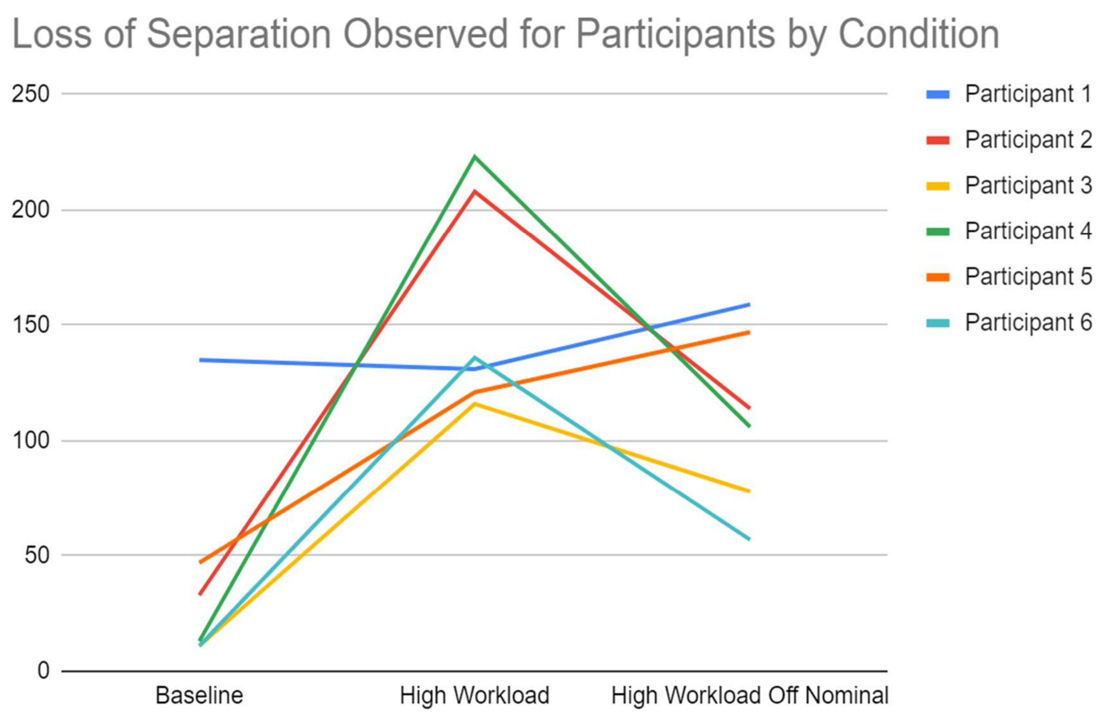
Demographic Questionnaire Work Experience Responses

Participant	Normal or Corrected to Normal Vision	Normal Color Vision	Have you used conflict detection automation in real world operations	Have you communicated with pilots using CPDLC data link in real world operations
SS01	Yes	Yes	Yes	No
SS02	Yes	Yes	Yes	No
SS03	Yes	Yes	No	No
SS04	No	Yes	No	No
SS05	Yes	Yes	Yes	No
SS06	Yes	Yes	Yes	No

APPENDIX D

LOSS OF SEPARATION OBSERVED FOR PARTICIPANTS BY CONDITION

Appendix D



Loss of Separation Observed for Participants by Condition
250
200
150
100
50
0
Baseline
High Workload
High Workload Off Nominal
Participant 1
Participant 2
Participant 3
Participant 4
Participant 5
Participant 6

APPENDIX E

PAIRWISE COMPARISONS FOR CLCD BETWEEN CONDITION

Appendix E

Pairwise Comparisons for CLCD Between Condition

				95% CI for Difference	
Condition (1)	Condition (1) Compared with:	Mean Difference	Std. Error	Lower Bound	Upper Bound
Baseline	High Workload	-14.333*	2.741	-24.019	-4.648
	High Workload Off Nominal	-16.667*	2.472	-25.403	-7.930
High Workload	Baseline	14.333*	2.741	4.648	24.019
	High Workload Off Nominal	-2.333	2.917	-12.644	7.977
High Workload Off Nominal	Baseline	16.667*	2.472	7.930	25.403
	High Workload	2.333	2.917	-7.977	12.644

