

## THE IMPACT OF SECTOR CHARACTERISTICS AND AIRCRAFT COUNT ON AIR TRAFFIC CONTROL COMMUNICATIONS AND WORKLOAD

Esa M. Rantanen, Philip W. Maynard and Deniz Özhan

University of Illinois at Urbana-Champaign, Aviation Human Factors Division  
Savoy, Illinois, USA

Means of communication between pilots and controllers is one of the fundamental principles of air traffic control (ATC). Consequently, air-ground communications will both reflect the taskload imposed on the controller as well as drive the workload experienced by the controller. Therefore, analysis of ATC communications could potentially reveal a very rich and detailed picture of the demands placed on a controller in a given sector and traffic situation. This paper reports analysis of ATC voice data obtained from three different sectors at the Indianapolis air route traffic control center (ZID ARTCC). The main purpose of this analysis was to examine how different sector characteristics and the busy and slow periods within the sectors differed explicitly in terms of pilot-controller communications and implicitly in terms of controller taskload and workload. Measures derived from the voice data were also compared to metrics reflecting ATC sector complexity that were derived from the output of the objective activity and taskload assessment program, POWER, developed by the FAA.

### Introduction

Mental workload is frequently cited as the most critical characteristic of air traffic controller's task (e.g., Hopkin, 1995; Wickens, Mavor, & McGee, 1997; Wickens, Mavor, Parasuraman, & McGee, 1998). Mental workload experienced by controllers is frequently cited as a limiting factor to the capacity of the entire national airspace system (NAS) and evaluation of new technologies and forms of automation is often focused on their impact on controller workload in particular. Mental workload, however, is a complex and multidimensional theoretical construct, influenced by numerous interacting factors (Meshkati, 1988; Vidulich, 2003), and with a substantial subjective component. Hence, workload is covert, individually experienced by controllers and not directly measurable.

Controller workload could, however, possibly be inferred from other, overt, aspects of controllers' work. For example, there has been much research activity to quantify taskload by measurable characteristics of traffic in an ATC sector or by the infrastructure of the sector itself, collectively known as sector or traffic complexity or dynamic density. Much of air traffic controllers' work also involves spoken communication. Presently, virtually all control actions must be communicated to pilots via voice radio. Hence, voice communications are intuitively and unsurprisingly an attractive method for examining controller workload. However, little research has been done to validate and quantify the putative relationships between sector complexity and workload on one hand, and controller communications and workload on the other. The purpose of the research re-

ported in this paper was to examine the relationship between various sector complexity measures and ATC communications, and thus attempt to bracket controller workload.

### ATC Complexity and Workload

Dynamic density is used in a variety of contexts in the literature and does not necessarily correspond to a single metric, but Laudeman, Shelden, Branstrom, and Brasil (1998) and Sridhar, Sheth, and Grabbe (1998) have reported an equation for this construct. The index sums nine specific variables, each multiplied by a weight derived from regression analysis of controller activity data and subjective ratings. No definition was provided for traffic density, however. Another complexity metric, risk index, is an index of collision risk (Knecht, Smith, & Hancock, 1996) and it has also been referred to as dynamic density (Smith, Scallen, Knecht, & Hancock, 1998). It is derived from two directly measurable variables, (1) number of aircraft at a given altitude,  $N$ , and (2) distance from the  $i$ th to the  $j$ th aircraft,  $d_{ij}$ . Other complexity metrics include predictive workload measures based on work done at NASA Ames research center by Chatterji and Sridhar (2001), and work by Wyndemere, Inc. (1996; see also Pawlak, Brinton, Crouch, & Lancaster, 1996 and Pawlak & Brinton, 1996). Summaries of the many metrics are provided by the FAA (2000) and Kopardekar and Magyarits (2002).

No theoretical foundations for measurement of these constructs could be established from the ATC research literature, as such were not provided. Instead, it often seems to be the case that validity of infer-

ences made about covert, not directly measurable constructs is based only on the authors' proclamation that by measuring A (a directly measurable variable) they were in fact also measuring B (a covert, only indirectly measurable variable). Hence, much research remains to be done to create and validate a theoretical framework for establishing rigorous and reliable connections between directly measurable variables and indirect constructs of interest.

