DEVELOPMENT OF AN EMPIRICALLY-BASED INDEX OF AIRCRAFT MIX

Elaine M. Pfleiderer
FAA Civil Aerospace Medical Institute

The present study is part of an ongoing effort to identify objective predictors of subjective air traffic controller workload. The study begins with a comparison of the salient variables governing en route controllers' perceptions of the performance capabilities of a sample of aircraft and the actual performance of the aircraft in the en route environment. A group of 24 Certified Professional Controllers (CPCs) from Kansas City (N = 17) and Boston (N = 7) en route centers provided estimates of cruising speed, climb, and descent rates for a sample of 24 aircraft types. A matrix of squared Euclidean distances derived from summary measures (i.e., means of estimated speed, climb, and descent rates) was used to construct a classical multidimensional scaling (CMDS) model representing controllers' perceptions of the performance capabilities of each aircraft type. A second matrix was derived from means of speed, climb, and descent rates for the same 24 aircraft types computed from a sample of live air traffic data collected from the Kansas City and Boston en route centers. This matrix was used to construct a second CMDS model representing actual aircraft performance. Interpretation of the dimensions of the CMDS model of ATC estimates suggested that Dimension 1 was related to engine type, whereas Dimension 2 was primarily associated with aircraft weight class. In the model of SAR data, both engine type and weight class were predominantly associated with Dimension 1. Results are used to develop a measure of aircraft mix (i.e., the mix of aircraft with different performance characteristics) to be added to a suite of controller activity and taskload measures.

Introduction

Aircraft mix has been proposed as one of the traffic characteristics that contributes to sector complexity in en route air traffic control (Robertson, Grossberg, & Richards, 1979; Federal Aviation Administration [FAA], 1984; Grossberg, 1989; Mogford, Murphy, Roske-Hofstrand, Yastrop, & Guttman, 1994). "Sector complexity" describes static and dynamic characteristics of the air traffic control environment that combine with controller taskload (i.e., the air traffic events to which the controller is exposed) to produce controller workload (i.e., the controllers' reaction to and perceived effort involved in managing these events) (Grossberg, 1989; Manning, Mills, Fox, Pfleiderer, & Mogilka, 2001). As changes are introduced into the air traffic control environment such as the recent implementation of the Display System Replacement (DSR), or the proposed introduction of "free flight" (Radio Technical Commission for Aeronautics [RTCA], 1995) it becomes increasingly important that measures are developed to evaluate the impact of these changes on performance. In spite of a growing body of work dedicated to the measurement of workload, taskload, sector complexity, and controller performance (for a list of 162 of these measures, see Hadley, Guttman, & Stringer, 1999) little attention has been focused on quantifying aircraft mix. This is possibly because, until recently, aircraft mix had not been clearly defined.

Pfleiderer (2000) conducted an investigation of the salient features of aircraft mix as it relates to aircraft performance characteristics. For this analysis, 30 Certified Professional Controllers (CPCs) from various Air Route Traffic Control Centers (ARTCCs) across the United States provided average speed, climb, and descent rate estimates for a sample of 30 distinct aircraft types. A matrix of squared Euclidean distances derived from summary estimates (i.e., means of speed, climb, and descent) was used to construct a classical multidimensional scaling (CMDS) model of the aircraft. Multiple regression interpretation of the two-dimensional solution revealed that Dimension 1 was related to engine type, whereas Dimension 2 was associated with weight class. The results of the analysis were interpreted as evidence of performance-based prototypes. However, it was also evident from the position of the elements (i.e., aircraft types) in the derived stimulus space that it might be possible to develop a measure of aircraft mix using these two easily-obtained variables.

The present study is a continuation of that investigation (i.e., Pfleiderer, 2000). Phase I was designed to determine whether controllers' perceptions of aircraft performance and the actual recorded performance of aircraft were comparable (i.e., would demonstrate similar dimensionality in repeated CMDS analysis). For this analysis, a matrix of squared Euclidean distances of controller estimates of mean speed, climb, and descent rates for 24 distinct aircraft types was compared with a matrix of mean speed, climb, and descent rates of the same aircraft types calculated from routinely-recorded System Analysis Recording (SAR) data. It was expected that the two dimensions noted in the previous CMDS analysis of controllers' perceptions would be the same as those in the SAR sample, but it was possible that the two matrices might differ with regard to the relative salience and importance of each dimension. Characteristics of the CMDS model of SAR data could be used to confirm, amend, or replace previously-gathered information regarding the salient features of aircraft mix.

