RELATIONSHIPS BETWEEN MEASURES OF AIR TRAFFIC CONTROLLER VOICE COMMUNICATIONS, TASKLOAD, AND TRAFFIC COMPLEXITY

Carol Manning, Cynthia Fox, & Elaine Pfleiderer

FAA Civil Aerospace Medical Institute, Oklahoma City, OK

A study was conducted to replicate a previous study in which air traffic control taskload measures were found to predict subjective workload better than a set of communications measures. In this study, two measures of the timing of communications were computed along with two principal components describing communication content. Five principal components were derived to describe ATC taskload. These variables were used to predict criterion traffic complexity ratings. Although the taskload measures appeared to predict complexity ratings better than the communication measures, it is difficult to formulate a definite conclusion from the data. This study too had limitations, discussed here, that should be addressed in future research.

Introduction

It has been proposed that the amount and type of voice communications between air traffic controllers and pilots is a good indicator of a controller's workload [1, 2]. Reasons for this assertion are that pilot/controller communications increase 1) as a function of the number of aircraft in the airspace, 2) when the air traffic situation is complicated (such as when weather is present), and 3) in airspace utilizing complex procedures (such as arrival and departure sectors). However, a recent study [3] found that a set of objective taskload measures (called Performance and Objective Workload Evaluation Research, POWER) [4] predicted air traffic controller subject matter expert (SME) ratings of other controllers' subjective workload about as well as an extensive set of voice communication measures. This result suggests that extracting objective taskload measures derived from routinely recorded ATC data may replace the need to analyze voice communication measures for the purpose of estimating subjective workload. Of course, there are other reasons to analyze pilot/controller communications that do not involve assessing subjective workload.

Some limitations associated with the study described in [3] might affect the generalizability of their results. First, the data came from a restricted number of traffic samples and sectors in the Kansas City Center, so the results might fail to generalize to other facilities or sector types. The results might also be different if more data were available for analysis. Second, the traffic samples consisted of a generally low level of activity. It is possible that results might not generalize to sectors and traffic samples with high activity levels. Third, subject matter experts (SMEs) rated the subjective workload that they thought someone else had experienced during a traffic sample, based on their observation of SATORI re-creations [5]. Perhaps rating someone else's workload rather than one's own fails to capture some important components of subjective workload.

An opportunity arose to attempt to replicate the communications study [3] when data from a joint FAA Technical Center/NASA project [6] became available to us. These data were collected from several en route facilities to assess the ability of several metrics to measure Dynamic Density (DD), defined as "the air traffic complexity or difficulty of a situation" [6]. During the DD study, traffic samples from the facilities were re-created using SATORI [5]. Controllers and supervisors from the respective sectors observed re-creations of the traffic samples and provided subjective ratings of their complexity. In the original study, these ratings were used as criteria against which the DD metrics were compared.

For the present study, SATORI files from the DD study were used to prepare and code transcripts of voice communications that occurred in the traffic samples. The original System Analysis Recording (SAR) files for the DD traffic samples were analyzed using the Data Analysis and Reduction Tool [7] to produce POWER measures [4]. The subjective ratings of traffic sample complexity provided for the DD study also served as criterion measures for this study. Instead of predicting complexity ratings using alternative DD metrics, we used communications and taskload measures to predict the complexity of the traffic samples. While the traffic complexity ratings were not deliberately obtained to assess subjective workload, it is likely that these measures of the situations' difficulty were similar enough to subjective workload to be considered acceptable equivalents.

The purpose of this study was to examine the relationship between measures of communication, objective taskload, and subjective complexity. We proposed two hypotheses about the relationships between these measures. First, we expected that measures derived from the communication transmissions would be significantly correlated with each other and with activity, as measured both by complexity ratings and the POWER taskload measures. As the traffic situation gets busier, more communications should occur, and more time should be spent communicating. Second, we expected that the measures associated with the communications would not make a unique contribution to the prediction of complexity, over and above the contribution provided by the POWER taskload measures. Thus, we expected that the POWER taskload measures alone would account for most of the variance in the complexity ratings and this prediction would not improve significantly by adding communications measures to the set of predictors.

