Monitoring Human Performance in Real-Time for NAS Safety Prognostics

Sarah V. Ligda [1], Mariah J. Harris[1], Christopher S. Lieber[1], and Nancy J. Cooke [2]
Arizona State University Polytechnic School, Mesa, AZ, 85212, USA

The present research explores how real-time communication data can be used to predict human performance of Air Traffic Management personnel, and ultimately risk in the U.S. National Airspace System (NAS). In three 25-minute simulated Human-In-The-Loop scenarios of varying complexity, two of twelve retired air traffic controllers worked arrival flows landing at Sky Harbor International (KPHX) under current day operations. During these scenarios, data were collected from several dimensions of operational performance defined by breaches in separation minima, controller-pilot radio communication, and both indirect and subjective measures of workload. Relationships among these variables are presented with focus on communication patterns associated with declines in operational performance during high workload. Preliminary results suggest more frequent separation breaches in both high workload conditions, one with with over three times as many separation breaches as the lower workload condition. We discuss next steps in linking the complex multifactorial dynamics associated with human performance to assessing those dynamics in real-time.

I. Introduction

A. Background

As an emergent property, safety of the NAS arises from interactions between many elements, ranging from those attributable to humans, technology, equipment, maintenance airspace, navigation, weather, terrain, and wildlife. In a system of systems, these interactions often occur between elements of different system components. Harris and Stanton (2010) aptly describe these components as selectively open systems: each needs to interact with other components, exchange resources and information, and operate under broad regulations to achieve overall system of system objectives [1]. However, each has its own operational, structural, and managerial independence with capabilities built to achieve its own set of local objectives, all which must comply with system of system constraints. A regional airliner built for economy flights has a unique set of objectives compared to a busy international airport, but both must exist and interact in compliance with larger FAA regulations. Accidents result from interactions that are insufficiently controlled by constraints, such as regulations [2] If an accident results from an incompatible interaction between an aircraft's landing gear and the pavement surface at an airport, the constraints to prevent that incompatibility are inadequate, not enforced, or not executed properly.

In 2016, more than 26 million flights carrying nearly 947 million passengers operated in the U.S. National Airspace System (NAS), a complex system of systems regulated by the Federal Aviation Administration [3]. To meet future growth rates of about 2% per year, advanced technologies, services, and procedures are being developed and implemented in the NAS under the Next Generation Air Transportation System program [4]. New capabilities to modernize the U.S. National Airspace System (NAS) rely upon real-time data streams derived from many sources across the NAS. These data streams can be integrated with other available sets of information to model components associated with loss of NAS safety. This then will support real-time risk mitigation by pinpointing areas in which the probability of safety loss is elevated. With these capabilities, new and existing sources of real-time data will be available and provide opportunities for system-wide diagnostic health information and prognostic risk assessment via data fusion [5]. *Our current focus is to identify available information sources that relate to human performance and may influence safety.*

---

[1] Graduate Research Associate, Human Systems Engineering, ASU POLY Campus Mailcode 2880
[2] Professor and Graduate Program Chair, Human Systems Engineering, ASU POLY Campus Mailcode 2880

Human performance is a critical component associated with safe NAS operations. Historically, human operators have often been found to be determinants of accidents that involve 14 Code of Federal Regulations (CFR) Part 121 air carrier operations, the regulatory class under which commercial airliners operate and the public most frequently uses [6, 7]. For instance, "personnel" is cited as a broad factor of approximately 80% of 446 air carrier accidents between 1997 and 2006, with "environment" cited in roughly 40% of accidents and "aircraft" cited in about 20% [8].

Human performance can be influenced by a variety of factors internal to the person (e.g., situation awareness, fatigue) or in the environment (e.g., workload). In identifying data sources for a real-time safety prognostic system for the NAS, the challenge is to identify a source of data that is indicative of human performance deficits in a predictive manner and in real-time. Specifically, we are restricted to sources of data that are permitted to be collected and processed in practice and in real-time (i.e., from current, real-world NAS operations) to ultimately model human performance.