Phase II focused on the development of an index of aircraft mix based on the results of the Phase I multidimensional scaling analyses. Because multidimensional scaling translates patterns of responding into patterns of elements in a dimensional space, it should be possible to assign base values to aircraft and then calculate distances representing differences in performance capabilities to compute an aircraft mix index.

Finally, the aircraft mix index was computed for all aircraft present in a particular traffic sample. If the index has sufficient variability and precision, it should be able to discriminate between low-altitude sectors (i.e., sectors with a high probability of aircraft with disparate performance capabilities) and high-altitude sectors (i.e., sectors with a low probability of aircraft mix due to the relatively lower service ceilings of many piston-driven aircraft). If the aircraft mix index passes the "discriminability" test, future research will be conducted to determine whether or not it adds unique information to an existing suite of Performance and Objective Workload Evaluation Research (POWER) measures (Mills, Pfleiderer, & Manning, 2002). It is possible that the complexity associated with aircraft mix is redundant with other variables. It is also possible that aircraft mix is characteristic of so few sectors so as to be of little use within the larger suite of measures. One thing is certain: Aircraft mix's relative contribution to sector complexity and controller workload cannot be assessed until it has been quantified.

Phase I: Method

Design and Procedure

Multidimensional scaling refers to a group of descriptive procedures that transform proximity data into mapped elements in one or more spatial dimensions (Kruskal & Wish, 1978). In this application, two matrices of dissimilarity measures were analyzed: One matrix was based on summary controller estimates of aircraft performance, the other was based on summary measures of aircraft performance derived from SAR data.

Controller Estimate Matrix. This matrix represents a subset of the data used in a previous study (Pfleiderer, 2000) in which 30 Certified Professional Controllers (CPCs) provided estimates of average cruising speed, climb rate, and descent rate for each of 30 distinct aircraft types. In the present study, mean speed, climb, and descent rate estimates were calculated from estimates provided by 24 of the original 30 controllers for 24 of the original aircraft types. The subset of 24 controllers was selected from the larger sample because these CPCs met currency requirements at the same ARTCCs represented in the SAR sample: Kansas City (N = 17) and Boston (N = 7). The aircraft list was reduced because 6 of the 30 aircraft types did not appear in the Kansas City or Boston airspace during the time sampled. For more information about matrix construction, participants' professional experience, detailed descriptions of the materials used to collect estimates, and other points of methodology, see Pfleiderer (2000).

SAR Data Matrix. The information used to construct this matrix was recorded at the Kansas City and Boston centers. The Kansas City sample consisted of 168 hours of continuous SAR data recorded from January 19, 1999 through January 24, 1999. The Boston sample comprised a total of 27 hours of SAR data, recorded on March 16, 1998 from 14:00 to 20:59 ZULU (7 hours); March 17, 1998 from 14:00 to 20:59 ZULU (7 hours); March 19, 1998 from 15:00 to 19:59 ZULU (5 hours); and March 20, 1998 from 15:00 to 22:59 ZULU (8 hours). Raw data were extracted through the use of "log" and "track" reports produced by the Data Analysis Reduction Tool (DART). Within the sample time frame, 7095 flights corresponded to the selected aircraft types. The modal flight duration of these flights was 27 minutes, which translates to approximately 270 speed and altitude updates (observations) per flight.

Aircraft type was derived from designators (alphanumeric labels that indicate the make and model of an aircraft) that are printed on the flight progress strip (FPS) and appear within the flight plan readout display. The contents of both flight progress strips and flight plan readouts are recorded by the Host system and output in the DART log report.

Mean climb and descent rate estimates were calculated from altitude information recorded in the DART track reports. Climb and descent rate estimates represent the amount of change divided by duration of change for all detected altitude changes converted into feet per minute (fpm) and then averaged across changes for each aircraft (set to missing if no altitude changes were detected). For example, from 8:12:08 to 8:17:02 flight XMPL01 climbed from 26,400 feet to 33,000 feet – a total of 6,600 feet in 4 minutes and 54 seconds (1,454 fpm). From 8:30:00 to 8:35:00, XMPL01 climbed from 33,000 feet to 35,000 feet – a total of 2,000 feet in 5 minutes (400 fpm). The climb rate estimate for XMPL01 would then be 927 fpm (the mean of the two changes.) Mean climb and descent rate estimates were calculated in this manner for each flight and then averaged across flights for each designator. Of course, not all flights made altitude changes during the time sampled, and so the number of observations used to calculate mean climb and descent rates varied between aircraft designators.