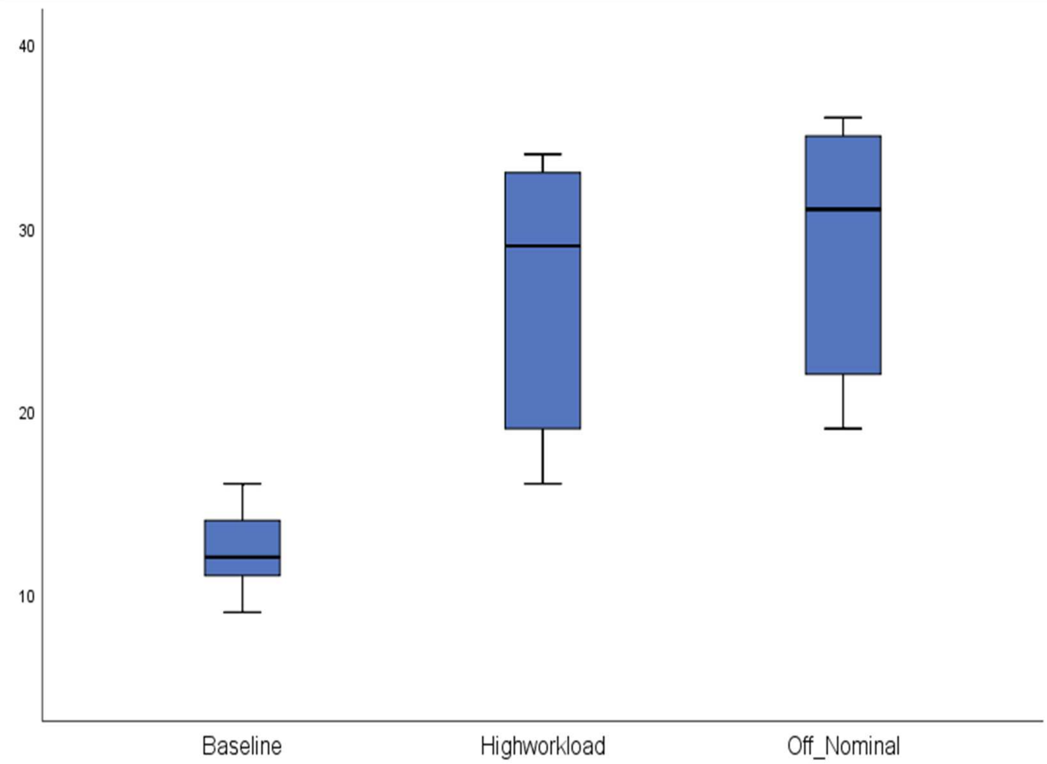
*The mean difference is significant at the .05 level.