Method

This study examined statistical relationships between measures of pilot/controller communications, objective ATC taskload, and subjective complexity. The communication and taskload measures were obtained from samples of routinely-recorded ATC data for a set of traffic samples. The complexity ratings were provided by subject matter experts (SMEs) who observed graphical re-creations of the traffic samples and rated their complexity. Each component of the study is discussed in more detail below.

Traffic Samples

SAR data and associated voice communication files were obtained for 18 traffic samples recorded between July and September of 2001 at 9 sectors (sectors 3, 4, 6, 19, 20, 22 38, 39, and 41) at the Atlanta Air Route Traffic Control Center (ARTCC). Half were high-altitude and half were low-altitude sectors.

Participants

Participants were controllers and supervisors who worked in the areas of specialization where the sectors were located. At least two controllers and two supervisors provided complexity ratings for each traffic sample.

Subjective complexity estimates

Six participants, three CPCs and three Area Supervisors, were asked to provide air traffic complexity ratings, on a 1-7-point scale (1=low, 7=high), using an electronic keypad. The method used to obtain the complexity ratings was similar to that used to obtain subjective workload ratings in real time using the Air Traffic Workload Input Technique (ATWIT) [8]. The ATWIT presents auditory and visual cues that prompt a controller to press one of seven buttons within a specified amount of time to indicate the amount of mental workload experienced at that moment. In this study, instead of using ATWIT to rate their own workload, the participants instead rated the complexity of the traffic sample they had just observed. Participants were prompted every two minutes during each traffic sample to provide a complexity rating.

Objective taskload measures

ATC data were extracted from the SAR files using the Data Analysis and Reduction Tool [7]. The resulting files were processed by the POWER program [4], which uses data from a subset of the DART files to compute a set of taskload measures.

The POWER measures included information about the number of controlled aircraft; handoffs made and accepted; controller data entries and data entry errors; numbers and timing of aircraft heading, speed, and altitude changes; and the average time aircraft were under control. In all, 15 POWER measures were utilized in this study.

Communication measures

Communication measures were based on transmissions obtained from voice tapes associated with the traffic samples. The measures analyzed in this study included the total number of transmissions, total time spent communicating during a traffic sample, and time required for individual transmissions (for all speakers, and analyzed separately for the controller and all other speakers). The transmissions were also categorized by content. Content categories were based on a set derived previously [9]. The content categories were 1) Address, 2) Courtesy, 3) Instructional Clearances, 4) Frequency Changes, 5) Advisory/Remark, 6) Request, 7) Readback/Acknowledgment, and 8) Non-codable. Due to concerns expressed during the ATM2001 R&D Seminar, the overall clearance category derived in [9] was separated into two sub-categories, Instructional Clearances and Frequency Changes, to distinguish between clearances instructing an aircraft to proceed (likely to result in higher workload and complexity) and more routine instructions for the pilot to change the radio frequency when leaving the sector (likely to result in lower workload and complexity). More than one message content may have occurred during a transmission; all were coded. Communications were not otherwise coded by specific message types (e.g., altitude or heading clearance) and errors (e.g., transposed numbers/letters) were also not coded.

Procedure

All transmissions that occurred during each traffic sample were transcribed. Message content of each transmission was categorized, along with the identity of the speaker (i.e., controller, pilot, other speaker) and the start and stop times. These data were used to compute the total number of communications and total time spent communicating during each 6-minute period, the mean time for individual transmissions, and content codings for each transmission.

Fifteen POWER measures were computed for the five 6-minute segments included in each of the 18 experimental traffic samples. Complexity ratings, which were obtained every 2 minutes, were averaged across three 2-minute periods to obtain a six-minute average complexity. Then the 6-minute complexity ratings were averaged across all participants.

We started with 90 observations (18 traffic samples x 5 6-minute segments per traffic sample.) Some data were missing from the SAR or communications files (or both). If any missing data were present in any of the files, then the associated 6-minute segment of the traffic sample was eliminated from analysis. Eleven such segments were excluded, leaving only 79 observations available for analysis.

Results

Complexity

The mean complexity rating, when averaged across the 79 six-minute time periods, ranged between 1.0 and 6.56. The average complexity rating was 3.42 (SD = 1.27). As was true in [3], this value was significantly lower than 4 ($t(78) = -4.10$, $p < .001$). As the value of 4 is the midpoint of the 7-point scale used to measure complexity, and because about 64% of the average complexity ratings were less than or equal to 4, the results suggest that participants thought the average complexity for these traffic samples was lower than average.