Much of the early ATM human factors research investigated the criteria of controller performance that were thought to support system safety such as operational errors (OE). For instance, Kinney, Spahn, and Amato (1977) examined safety reports to better understand specific behavioral indicators associated with system errors in Center controllers, those responsible for sectors above roughly 20,000 feet [9]. The authors found attention, judgment, or communication directly caused 95% and contributed to 71% of a total of 564 errors between 1974 and 1976. Further analyses of reports show that in 1993 and 1994, controller-pilot communications contributed or caused 27% in 1993 and 26% in 1994 of legal loss of spatial separation between aircraft, a main OE violation [9]

A more recent literature review examined the relationships between nine human factors constructs most often linked to controller performance: stress, mental workload, communication, attention, vigilance, fatigue, situation awareness, trust and teamwork [10]. Within the sets of publications under review, the three most frequent factors found to impact performance are workload, stress, and fatigue. Manning and colleagues at FAA's Civil Aerospace Medical Institute describe the construct of cognitive workload as broad and well researched, but often difficult to directly assess in air traffic studies without subjective methods like questionnaire ratings or reaction times to real-time probes.

Alternatively, many objective workload metrics differ from cognitive workload in that they are based on rates of operational task load activity, yet are limited in that they lack evaluation qualities of the human controller [11, 12, 13] The FAA researchers also report that analyses of cognitive and objective workload data are frequently correlated and are believed to be associated with controller performance and OE rates.

Given the importance of controller-pilot communication in the literature and the availability of communication data in real-time, we are investigating how communication data can serve as an indicator or human performance in the NAS. Leveraging previous work familiar to the authors, we can take advantage of the rich, multidimensional nature of communication data to identify patterns of controller-pilot communication in real time and associate it with human performance deficits [14, 15]. Further, it may be feasible to identify more specific contributors to performance loss such as workload, fatigue, and situation awareness, thereby guiding interventions.

B. Current Objectives

For this first human-in-the-loop simulation of our five-year project, our primary goal is to examine how communication patterns (e.g., content, volume, flow patterns) are associated with controller workload and human performance. To guide us in this work, we present and evaluate early findings on three types of data: 1) operational simulator performance as defined by breaches in separation minima; 2) real-time workload ratings recorded at three, 12, and 21 min in each 25 minute scenario; 3) audio recordings of radio controller-pilot transmissions at the time of the workload ratings. This will inform analyses with these measures and others, including facial recognition, heart rate variability, situation awareness probes, and operational efficiency, and guide future directions in real-time human performance modeling.

II. Method

A. Participants and Design

The first two of 12 anticipated participants were retired air traffic controllers with experience at a FAA Terminal Radar Approach Control (TRACON) facility within the previous 15 years. This was to ensure baseline familiarly with role and standard airspace system infrastructure. Neither was familiar with the KPHX TRACON airspace. Each was compensated $60 per hour of participation.

The study design had one within-subject fixed factor (Scenario Condition) and one random factor (Controller). The Scenario Activity factor varied the frequency and type of air traffic operations necessary in order to maintain separation minima. This is to prompt a broad range of communication, human factors, and performance data. There were three conditions: a baseline scenario with moderately high traffic density in nominal airspace (Baseline), and two in which the density of the arrival traffic substantially increased. In one of these two higher-density scenarios, airspace activity was under nominal conditions (High Workload Nominal). The other contained common off-nominal events to further increase workload (High Workload Off-Nominal). These were a pilot deviation, Phoenix Sky Harbor International airport (KPHX) runway switch, and moderate turbulence in several experimental arrival flows.

The dependent measures are various aspects of controller-pilot voice communications via radio transmissions (e.g., flow – who talks to whom, audio frequency, pitch, cadence, verbal pauses, content amount, content repeats). Other measures include controller heart rate, facial expression, subjective workload reporting, situation awareness, and loss of separation events. Participant performance will also be judged by a subject matter expert after scenarios were complete.

B. Materials

ATC Workstations

The Metacraft facility consists of a fixed-motion, mid-fidelity Air Traffic Management system with equipment for eight operational positions. These positions are configured to operate as either ATC radar stations or multi-aircraft pseudo-pilot stations. All experimental hardware, software, and various pieces of equipment for participants closely resemble and simulate current-day functions of basic a ATC radar workstation in a TRACON facility. This workstation simulated a primary radar screen, flight strips, automatic terminal information service (ATIS) text broadcast, and radio frequency equipment for voice communications with pilots. Early NextGen technologies were enabled to support electronic hand-off procedures and area navigation (RNAV) approaches (e.g., Automatic Dependent Surveillance—Broadcast, Performance Based Navigation). Controller-Pilot Datalink Communication functionality was disabled.