#### ATC Communications and Workload

A much better connection has been established between workload and communications. A comprehensive study by Casali and Wierwille (1983) manipulated communication load during a simulated flight task; in addition to normal ATC instructions, the subjects were required to perform a call sign recognition task, with target call signs embedded in sets of extraneous call signs of varying difficulty. Of 16 workload measurement techniques employed, eight were sensitive to communication load manipulations. These techniques included both subjective ratings and objective measures. Hence, it is quite clear that communications load is a workload driver. However, the data reported by Casali and Wierwille (1983) does not allow for a reverse relationship to be established, that is, estimation of workload by analysis of the communication load. Several reasons prevent this: first, the article did not report any overall measures of communication load, such as number and durations of communications, and second, there were several other sources of workload present in the experiment, for example, piloting of the simulator. It should also be noted that a communication task is very different for a pilot and a controller. A pilot typically needs to respond to only a small fraction of messages transmitted on the frequency (i.e., only to those addressed to him or her), whereas the ratio of messages controllers receive and transmit is close to one (i.e., controllers talk to all aircraft on frequency).

Hurst and Rose (1978) replicated an earlier study that had indicated that peak traffic and the duration of radio communications were good predictors of behavioral response of air traffic controllers working in air route traffic control centers. This study included 3,110 observations made on radar sectors at the 13 major radar control rooms in the U. S. Duration of radio communications compared to behavioral ratings were made by expert-observer controllers showed that the former were good predictors of the latter.

A very strong relationship between controller workload and communications load was established in a study by Porterfield (1997). This study used ATC

communications recorded from high-fidelity simulations and compared communication times to concurrently recorded subjective workload estimates (Air Traffic Workload Input Technique, ATWIT). The primary communications metric was average communication time per minute, calculated for 4-minute intervals to match ATWIT probes. A maximum coefficient of correlation of .88 indeed is very impressive, and the average communication time per minute also closely followed ATWIT ratings over a 15-minute period. However, the ATWIT ratings were generally very low, maximum ratings 3.5 on a scale from 0 to 7. At a workload rating 3.5 the communication load was 11 s per minute, or a proportion of .183.

Manning, Mills, Fox, Pfliegerer, and Mogilka (2002) analyzed 12 traffic samples from Kansas City Air Route Traffic Control Center (ZKC ARTCC). These traffic samples were viewed on SATORI (Rodgers & Duke, 1993) software, which recreated the traffic situations, by 16 ATC instructors who provided ATWIT workload estimates at 4-minute intervals. The samples were also processed by POWER software, which extracted a number of objective ATC taskload metrics from the data. Communications were quantified by the number of communication events and their durations, categorized by their content and speaker, as well as total communication times in 4-minute time epochs. The multitude of dependent variables was subjected to principal components analysis to reduce their number and like measures were combined to four taskload components. The results showed significant correlations between ATWIT ratings and total number and duration of communications ( $r = .62, p < .01$ ), and individual communication durations ( $r = .36, p < .05$ ), as well as number of instructional clearances ( $r = .65, p < .01$ ). The activity component of taskload, which combined number of aircraft, number of simultaneously controlled aircraft, and radar controller data entries, was also correlated with total number and duration of communications ( $r = .63, p < .01$ ), as well as with the number of frequency changes ( $r = .36, p < .05$ ) and instructional clearances ( $r = .52, p < .01$ ). Hence, it may be concluded that communication metrics may be a valid indicator of controller workload and taskload, although the r-values reported certainly leave other factors to be accounted for.

#### Availability of Data

Recent technological advances, particularly in area of digital technology, and the ATC modernization efforts potentially make available new sources for data as well as data collection and storage methods. An example of access to data from which various meas-

ures can be derived is the System Activity Recordings (SAR) that stores all flight and radar information in Air Route Traffic Control Centers (ARTCCs). These data can be further processed by two specific computer programs, the Data Analysis and Reduction Tool (DART) (Federal Aviation Administration [FAA], 1993) and the National Track Analysis Program (NTAP) (FAA, 1991), which produce a number of text-based output files. These files can be further analyzed by specialized computer programs, such as the Performance and Objective Workload Evaluation Research (POWER) (Mills, Manning, & Pfliegerer, 1999; Manning, Mills, Fox, & Pfliegerer, 2000). Currently, the POWER program derives over 40 separate measures that describe a variety of aspects of ATC.