Communication measures

Overall, 4,929 transmissions (or, on the average, about 62.4 (SD=15.28) per 6-minute segment) were recorded. Two thousand seventy-three of these (43%) were made by a controller, and 2,773 (57%) were made by another speaker (pilot or other controller.) The average number of transmissions that occurred during a 6-minute period was 62.39 (SD = 15.28). Controllers made, on the average, 26.24 (SD = 8.42) of these transmissions, while other speakers made 35.1 (SD = 9.66).

On the average, the total amount of time spent communicating during a 6-minute period was 214.11 seconds (SD = 49.69), or about 59% of the total time available. In [3], only about 29% of the total time was spent communicating. Controllers spent, on the average, 103.56 seconds (SD = 31.63) speaking during each 6-minute period, while others spoke for an average of 108.29 seconds (SD = 26.50).

The average duration of each individual communication event was 3.43 seconds (SD = 1.93). This is longer than the previous finding of an average of 2.86 seconds (SD = 0.63) from [3]. Controllers' transmissions lasted, on the average, 3.95 seconds (SD = 2.10), while other speakers' transmissions lasted, on the average, 3.09 seconds (SD = 1.69).

Table 1 shows correlations of communication measures computed during the 4-minute segments. As in the previous study, the total number of communications was highly correlated with total time spent communicating. The total number of communications also had a significant negative correlation with mean time for individual communications. However, total communication time was not significantly correlated with mean time for each communication. Because total number of communications and total communication time were so highly correlated, total number of communications was eliminated from further analysis.

Table 1. Correlations of communication measures.

	Total N	Total time	Mean time
Total number of communication	1.0		
Total communication time	.85**	1.0	
Mean time for single communication	-.36**	.17	1.0

** Correlation is significant at p <.01 level

Communication content

The average number of communication topics that occurred in a 6-minute period is shown in Table 2. Because each transmission may have included more than one topic of conversation, transmissions may have included more than one content category. Thus, the number of times the content categories occurred in a 6-minute time period could be greater than the total number of transmissions.

Table 2. Descriptive statistics for communication content categories.

Communication content	Mean	SD
Address	54.6	15.7
Instructional clearances	12.30	6.9
Frequency changes	14.41	6.5
Advisory	19.04	8.0
Readback	20.29	7.7
Courtesy	9.09	4.8
Request	3.66	2.7

Addresses occurred most frequently, on the average, about 16 times in a 6-minute period. Readbacks occurred about 20 times and advisories occurred about 19 times per period. Requests and courtesies occurred least often. Non-codable communications were not reported here and were excluded from all subsequent analyses.

A Principal Components Analysis (PCA) with Varimax rotation was conducted to identify a smaller set of components that would describe the relationships between the communication content categories more concisely. The PCA produced two components with eigenvalues greater than 1. These components accounted for 67.8% of the variability in the data set. The rotated component matrix is shown in Table 3. For ease of interpretation, correlations less than .3 were excluded from the table.

Table 3. Rotated component matrix for communication content categories.

Communication content	Comp 1: Readback/ Instructional clearance	Comp 2: Courtesy/ Frequency Change
Readback	.93	
Instructional clearance	.91	
Address	.86	.43
Advisory	.46	.58
Request	.40	.52
Courtesy		.85
Frequency change		.60

Note: Correlations less than .4 are not shown.

Readbacks and instructional clearances had the highest correlation with component 1, although addresses, advisories, and requests were also correlated. Thus, component 1 was labeled Readback/Instructional Clearance. Courtesies and frequency changes had the highest correlation with component 2, although addresses, advisories, and requests were also correlated. Thus, component 2 was labeled /Courtesy/Frequency Change. These results are similar to those obtained in the Manning (2002) study, although 3 components were identified. However, two of the three components in that study were very similar to the two found here.

Taskload

Table 4 shows descriptive statistics for the 15 POWER measures used in the study, averaged across the 6-minute periods in each traffic sample. Some of the POWER measures included in this study are different than those included in [3] because some of the variables were recoded. For example, heading, speed and altitude changes, duration, and amounts replaced average the measures of heading, speed, and altitude variability used in [3]. Also, some new variables were added (for example, measures of visual clustering and horizontal, vertical, and Euclidean distances between aircraft). Although

POWER computes several, for the purpose of this study, individual data entries were excluded from the analyses and only total data entries and data entry errors were included here.