Data Collection Instruments

Sets of airspace data were collected and recorded through the Metacraft system, including aircraft state, sector information, and airport activity dimensions. Audio recordings of the radio transmissions between participant controller and pseudo-pilot confederates were captured via radio headsets devices with push-to-talk functionality. These recordings were time stamped and documented using natural language processing software for further editing and analysis. Other non-intrusive recording software (e.g., screen recordings, keyboard press recordings, cursor movement and click recordings) were used but hidden from participants.

To collect facial expression and heart rate variability, an iMotion system with an electrocardiogram (ECG) headset, ECG software, facial expression software, an external camera (Logitech C920 HD Pro Webcam) and a Dell Alienware laptop were used to collect electrocardiogram and facial expressions data. Likert-type situation awareness questions and workload ratings using a modified version of the Situation Present Assessment Method [4] appeared on a touchscreen tablet every three min during scenarios. All hardware devices are commercially available.

C. Procedure

The testing took place in the Metacraft simulator facility located on the second floor of the Simulator Building at Arizona State University Polytechnic campus. A high-level conceptual introduction and hands-on training were provided to become familiar with the mid-fidelity ATC radar workstation and the environmental real-world constraints specific to the Phoenix terminal airspace. Additionally, to increase the difficulty of the three scenarios, five mandatory rules were communicated to each participant to be enforced throughout testing:

1) *You must accept all handoffs from center. 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 filed destination. No alternate airports. You may only hand off to final approach.*
4) *Keep aircraft in your airspace. No handoffs (except to 120.9 sector) and no pointouts.*
5) *You must not declare emergencies.*

During the hands-on training sessions and data collection scenarios, participants sat at a mid-fidelity ATC radar TRACON controller station simulator while performing standard current-day ATC specialist (controller) operations in terminal airspace closely modeled after P50 approach sectors that service KPHX.

Once participants were trained and familiarized with the task and airspace, the three counterbalanced 25-minute data collection scenarios commenced. Three confederates operated Metacraft stations configured for multi-aircraft pseudo-pilot operation. These confederates provided air traffic within the terminal P50 airspace in order to increase traffic levels and provide communication sector check-ins (e.g., “Phoenix approach, this is United 305 with you on the HYDDR1 arrival at one one thousand feet”) and other responses to commands from the participant controller. After the third scenario was complete, the controller completed a paper-and-pen questionnaire evaluating the workload, difficulty, and other related aspects of the simulation. This entire process lasted between three and four hours per participant.

III. Results

To better integrate data sets in this and future simulations and gain overall insight on relationships between communication and human performance during both nominal and off-nominal ATC operations during high workload periods, we provide preliminary analyses collected from two experienced TRACON controllers who worked three 25-minute arrival scenarios varying in degrees of airliner traffic density and complex but relatively common off-nominal events. We analyze dimensions of loss of separation events (defined by breaches in separation minima of less than 5 nm radius and +/- 1000 foot vertical distance), aspects of content and non-content controller-pilot communication exchanges, and real-time workload ratings collected at three, 12, and 21 minutes (180, 720, and 1260 seconds, respectively) from the start of each scenario.

A. LOS analyses: Proximity at closest point, frequency, and duration of separation breaches

In all six of the arrival scenarios, multiple pairs of aircraft lost separation while in the controller’s airspace. This was expected as the airspace configuration, traffic density, and stringent restrictions of the simulation were predicted to be especially challenging to assess behavior at various degrees of difficulty and workload. Although the traffic density in Baseline is half of what it is in either of the two High Workload conditions (15 compared to 30 aircraft), both participants verbally confirmed in debrief sessions that all scenarios were difficult, particularly the last two-thirds. This is also observed in the per minute count of separation breaches presented in Fig. 1.