APPENDIX F

BOX PLOT FOR CLCD MEAN VALUES BY CONDITION

Appendix F

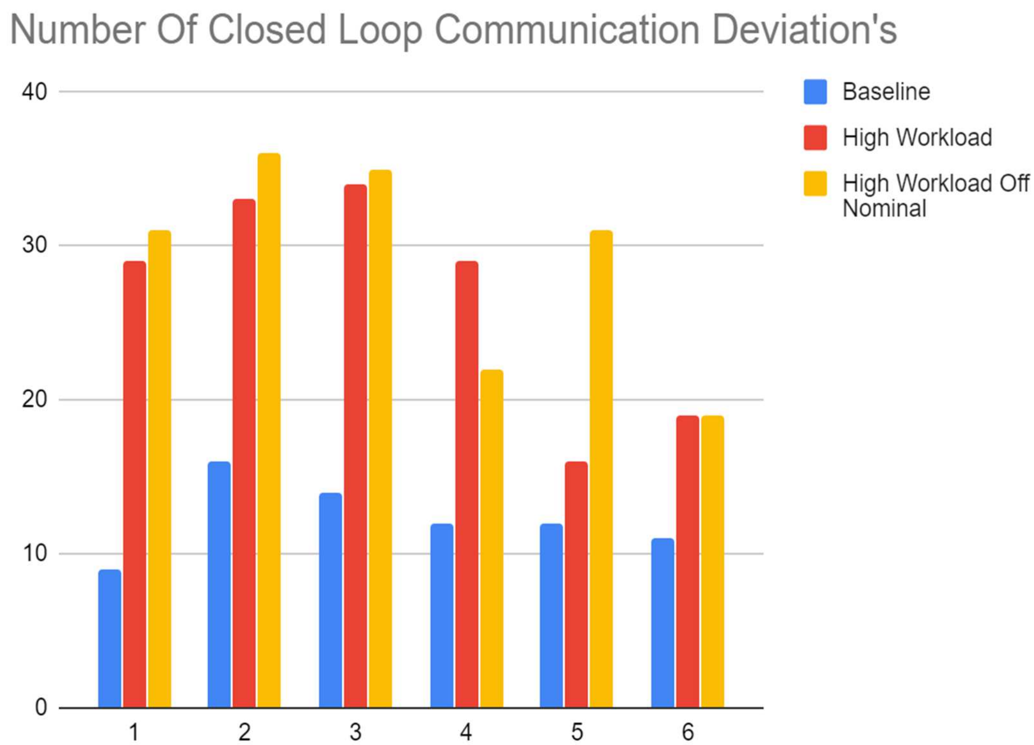
Box Plot for CLCD Mean Values by Condition



APPENDIX G

NUMBER OF CLCD BY CONDITION

Appendix G
Number of CLCD by Condition



APPENDIX H

NON-ADJUSTED WL RATINGS

Appendix H

Non-adjusted WL ratings

Participant	Trial Type	0 minutes	9 Minutes	18 Minutes	Avg
SS01	Baseline	1	6	7	4.67
SS01	High Workload	3	7	7	5.67
SS01	High Workload Off Nominal	3	7	7	5.67
SS02	Baseline	1	2	4	2.33
SS02	High Workload	1	7	7	5
SS02	High Workload Off Nominal	1	7	7	5
SS03	Baseline	1	-	1	1.5
SS03	High Workload	1	-	-	1
SS03	High Workload Off Nominal	2	-	7	4.5
SS04	Baseline	1	4	1	0
SS04	High Workload	1	5	-	3
SS04	High Workload Off Nominal	1	-	6	3.5
SS05	Baseline	2	5	3	3.3
SS05	High Workload	2	-	-	2
SS05	High Workload Off Nominal	2	-	-	2
SS06	Baseline	2	4	2	2.67
SS06	High Workload	1	-	7	2.67
SS06	High Workload Off Nominal	-	-	7	7
Total Avg Across All Trials	Baseline	-	-	-	2.76
Total Avg Across All Trials	High Workload	-	-	-	4.33
Total Avg Across All	High Workload Off	-	-	-	4.75

Trials	Nominal

APPENDIX I

ADJUSTED WL RATINGS

Appendix I

Adjusted WL Ratings

Participant	0 minutes	9 Minutes	18 Minutes	Avg
SS01	1	6	7	4.67
SS01	3	7	7	5.67
SS01	3	7	7	5.67
SS02	1	2	4	2.33
SS02	1	7	7	5
SS02	1	7	7	5
SS03	1	7	1	3
SS03	1	7	7	5
SS03	2	7	7	5.33
SS04	1	4	1	2
SS04	1	5	7	4.33
SS04	1	7	6	4.67
SS05	2	5	3	3.33
SS05	2	7	7	5.33
SS05	2	7	7	5.33
SS06	2	4	2	2.67
SS06	1	7	7	5
SS06	7	7	7	7
Total Avg Across All Trials				3
Total Avg Across All Trials				5.05
Total Avg Across All Trials				5.5

BIOGRAPHICAL SKETCH

Christopher Lieber graduated from University of Arizona in 2016 with a Bachelor of Science in Psychology and a Minor in Family Studies and Human Development. During his studies he worked intermittently as a paid and volunteer Firefighter/EMT. Christopher Lieber is now a Human Systems Engineering PhD student and Ira A. Fulton Schools of Engineering Dean's Fellow at Arizona State University, where he studies under Professor Nancy J. Cooke at the Cognitive Engineering Research on Team Tasks (CERTT) lab.