Table 4. Descriptive Statistics for 24 POWER measures computed at 6-minute intervals.

Power Measures	Mean	SD
Total N aircraft controlled	13.92	3.53
Average time aircraft under control	244.05	38.64
Visual clustering	6.00	2.90
Average N altitude changes	6.72	2.86
Average altitude change amount	33.37	13.35
Average N heading changes	4.57	3.18
Average heading change amount	38.84	30.73
Average N speed changes	6.87	3.89
Average speed change amount	51.82	24.74
Total N handoffs accepted	4.58	1.89
Total N handoffs initiated	4.38	2.43
N Radar controller data entries	29.46	10.59
N Radar controller data entry errors	1.65	1.62
N Data controller data entries	8.34	6.81
N Data controller data entry errors	0.46	0.78

A second PCA was conducted to identify a smaller set of variables that would describe the relationships between the POWER measures more concisely while reducing their number. The results of this analysis should be interpreted with some caution because some of the 79 6-minute periods contained the same aircraft and so the observations are not completely independent. Nine variables were eliminated from this analysis so that the ratio of the number of available observations to the number of variables analyzed would be near 5, the number necessary to produce a stable result. The PCA produced five components with eigenvalues greater than 1. Table 5 contains the rotated component matrix for the 5 components that accounted for about 72.4% of the variance in the data. The entries in the table are correlations between each POWER measure and the components derived from the analysis, transformed using the Varimax rotation method. For ease of interpretation, correlations less than .4 were excluded from the table.

Component 1 was related to the average number and duration of speed changes, the average number and amount of heading changes, and number of altitude changes. To a lesser extent, component 2 was also related to visual clustering and control duration. Component 2 was labeled Low Altitude Maneuvers because changes in speed and heading consistent with arrivals and departures. This component is similar to one found in the data from [3] that came from Kansas City Center.

Component 2 was primarily related to the number of aircraft controlled, the number of handoffs initiated (indicating aircraft leaving the sector), and visual clustering. To a lesser extent, the component was also related to the R- and D-controller data entries. This component was labeled Activity because higher values for these measures were associated with more aircraft in the sector, requiring more controller activity. A similar component was the most important component in [3].

Component 3 was primarily related to the number of D controller data entry errors and D data entries. Thus, Component 4 was called D Controller Activity. A similar component was also present in the data from [3].

Component 4 was primarily related to R-controller data entry errors, R-controller data entries, and was negatively related to the amount of time aircraft were under control. These conditions were consistent with busy R-controllers moving aircraft quickly, and making more data entry errors. Component 4 was called Overload. A similar component was present in the data from [3].

Component 5 was related to a smaller amount of altitude change and more handoffs accepted. This is consistent with overflight aircraft.

Table 5. Rotated component matrix for 15 POWER measures.

Power Measures	C1	C2	C3	C4	C5
Average amount speed change	.89				
Average N speed changes	.88				
Average N heading changes	.81				
Average amount heading change	.78				
Average N altitude changes	.63				
Visual clustering	.49	.65			
Average time aircraft under control	.42			-.68	
Average N aircraft controlled		.76			
Average N handoffs initiated		.66			
N Data controller data entries		.47	.68		
N Radar controller data entries		.41		.64	
N Data controller data entry errors			.88		
N Radar controller data entry errors				.75	
Average amount altitude change					-.82
Average N handoffs accepted					.65

Prediction of complexity

Table 6 shows correlations between the average complexity ratings, the five POWER taskload components, the two communication timing variables, and the two communication content components. By definition, the orthogonally-rotated principal components are unrelated, so their inter-correlations are 0. The mean complexity rating was significantly related to four of the five POWER taskload components and all the communication measures except for the mean time for a single transmission. The Low Altitude Maneuvers and Activity components were related to all communications measures except for the mean time for a single transmission.