Mean speed estimates were calculated by first computing the mean of all ground speeds recorded in the DART track report for each flight (distinguished by a unique Aircraft Identifier [AID] Computer Identifier [CID] combination) and then averaging across flights for each designator. The number of updates per flight varied as a function of control time. However, the number of speed estimates used to compute the average speed for each designator is equal to the number of aircraft corresponding to that designator. Please note that, unlike the computation of mean climb and descent rates, mean speed calculations did not involve speed changes. The measure simply represents the average speed for each aircraft type based on the average ground speed for all individual aircraft of that type.

Squared Euclidean distances were calculated from mean speed, climb, and descent rates using SPSS procedure PROXIMITIES. Distances in the resulting matrix represented each aircraft's performance relative to other aircraft in the sample.

Variables for Interpretation of CMDS Models. A separate set of variables was collected for the purpose of interpreting the dimensions of the CMDS models. The engine number, engine type, and weight class of each aircraft was obtained from information provided in Appendix A of 7110.65N, the most recent version of the *Air Traffic Control* (FAA, 2002).

Phase I: Results and Discussion

In general, the configurations (Figures 1 and 2) were similar. In both models, Group A consisted primarily of piston-driven aircraft. The exception to this was the C208 (Cessna Caravan), a turboprop that did not perform like other turboprops. (As a point of interest, most of the controllers in the sample misclassified the C208 as a piston-driven aircraft.) Group B consisted entirely of turboprops. In both configurations, all aircraft positioned to the right of the dashed gray line are jets. However, in the model of ATC estimates (Figure 2) high-performance jets (Group D) are clearly distinguished from other jet aircraft. In the SAR data model (Figure 1), jets formed a single, loosely-knit group (Group C).

Perhaps the most striking difference between the configurations had to do with weight class. Most of the aircraft types in the top portion of the stimulus configuration of ATC estimates (Figure 2) are of the Small and Large weight classes: Heavy aircraft are positioned in the bottom portion of the configuration. In the stimulus configuration of SAR data (Figure1), Heavy aircraft are scattered throughout the cluster of jets (Group C).

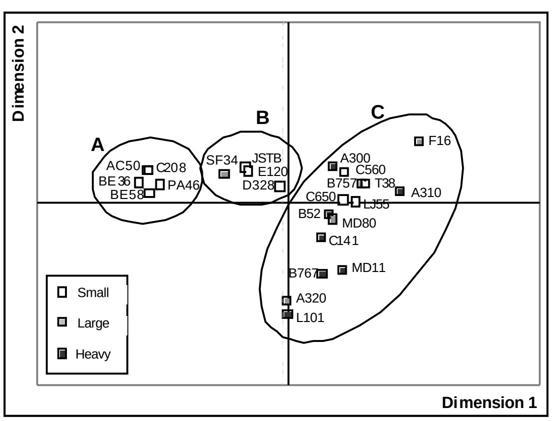

Figure 1. Derived Stimulus Configuration of the Two-Dimensional Multidimensional Scaling Model of SAR Data

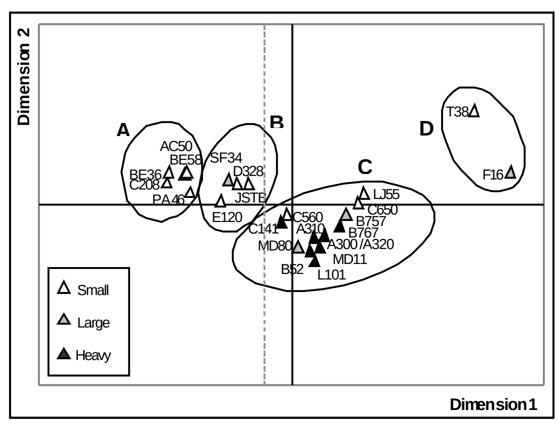

Figure 2. Derived Stimulus Configuration of the Two-Dimensional Multidimensional Scaling Model of ATC Estimates

Table 1. *Summary of Multiple Regression Analysis Interpretation of the Characteristics of the Two-Dimensional CMDS Model of SAR Data*

Criterion	R	R^2	F	p	$ß_1$	$ß_2$
Engine Type	.94	.89	85.92	.00	.87*	-.25*
Weight Class	.80	.65	19.80	.00	.62*	-.44*
Engine Number	.31	.09	1.65	.22	.20	-.28

* $p<.01$

Table 2. *Summary of Multiple Regression Analysis Interpretation of the Characteristics of the Two-Dimensional CMDS Model of ATC Data*

Criterion	R	R^2	F	p	$ß_1$	$ß_2$
Engine Type	.94	.88	80.34	.00	.80*	-.48*
Weight Class	.86	.74	29.16	.00	.48*	-.70*
Engine Number	.43	.18	2.37	.12	.08	.-42

* $p<.01$

The most objective technique available for dimensional interpretation is the regression method in which variables believed to correspond with the stimulus configuration are regressed over coordinates. For this application, engine type was coded according to performance capabilities associated with each engine type, from lowest (piston-driven) to highest (jet propelled). Weight class was also coded into three ordered levels: Small, Large, and Heavy (for a specific description of aircraft weight class categories, see FAA, 2002).