Although a number of POWER measures have been shown to correlate with other sector complexity and workload measures, their relationship with controller performance is less clear (Manning, et al., 2000). On the other hand, ATC voice data has been shown to be a good indicator of controller workload (Hurst & Rose, 1978; Porterfield, 1997; Manning et al., 2002), but they remain difficult to obtain and painstaking to analyze. If a valid relationship between certain complexity metrics and communication measures could be established, however, that would allow bypassing analysis of voice data in favor of mostly automatic data collection via POWER and similar tools.

### Method

Data from three sectors from the Indianapolis air route traffic control center (ZID ARTCC) were selected for POWER analysis. The selection criterion for these sectors was simply that they should be very different from each other with unique characteristics in terms of traffic patterns and load. A senior supervisor from ZID chose the sectors based on these requirements and his expert judgment; the sectors were River (26, RIV) low-altitude sector, Dayton (88, DAY) high-altitude sector, and Wabash (99, WAB) super high-altitude sector. Two one-hour samples from each sector were obtained, one from busy and one from slow time of day.

Analysis of these data by POWER yielded many variables pertinent to sector complexity. In addition, voice data from the same samples were obtained and converted to wav files. These files were analyzed by SPWave program (SPWave is freeware and can be downloaded from <http://www.itakura.nuee.nagoya-u.ac.jp/people/banno/spLibs/spwave/>). This program allowed for visualization of the voice data as a spectrogram, and a zoom capability allowed for very ac-

curate determination of transmission begin and end times. The data were coded (but not transcribed) and entered into an Excel spreadsheet. From the coded data a total of 53 variables were derived. These data were then compared to a number of complexity metrics that could be derived from the POWER output. Both voice and POWER data were examined in 10-minute epochs within the 1-hr samples.

### Results

There were a total of 53 separate variables that were derived from the voice data. The results reported here, however, only pertain to those variables that either have been shown to correlate with controller workload and those that showed significant differences between the different ZID sectors. Furthermore, total number and duration of communications were highly correlated, as might be expected ( $R$ -squared = 0.854) and therefore only communication duration is discussed here.

#### Differences Between Sectors

As Porterfield (1997) and Manning et al. (2002) had discovered, communication time was a good predictor of workload (subjective ratings) and it was therefore of interest to examine whether the three ZID sectors differed from each other in this respect (see Fig. 1). An ANOVA on the proportion of controller communication time showed nearly significant (at  $\alpha = .05$ ) differences between sectors,  $F(2, 29) = 2.90$ ,  $p = .071$ , and significant differences between busy and slow times,  $F(1, 29) = 20.31$ ,  $p < 0.001$ . The interaction between sector and time (busy or slow) was not significant. These results, however, should be moderated by the small sample size, with only 6 data points (epochs) per condition.

Number of instructional clearances has also been associated with controller workload (Manning et al., 2002) and clear differences were found between the sample ZID sectors (Figure 2). An ANOVA showed significant differences between sectors,  $F(2, 28) = 7.07$ ,  $p < .0005$ , and between times,  $F(1, 28) = 19.09$ ,  $p < .0005$ . The interaction between sector and time was not significant, however.

Finally, we examined the number of frequency changes between sectors, as this variable has also been shown to correlate with workload. No statistically significant differences between sectors in the ZID sample were found, however, but time had a significant effect on the number of frequency changes,  $F(1, 27) = 17.51$ ,  $p < .0005$ . This results is not surprising, as number of frequency changes

strongly correlates with the number of aircraft in the sample, which clearly is the main difference between busy and slow times.