Table 6. ***Correlations of complexity rating, taskload components, and communications measures.***

	Complex	Low alt man.	Activity	D activity	Over load	Over flight	Mean comm time	Total comm time	Read back	Courtesy
Complexity ratings	1.00									
P Comp 1 - Low Alt Maneuvers	.22*	1.00								
P Comp 2 – Activity	.62**	0	1.00							
P Comp 3 - D activity	.37**	0	0	1.00						
P Comp 4 – Overload	.14	0	0	0	1.00					
P Comp 5 - Overflights	.33**	0	0	0	0	1.00				
Mean xmit time	.18	.04	-.05	.10	-.17	.07	1.00			
Total comm time	.55**	.33**	.36**	.14	-.16	.05	.17	1.00		
C Comp 1 – Readback	.51**	.50**	.38**	.10	.11	.16	-.04	.76**	1.00	
C Comp 2 - Courtesy	.37**	-.32**	.36**	.20*	.18	-.14	.09	.45**	0	1.00

** Correlation is significant at the $p < .01$ level.
* Correlation is significant at the $p < .05$ level.

Additional analyses assessed the effectiveness of alternative multiple regression models in predicting the complexity ratings. Predictors were the five POWER taskload components, total communication time, average time for a single communication, and the two communication content components.

Table 7 shows the results of these analyses. The first row shows the multiple correlation for the "full" regression model containing all the predictors mentioned above.

The multiple correlation of the full regression model with the average complexity rating was R = .88, accounting for about 77% of the variance in the complexity ratings. Succeeding lines show multiple correlations between alternative "reduced" regression models containing fewer than the total number of predictors. The column containing F for the test of R^2 change compares the relative effectiveness of an alternative (reduced) model with the full model for predicting the average complexity rating. If the probability is greater than .05 that the change in R^2 between the two models is significantly different from 0, then the reduced model is considered to be as effective as the full model. On the other hand, if the probability is less than or equal to .05 that the change in R^2 between the two models is significantly different from 0, then the reduced model is not considered to be as effective as the full model.

Table 7. Comparison of alternative multiple regression models predicting complexity ratings.

Regression model	**R**	R^2	R^2 change	*F* for test of R^2 change	df	p
1. Full model	0.875	0.766	N/A			
2. All POWER components	0.836	0.699	0.07	4.911	4, 69	0.002
3. POWER Low alt, activity, Dside, overflight components, courtesy comm component, time for single transmission	0.862	0.743	0.02	2.216	3, 69	0.094*
4. All Comm measures	0.655	0.429	0.34	19.819	5, 69	0.000
5. Communication times	0.559	0.312	0.45	19.058	7, 69	0.000
6. Communication content	0.629	0.396	0.37	15.560	7, 69	0.000
7. Activity, readback, D-side	0.767	0.588	0.18	8.701	6, 69	0.000

In the first comparison (Line 2), the reduced model containing only the POWER taskload components had an R^2 of .70, compared with the full model's R^2 of .77. The F computed to assess the R^2 change of .07 had a value of 4.91, and the probability was .002 that the change in R^2 was greater than 0. Thus, the reduced model containing all the POWER taskload components did not predict complexity ratings as well as the full model. Another example is shown on line 3 of Table 6. The reduced model containing three of the five POWER taskload components, the courtesy communication component, and average time for a single transmission had an R^2 of .74, compared with the full model's R^2 of .77. The F statistic computed to compare the R^2 change of .02 had a value of 2.216, and the probability was .094 that the change in R^2 was greater than 0. Thus, the reduced model containing three taskload components and two communication measures predicted the complexity ratings as well as the full model. Four other regression model comparisons are shown in Table 6. None of them predicted the complexity ratings as well as the full model.

Discussion and Conclusions

In a previous study [3], we found that communications measures add only a little to the prediction of subjective workload, over and above the contribution of a set of routinely recorded measures of ATC taskload. Based on the results of [3], we developed a set of related hypotheses that we hoped could be demonstrated for a different set of SAR and ATC voice communications data.

We hypothesized that measures derived from air traffic control communication transmissions would be significantly correlated with each other and with activity, as measured both by complexity ratings and the POWER taskload measures. Second, we expected that the communications measures would not make a unique contribution to the prediction of complexity, over and above the contribution provided by the POWER taskload measures.

Before discussing the results of the study, it is necessary to repeat that the data analyzed were collected for a different purpose than this study. Complexity ratings were obtained to provide a criterion against which to compare a set of Dynamic Density measures. While we presumed that complexity is likely to be related to subjective workload, we were not able to demonstrate this. Thus, the relationships found here may not accurately represent the relationships that might have been found if subjective workload ratings had been provided instead, and had been provided by the controllers who actually worked the traffic.