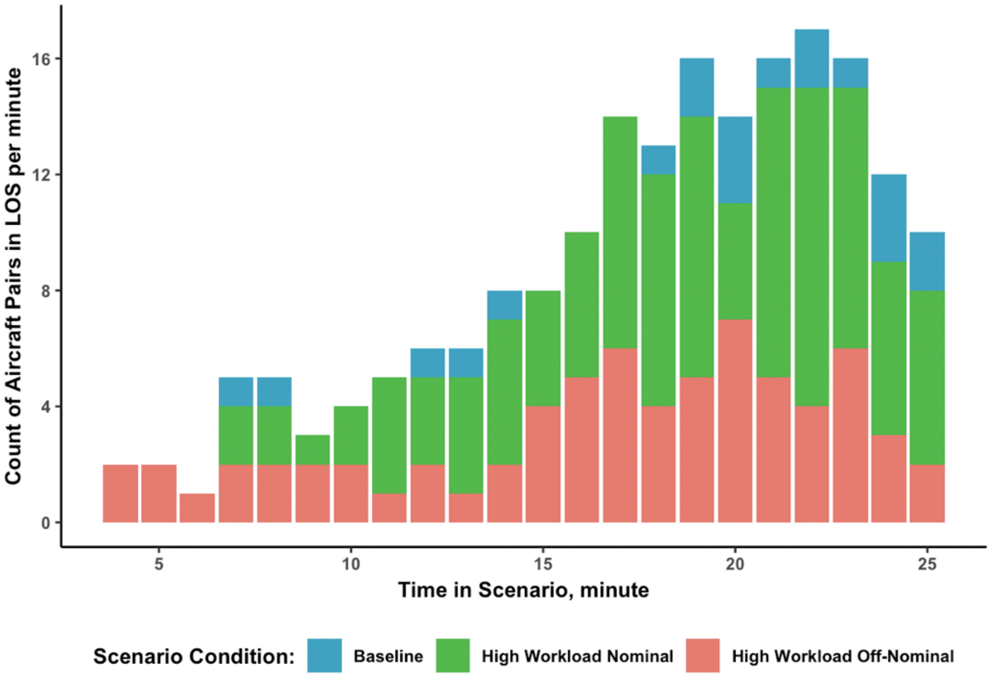

Fig. 1 Total count of aircraft pairs in loss of separation per minute over two ATC participants

Also noteworthy is that even though the Baseline condition is designed to produce a constant workload and perceived difficulty for the controller throughout the 25 minute session - unlike the two High Workload conditions in which difficulty is expected to increase - preliminary findings show that the resulting LOS count increased over time. This is likely due to the initial conflict-free traffic configuration of all three scenarios.

Initial overall trends in operational performance reveal differences between Baseline and High Workload conditions, particularly the safety outcomes in the High Workload Nominal scenario. This difference is most apparent in the overall frequency with which aircraft pairs lost separation during each scenario: over three times as many separation breaches occurred in the High Workload Nominal scenario as Baseline as can be seen in Fig. 2.

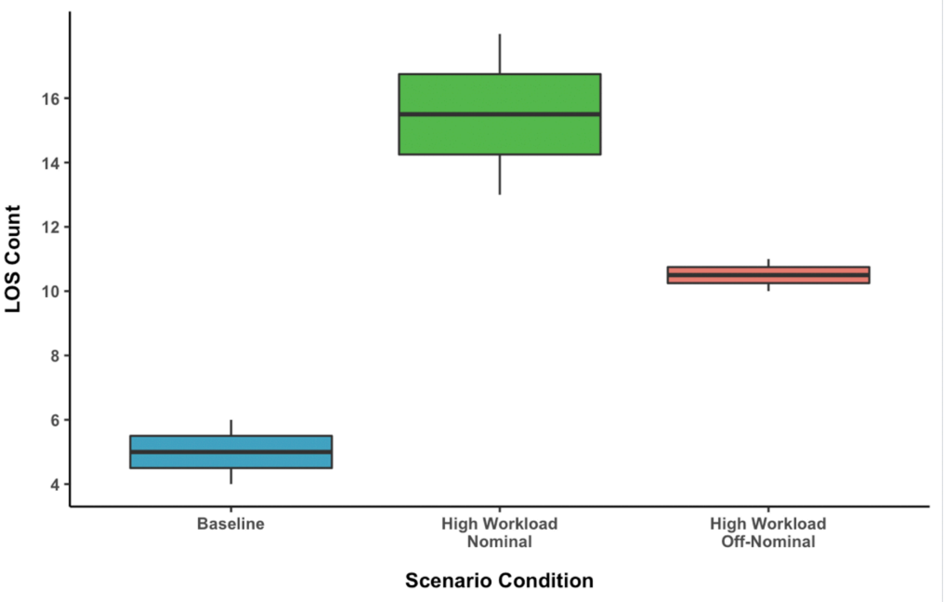

Fig. 2 Frequency of losses of separation between aircraft pairs as a function of condition.