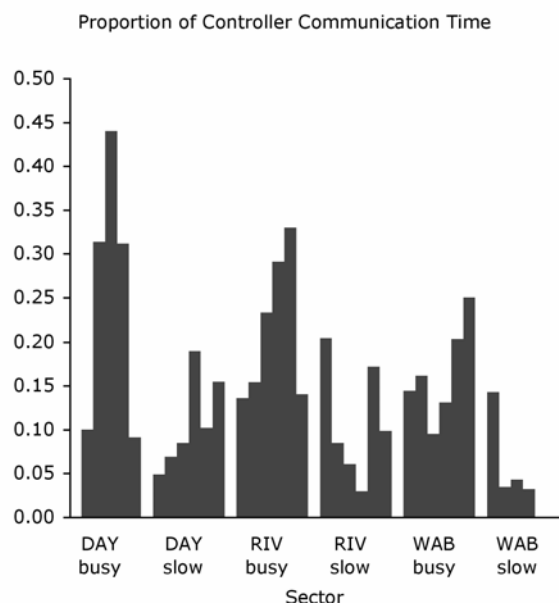


Figure 1. Proportion of controller communication time in the six samples from ZID. Note that the maximum in DAY sector during busy time approaches 50%, meaning that the controller was speaking for almost half of the time during the 10-minute epoch. WAB had much lower communication load than the other two sectors.

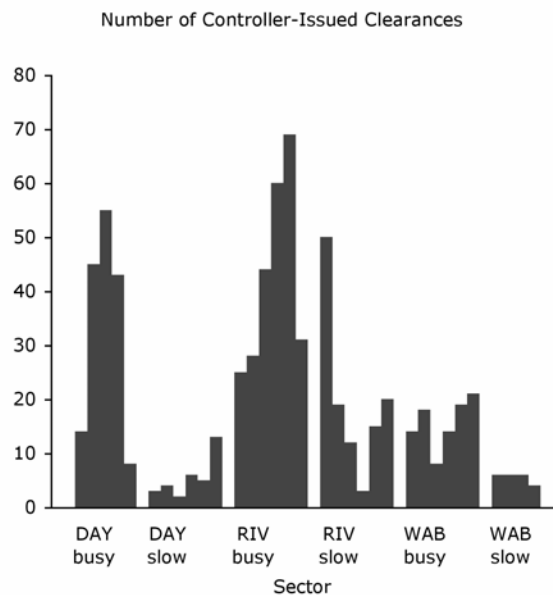


Figure 2. Number of controller-issued clearances shows significant differences not only between times but also between sectors. RIV is clearly in a class of its own, as might be expected for a feeder sector.

#### Aircraft Count and Communications Load

Given that communication time has been found to be a good predictor of workload (Manning et al., 2002; Porterfield, 1997), we examined correlations between the communication time recorded from the ZID voice data and POWER metrics from the same samples. Best correlation was found between the sum of three controller activity metrics (altitude changes + heading changes + number of handoffs) and controller communication time. The premise was that as aircraft altitude and heading changes currently necessitate a clearance, as does handoffs, they can be combined into an index that captures most of controller activity (Activity Count). The results are depicted in Figure 3 below. A regression analysis showed a significant relationship between activity count and communication time,  $F(1, 31) = 43.66$ ,  $p < .0001$ ,  $R\text{-squared} = .5848$ .

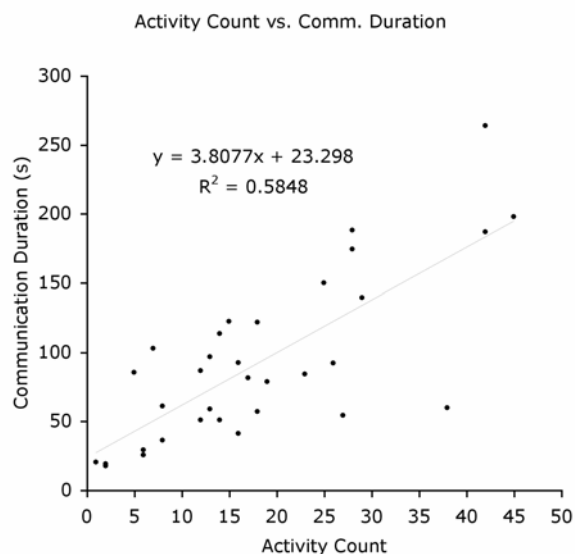


Figure 3. Activity count, which is a sum of three POWER metrics (altitude changes + heading changes + number of handoffs) regressed against controller communication duration, which in turn has been shown to be a good predictor of workload.