Nevertheless, we found that several communications measures based on times and content of communications were inter-related except for the average time to make an individual transmission. The other three related communications measures were also related to average complexity ratings, and to two of the five POWER taskload components. So the communications measures, for the most part, were related to each other and to ATC activity.

We also found that the prediction of complexity ratings using the available data was complicated. Unlike the results of [3], the POWER components alone did not statistically predict complexity ratings as well as the full model containing all variables. The only reduced model we found that predicted complexity ratings as well as the full model contained both taskload and communications measures and excluded very few of either. However, in this study, the statistical prediction was influenced by a larger number of observations, and so nearly every regression model comparison was significant and few reduced models were equivalent to the full model. Thus, we did not find that taskload measures alone were adequate to predict traffic complexity, but neither were the communication measures.

If this type of analysis is conducted in the future using even larger sample sizes, it may be necessary to approach the problem from a different perspective. For example, while the F test assessing the significance of the R^2 change in the model containing all five POWER components was significant, indicating that the models were not equivalent, the actual change in R^2 was .07, or only 7% of the variance in the data.

We would propose that, given a reasonable sample size, a change in R^2 of 10% or less might be considered an adequate indication of the similarity of regression models. If that "rule of thumb" were adopted, then a model containing only the POWER components might be considered an acceptable replacement for the full model, while another model containing only the communication measures might not be considered acceptable. In this case, we might conclude that the taskload measures alone might replace the need for using communications measures to predict subjective workload.

Additional concerns must eventually be addressed in this type of research. As mentioned earlier, data were analyzed for five periods within each traffic sample (except for periods excluded due to missing data.) That type of analysis introduced the possibility that data from the same traffic sample were not statistically independent. As more data become available, it will be possible to use entire traffic samples (rather than subsets) as observations, eliminating the possibility of lack of independence.

We had hoped that analyzing a different set of data would confirm or deny the results found in [3] which suggested that communications measures did not add anything to the prediction of subjective workload over and above the contribution of taskload measures. The results of this analysis suggest that taskload is a better predictor of complexity than communications, but do not provide a definitive answer. Additional SAR and communications data obtained from Ft. Worth Center for the DD study [6] will be available for analysis soon. Perhaps they will help clarify this matter.

References

[1] Morrow, D. & M. Rodvold, 1998, Communication issues in air traffic control. In M. W. Smolensky & E. S. Stein (Eds.) *Human Factors in Air Traffic Control*. San Diego, CA: Academic Press, pp. 421-456.

[2] Porterfield, D.H., 1997, *Evaluating controller communication time as a measure of workload,* 7, International Journal of Aviation Psychology, pp. 171-182.

[3] Manning, C. A, S. H. Mills, C. M. Fox, E. M. Pfleiderer, & H. J. Mogilka, 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.

[4] Mills, S. H., E. M. Pfleiderer, & C. A. Manning, 2002, *POWER: Objective Activity and Taskload Assessment in En Route Air Traffic Control*, (DOT/FAA/AM-02/2), Washington, DC, FAA Office of Aerospace Medicine.

[5] Rodgers, M. D., & D. A. Duke, 1993, SATORI: *Situation Assessment Through Re-creation of Incidents*, 35(4), The Journal of Air Traffic Control, pp. 10-14.

[6] Federal Aviation Administration, William J. Hughes Technical Center, NAS Advanced Concepts Branch, ACT-540, 2001, *Dynamic Density metric development and validation: Research Management Plan*, Version 2.0. Atlantic City, NJ: Author.

[7] Federal Aviation Administration, 1993, Multiple Virtual Storage (MVS), User's Manual, Data Analysis and Reduction Tool (DART), (NASP-9247-PO2), Washington, DC, Author.

[8] Stein, E.S., 1985, *Air traffic controller workload: An examination of workload probe,* (DOT/FAA/CT-TN84/24), Atlantic City, NJ, Federal Aviation Administration Technical Center.

[9] Prinzo, O. V., T. W. Britton, & A. M. Hendrix, 1995, *Development of a coding form for approach control/pilot voice communications*. (DOT/FAA/AM-95/15), Washington, DC: FAA Office of Aviation Medicine.