The lower overall LOS count in High Off-Nominal when compared to High Nominal is an interesting finding, especially considering the number of total aircraft in the experimental airspace was the same in both High Workload conditions with similar traffic configurations. However, further investigation shows preliminary discrepancies in both the proximity and duration of separation losses between these conditions. In the Off-Nominal scenario, over a quarter of separation breaches at closest point were at a distance less than one nm, more than double that of both Baseline and High Nominal. Both High Workload scenarios contained breaches lasting over five minutes whereas Baseline's breaches were comparatively shorter.

Table 1 Proximity at closest point and duration of LOS by scenario condition

	Baseline (mean = 5 LOS)	**High Workload Nominal (mean = 15.5 LOS)**	**High Workload Off-Nominal (mean = 13 LOS)**
LOS Proximity at Closest Point	<1 nm: 10%	<1 nm: 12.9%	<1 nm: 28.6%
	1-2 nm: 10%	1-2 nm: 9.7%	1-2 nm: 14.3%
	2-3 nm: 30%	2-3 nm: 16.1%	2-3 nm: 9.5%
	3-4 nm: 20%	3-4 nm: 22.6%	3-4 nm: 33.3%
	4-5 nm: 30%	4-5 nm: 38.7%	4-5 nm: 14.3%

LOS Duration	<1 min: 70%	<1 min: 22.6%	<1 min: 38.1%
	1-3 min: 20%	1-3 min: 38.7%	1-3 min: 28.6%
	3-5 min: 10%	3-5 min: 12.9%	3-5 min: 19.0%
	--	6-9 min: 25.8%	8-10 min: 14.3%

B. Communication and Workload analyses

Subjective workload ratings were assessed at 3, 12, and 21 minutes (180, 720, and 1260 seconds) into each scenario. Table 2 presents descriptive statistics of these workload ratings as well as a summary of communication exchanges surrounding the real-time probes. We extracted communication samples from the same participant and condition.

Table 2 Summary of real-time workload probe ratings and communication exchanges

Workload Probe Initiated at 3 min	**Workload Probe Initiated at 12 min**	**Workload Probe Initiated at 21 min**
2.25 sec to press “Ready” 1.6 workload rating 1 low – 7 high (6 data points)	21.45 sec to press “Ready” 6.0 workload rating 1 low – 7 high (6 data points)	11.85 sec to press “Ready” 6.5 workload rating (1 low – 7 high) (6 data points)
Controller-Pilot Communication (2:20-3:40 min)	**Controller-Pilot Communication (11:20-12:40 min)**	**Controller-Pilot Communication (20:20-21:40 min)**
Content Summary 40 seconds prior to Workload Probe, controller issued strategic altitude command to JetBlue reduce speed when at 11,000. Pilot repeats transmission as confirmation. 70 seconds later, JetBlue 475 gives conformation that they’ve reached 11,000 and are reducing speed.	**Summary** Speakers talking over each other frequently, increased speed, requests for repeats from pilots. Controller speaking to 6-8 pilots within this 80 second period. Appears only tactical information is exchanged (no planning comm). Multiple sets of information exchanged in one transmission	**Summary** Speakers talking frequently, increased speed from 3 min period, but slower than 12 min periods, requests for repeats from pilots. Controller speaking to 5-7 pilots within this 80 second period. Multiple sets of information exchanged in one transmission Appears only tactical information is exchanged (no planning comm)
Non-Content Summary Relatively fast speed, clear cadence. ATC: clear enunciation, stress, and slightly reduced speed on on essential information (“reduce speed to two one zero”)	**Non-Content Summary** Louder, faster enunciation from all speakers, stress on multiple utterances Two counts of conflict alerting	**Non-Content Summary** Louder, not as fast enunciation from all speakers, stress on multiple utterances Four counts of conflict alerting
Communication Sample ATC: jet blue four seventy five reaching one zero thousand reduce your speed to two one zero JBU475: roger that as soon as we hit ten we’re going down to two one zero	**Communication Sample** ATC: delta sixty three ninety two, reduce speed to one nine zero, maintain altitude at eight thousand DAL6392: reducing speed to one niner zero, maintaining at eight delta sixty three ninety two	**Communication Sample** ATC: delta sixty three ninety two, reduce speed to one nine zero, maintain altitude at eight thousand DAL6392: reducing speed to one niner zero, maintaining at eight delta sixty three ninety two

Controller Rate of Speech 5.7 syllables per second	**Controller Rate of Speech** 7.6 syllables per second	**Controller Rate of Speech** 6.7 syllables per second

IV. Discussion

The sets of operational performance and controller-pilot communication exchanges between arrival controllers and confederate pilots in our observations thus far is insightful in several respects. First, it appears that the configuration and structure of the three scenarios is indeed producing greater workload and difficultly in all three levels of our within-subject factor as confirmed by several types of measures: verbal communication (quantity of communication exchanges, speed with which content is delivered, information exchanged in those communications, and strategic vs. tactical intent); real-time workload ratings (indirect measure of ready latency and subjective response of workload rating); and discussions in post-simulation debriefings.