However, only slightly poorer results were obtained from regression of aircraft count and controller communication duration. A linear regression yielded a significant relationship between these variables,  $F(1, 31) = 37.83$ ,  $p < .0001$ ,  $R\text{-squared} = .5496$  (Figure 4). Since aircraft count is much easier to obtain from data, this metric appears to suffice for an indicator of controller workload, inferred from communication duration.

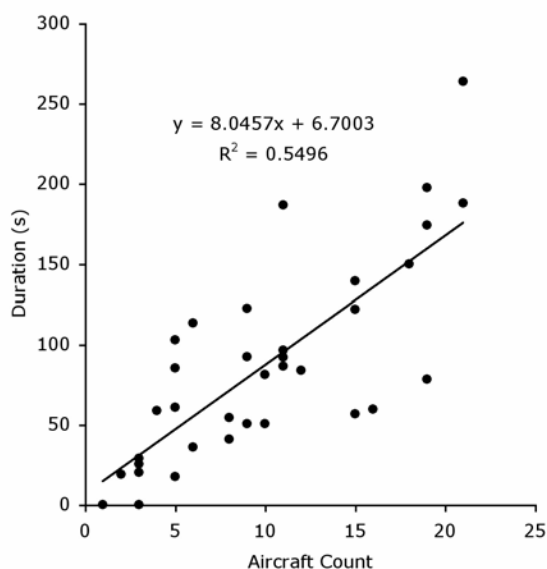


Figure 4. Relationship between aircraft count and cumulative controller communication duration was positive and statistically significant. The correlation coefficient was only slightly poorer than that obtained from activity count.

#### Discussion

Analysis of ATC voice data from three different sectors of ZID ARTCC revealed substantial differences between the sectors as well as between busy and slow times within the sectors. Given that communication duration and number of clearances issued have been shown to be workload drivers, we may conclude that the sample sectors can indeed be ranked in terms of workload imposed on the controller. In this respect, it appears that the high-altitude DAY and low-altitude RIV sectors were much more demanding than the superhigh-altitude WAB sector. Furthermore, it also appears that a simple metric of controller taskload, that is, aircraft count, correlated nearly as well with communication duration as did the more complex activity count, clearly favoring the use of the former as an indicator of controller workload. There are, however, several caveats that should be considered when assessing the validity of these conclusions.

First, a number of POWER metrics clearly differentiated between the sectors of different characteristics, revealing important factors that might affect controller taskload (e.g., maximum number of aircraft under controller's responsibility at any one time, proportion of aircraft changing altitude, handoff acceptance latency) that were not reflected in voice data. Equally important is to consider metrics that remained essentially invariant between sectors (e.g., number of air-

craft), as these may reflect taskload factors that are independent from sector characteristics.

Second, although the POWER output included many parameters that were also part of the proposed airspace complexity and dynamic density measures as reviewed before, none of these metrics could be fully calculated for the sample sectors. Those complexity variables that were computed, that is, proportion of climbing and descending aircraft, average vertical distance between aircraft pairs, and aircraft density, did not show particularly strong correlations with communications measures. Finally, it must be acknowledged that the sample size in this study was quite small, with a maximum number of data points of 36 (3 sectors x 2 samples x 6 10-minute epochs) only marginally sufficient for regression analysis.

Nevertheless, this research may serve as an example for future validation efforts of various metrics of ATC complexity and controller taskload and workload. It should be kept in mind, however, that the aforementioned constructs are themselves complex and often involve multiple interactions, and hence simple measures may reveal only a partial picture of the situation.

#### Acknowledgments

This paper is based upon work supported by the Federal Aviation Administration under Collaborative Agreement No. 02-G-019. Dr. Carol Manning was the scientific/technical monitor. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the FAA.