According to Kruskal and Wish (1978), two conditions are necessary for satisfactory multiple regression interpretation of a dimension. First, the multiple correlations must be extremely high (correlations in the .90s are recommended, although those in the .70s will suffice). As shown in Tables 1 and 2, only engine type and weight class achieved the recommended degree of association with the dimensions. In general, the two data sets were remarkably similar: Correlations were in the .90's for engine type and in the .80's for weight class. However, the two models differed with respect to the relationship between the dimensions and these variables. Notice that in the model of SAR data, the standardized regression weights of both engine type and weight class are more closely associated with Dimension 1 than with Dimension 2. However, in the configuration derived from ATC estimates, the standardized regression weights of engine type were more closely associated with Dimension 1, whereas weight class was more closely associated with Dimension 2.

Phase I: Conclusions

The clusters of aircraft identified in the neighborhood interpretation of the stimulus configurations present the simplest means by which to code aircraft types for the aircraft mix variable. For the most part, these groups were defined by engine type. Though high-performance jets were not clearly distinguished from other jets in the configuration of SAR data, it seems reasonable to classify these aircraft separately in the computation of the aircraft mix index. On the average, the controllers who contributed estimates for the ATC sample had approximately 10 years of experience at their current ARTCCs. The SAR sample represented 195 hours of traffic. Given the concordance of the two matrices in other respects, it is possible that high-performance jets might have emerged as a separate group in the SAR configuration had the sample been large enough to better approximate the years of experience represented by the controllers in the ATC sample.

It is unlikely that the incorporation of weight class is crucial to the precision of the aircraft mix index. To begin with, weight class is a correlate of engine type (i.e., most piston-driven aircraft are Small, most turboprops are Large, all Heavy aircraft are jets). Because of the nature of this relationship, incorporation of the weight class dimension would only involve jet aircraft (i.e., separating jets into Heavy/other subgroups). However, the tight clustering of the jet aircraft in Group C of the stimulus configuration of ATC estimates (Figure 2) suggests that this differentiation is probably not necessary. Heavy aircraft may perform somewhat differently than other jet aircraft, but this difference appears to be only slightly perceptible to air traffic controllers (other than procedural considerations addressing the wake turbulence associated with Heavy aircraft and B757s).

Phase II: Method

Sample

The sample selected for testing the aircraft mix index consisted of SAR data from 15 high-altitude sectors and 13 low-altitude sectors within the Kansas City airspace. The Kansas City ARTCC was selected because of the availability of sector information for that particular center (e.g., sector strata, number of underlying airports, sector combinations). The data were recorded on Friday, December 22, 1999 from 15:15 to 16:15 (local time) when most sectors within the Kansas City en route center were open (i.e., sector combinations were minimal).

Procedure

Based on Phase I results, aircraft were assigned aircraft type codes with values ranging from one to four. Piston-driven aircraft were assigned a value of 1, turboprops a value of 2. With some exceptions, jet aircraft were assigned a value of 3. High-performance jets (i.e., aircraft types that perform within similar parameters as the aircraft in Group D of Figure 2) are coded as such in the system files of all en route Host computers. These file codes were used to assign a value of 4 to all high-performance jets in the sample. Then, aircraft type differences were calculated between pairs of aircraft within a given sector to create a half matrix of differences. For example, DAL589 is a commercial jet assigned a base code of 3. N149RJ is a turboprop with a base code of 2. The aircraft type difference between N149RJ and DAL589 is 1. The final step in the computation of the index involved summing all items in the half matrix.

For each minute of data, the aircraft mix index was calculated for all aircraft pairs at approximately 12-second intervals and stored in an array. At the end of each minute, the mean and standard deviation of the aircraft mix measure were calculated and sent to an array for the purpose of calculating the mean and standard deviation of the aircraft mix measure for each 15-minute interval processed.

Phase II: Results and Discussion

Computing aircraft mix at 15-minute intervals for high and low altitude sectors did not produce a normal distribution. For that reason, the Mann-Whitney *U* statistic (Mann & Whitney, 1947) was employed to examine whether the aircraft mix index was reliably different in high versus low altitude sectors. The Mann-Whitney *U* is a distribution-free statistic that tests the null hypothesis that two sets of observations were sampled from identical populations. The minimal assumption of the Mann-Whitney *U* is the independence of observations (Marascuilo & McSweeney, 1977). Given the fact that aircraft cannot be controlled by more than one sector at any given time, it logically follows that aircraft in one sector were independent of aircraft in another within each of the 15-minute intervals.