Furthermore, differences in controller's objective performance of safety as defined by breaches in separation minima is a particularly noteworthy conformation of workload and difficulty. The implications of this trend suggest that not only did the controllers *perceive* themselves to be experiencing higher degrees of workload and difficulty - as well as higher taskload suggested by quantity of communication exchanges - the impact of their higher workload and difficulty level produced objective deficits in safety performance. This finding is essential to our overarching objectives

Differences in the type of separation breaches between the three conditions in addition to a direct count should be explored and analyzed further. Breaches in separation in close proximity, such as those that reach less than one nm, or continue for over five or 10 minutes are more serious than others. Further research to understand the contributing factors and consequences of these breaches is required.

Also recommended for further examination is the type of communication exchanged between speakers. We noticed a shift from planning communication (strategic) in less busy periods to more tactical, immediate commands from controllers.

Acknowledgments

The research reported in this paper is supported by funds from NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Baiwa, Principal Investigator: Dr. Yongming Liu). The support is gratefully acknowledged.

References

[1] Harris, D. & Stanton, N. A. (2010). Aviation as a system of systems: Preface to the special issue of human factors in aviation. *Ergonomics, 53*(2), 145-148.

[2] Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42,* 237–270.

[3] FAA. (2017). Air traffic by the numbers. Washington, DC: Federal Aviation Administration.

[4] Joint Planning and Development Office. Next Generation Transportation System: Concept of Operation v.3.0. Government Printing Office, Washington D.C. (2010)

[5] Liu, Y. & Goebel, K. "Information Fusion for National Airspace System Prognostics." *PHM Society Conference*. Vol. 10. No. 1. 2018.

[6] 14 CFR 121.404. In: Code of Federal Regulations, Title 14 Aeronautics and Space, Parts 110-199, pp. 155-197, 2014.

[7] FAA Air Traffic Organization, "Air Traffic Organization 2015 Safety Report," Report No. 2016-AJI-091, 2016.

[8] National Transportation Safety Board, "Annual Review of Aircraft Accident Data: U.S. Air Carrier Operations Calendar Year 2006," Report No. NTSB/ARC-10/01, 2006.

[9] Kinney, G.C., Spahn, J., & Amato, R.A. (1977). *The human element in air traffic control: Observations and analyses of the performance of controllers and supervisors in providing ATC separation services* (Report No. MTR-7655). McLean, VA: METREK Division of the MITRE Corporation.

[10] Prinzo, O. V. (1996). *An analysis of approach control/pilot voice communications* (NTIS no. ADA317528). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

[11] Edwards, T., Sharples, S., Wilson, J. R., and Kirwan, B. (2012). Factor interaction influences on human performance in air traffic control: The need for a multifactorial model. *Work: A Journal of Prevention, Assessment and Rehabilitation, 41*(1), 159-166.

[12] Manning, C. A. (2000). *Measuring air traffic controller performance in a high-fidelity simulation* (NTIS No. ADA373813). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

[13] Manning, C. A., Mills, S. H., Fox, C. M., Pfleiderer, E. M., & Mogilka, H. J. (2001). The relationship between air traffic control communication events and measures of controller taskload and workload. *4th USA/Europe Air Traffic Management R&D Seminar.* Santa Fe, NM.

[14] Manning, C. A., Mills, S. H., Fox, C. M., Pfleiderer, E. M., & Mogilka, H. J. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload* (NTIS No. ADA401923). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

[15] Crutchfield, J., & Rosenberg, C. (2007). *Predicting subjective workload ratings: A comparison and synthesis of operational and theoretical models* (Report No. DOT/FAA/AM-07/6). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

[16] Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of Cognitive Engineering and Decision Making, 3,* 27-46.

[17] Gorman, J. C., Hessler, E. E., Amazeen, P. G., Cooke, N. J., & Shope, S. M. (2012). Dynamical analysis in real time: Detecting perturbations to team communication. *Ergonomics, 55,* 825-839.