#### References

- Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25(6), 623-641.
- Chatterji, G., & Sridhar, B. (2001). Measures of air traffic controller workload prediction. *First AIAA Aircraft Technology, Integration, and Operations Forum* (AIAA Paper 2001-5242). Los Angeles, CA, Oct. 16-18, 2001.
- Federal Aviation Administration (1991). *Multiple Virtual Storage (MVS); Subprogram Design Document; National Track Analysis Program (NTAP)*. (NASP-9247-PO2). Washington, DC: Author.
- Federal Aviation Administration (1993). *Multiple*

- Virtual Storage (MVS); User's Manual; Data Analysis and Reduction Tool (DART)*. (NASP-9114-H04). Washington, DC: Author.
- Federal Aviation Administration (2000). *Dynamic density—A review of proposed variables*. NAS Advanced Concepts Branch, ACT-540, William J. Hughes Technical Center.
- Hopkin, V. D. (1995). *Human factor in air traffic control*. Bristol, PA: Taylor & Francis.
- Hurst, M. W., & Rose, R. M. (1978). Objective workload and behavioural response in airport radar control rooms. *Ergonomics*, 21(7), 559-565.
- Knecht, W., Smith, K., & Hancock, P. A. (1996). A dynamic conflict probe and index of collision risk. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 106-110). Santa Monica, CA: HFES.
- Kopardekar, P., & Magyarits, S. (2002). Dynamic density: Measuring and predicting sector complexity. Paper presented at the 21<sup>st</sup> Digital Avionics Systems Conference, Irvine, CA.
- Laudeman, I. V., Shelden, S. G., Branstrom, R., & Brasil, C. L. (1998). *Dynamic density: An air traffic management metric* (NASA-TM-1998-112226). Moffett Field, CA: NASA Ames research Center.
- Manning, C. A., Mills, S. H., Fox, C. M., & Pfliegerer, E. (2000). Investigating the validity of performance and objective workload evaluation research (POWER). *Paper presented in the 3<sup>rd</sup> USA/Europe Air Traffic Management R & D Seminar*, Naples, Italy, June 13-16, 2000.
- Manning, C. A., Mills, S. H., Fox, C. M., & Pfliegerer, E. (2001). Investigating the validity of performance and objective workload evaluation research (POWER). DOT/FAA/AM-01/10. Washington, DC: Office of Aerospace Medicine.
- Manning, C. A., Mills, S. H., Fox, C. M., Pfliegerer, E. M., & Mogilka, H. J. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload*. (DOT/FAA/AM-02/4). Washington, DC: FAA Office of Aerospace Medicine.
- Meshkati, N. (1988) Toward Development of a Cohesive Model of Workload.. In Hancock, P.A. and Meshkati, N. (Eds.) *Human Mental Workload*. Elsevier.
- Mills, S. H., Manning, C., & Pfliegerer, E. M. (1999). Computing en route baseline measures with POWER. *Poster presented at the 10<sup>th</sup> International Symposium on Aviation Psychology*, Columbus, OH, May 3-6, 1999.
- Pawlak, W. S., & Brinton C. R. (1996). Issues in free flight: Results from controller-in-the-loop simulations. *Air Traffic Control Association Conference*. Alexandria, VA: ATCA.
- Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996). A framework for the evaluation of air traffic control complexity. *AIAA, Guidance, Navigation and Control Conference* (San Diego, CA, July 29-31, 1996; AIAA Paper 96-3856). Reston, VA: AIAA.
- Potterfield, D. H. (1997). Evaluating controller communication time as a measure of workload. *The International Journal of Aviation Psychology*, 7(2), 171-182.
- Smith, K., Scallen, S. F., Knecht, W., & Hancock, P. A. (1998). An index of dynamic density. *Human Factors*, 40(1), 69-78.
- Sridhar, B., Sheth, K. S., & Grabbe, S. (1998). Airspace complexity and its application in air traffic management. *Proceedings of the 2<sup>nd</sup> USA/Europe Air Traffic Management R&D Seminar*.
- Vidulich, M. A. (2003). Pilot actions and tasks: Selections, execution and control. In P. S. Tsang and M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 115-146). Mahwah, NJ: Lawrence Erlbaum.
- Wickens, C. D., Mavor, M., & McGee, J. (1997). *Flight to the future: Human factors of air traffic control*. Washington, DC: National Academy of Science.
- Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control*. Washington, DC: National Academy of Science