As shown in Table 3, the null hypothesis was rejected for all comparisons, indicating that the distributions of the aircraft mix index were reliably different in high- versus low-altitude sectors. The sum of ranks assigned to each of the original values for high and low altitude sector groups is consistent with the expectation that the mix of aircraft with different performance characteristics (ergo the aircraft mix index) would be higher in low altitude airspace.

Table 3. Mann-Whitney *U* Tests for Aircraft Mix Index (by Interval)

	N	**Mean Rank**	**Sum of Ranks**	***U***
Interval 1				
High Altitude	15	9.27	139.00	19.00*
Low Altitude	13	20.54	267.00	
Interval 2				
High Altitude	15	9.00	135.00	15.00*
Low Altitude	13	20.85	271.00	
Interval 3				
High Altitude	15	9.13	137.00	17.00*
Low Altitude	13	20.69	269.00	
Interval 4				
High Altitude	15	8.87	133.00	13.00*
Low Altitude	13	21.00	273.00	

* Asymptotic significance (2-tailed) <.01

Phase II: Conclusions

Because the aircraft mix index was able to reliably detect distribution differences in high- and low-altitude sectors, the measure warrants further investigation. Plans for future tests include conducting a CMDS analysis of a much larger sample of aircraft types to include helicopters and other rotorcraft to determine whether they fit into one of the existing aircraft categories or require the introduction of a separate code. Then, the aircraft mix index will be introduced into the current set of Performance and Objective Workload Evaluation Research (POWER) variables (Mills, Pfleiderer, & Manning, 2002) to determine whether or not the aircraft mix index adds unique information to that set. Each step in this process brings us closer to determining the relative contribution of aircraft mix to sector complexity. Constructing the elements that create sector complexity may help us understand the nature of controller workload, and thus provide insight into the relationship between controller workload and performance.

References

Federal Aviation Administration (1984). *Establishment and validation of en route sectors*. (FAA Order No. 7210.46). Washington, DC: Author.

Federal Aviation Administration (2002). *Air traffic control.* (FAA Order No. 7110.65N) Washington, DC: Federal Aviation Administration Air Traffic Operations Program. *Available online*: http://www.faa.gov/atpubs/ATC.

Grossberg, M. (1989). Relation of sector complexity to operational errors. *Quarterly Report of the Federal Aviation Administration's Office of Air Traffic Evaluations and Analysis*. Washington, DC: Federal Aviation Administration.

Hadley, G.A., Guttman, J.A., & Stringer, P.G. (1999, June). *Air traffic control specialist performance measurement database.* (Report No. DOT/FAA/CT-TN95/45).

Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling.* Beverly Hills, CA: Sage.

Mann, H.B., & Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, p. 50-60.

Manning, C.A., Mills, S.H., Fox, C., Pfleiderer, E., & Mogilka, H. (2001). *Investigating the validity of performance and objective workload evaluation research (POWER).* (Report No. DOT/FAA/AM-01/10). Washington, DC: Office of Aerospace Medicine.

Marascuilo, L.A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.

Mills, S.H., Pfleiderer, E.M., & Manning, C.A. (2002). *POWER: Objective activity and taskload assessment in en route air traffic control.* (Report No. DOT/FAA/AM-02/02). Washington, DC: Office of Aerospace Medicine.

Mogford, R.H., Murphy, E.D., Roske-Hofstrand, R.J., Yastrop, G. & Guttman, J.A. (1994). *Application of research techniques for documenting cognitive processes in air traffic control: Sector complexity and decision making.* (Report No. DOT/FAA/CD-TN94/3). Atlantic City, NJ: Federal Aviation Administration.

Pfleiderer, E. M. (2000). *Multidimensional scaling analysis of controllers' perceptions of aircraft performance characteristics*. (Report No. DOT/FAA/AM-00/24). Washington, DC: Office of Aviation Medicine.

Robertson, A., Grossberg, M., & Richards, J. (1979). *Validation of air traffic controller workload models.* (Report No. DOT/FAA/RD-79/83). Cambridge, MA: U.S. Department of Transportation Research and Special Programs Administration – Volpe National Transportation System Center.

RTCA (1995). *Final Report of RTCA Task Force 3: Free Flight Implementation*, Washington, DC: RTCA Incorporated.