Using Human-Machine Interaction Frequency as a Proxy Measure of Subjective Air Traffic Complexity

Bruno Antulov-Fantulin, Tomislav Radišić
Department of Aeronautics, Faculty of Transport and Traffic Sciences, University of Zagreb
Zagreb, Croatia
bantulov@fpz.hr

Matthieu Rummens
French Air Force Academy
French Air Force
Salon, France

***Abstract*— Subjective air traffic complexity scores have been used previously as a useful measure of air traffic controller workload. There were, however, difficulties in implementing such measurements for real-time workload assessment due to the extent of modifications needed on an operational ATM system. A solution is proposed here which requires only the minimum of HMI data to determine the subjective air traffic complexity. For this paper, an experiment has been conducted with licensed air traffic controllers who assessed air traffic complexity in real-time human-in-the-loop simulations. Simultaneously, basic human-machine interactions were recorded. Analysis of the simulation data showed that the human-machine interactions can be used, with some limitations, to detect increases in air traffic complexity and situations where the controller's workload capacity is exhausted.**

Keywords—HMI, Air Traffic Complexity, Workload Capacity

I. Introduction

The limited efficiency of current air traffic systems will require a next-generation of air traffic systems that are able to help air traffic controllers in their job. Today, systems in air traffic control already have a lot of tools that are helping and making the job of an air traffic controller more efficient and reliable. It is known and it has been researched in other papers that airspace capacity is equal to the workload capacity of the air traffic controller currently working on the airspace [1]. Air traffic controller workload is likely to remain the greatest functional limitation on the capacity of the air traffic management system.

One of the key factors contributing to air traffic controller workload is air traffic complexity. Given predicted traffic increases, as well as corresponding developments in air traffic control procedures and technologies, it is increasingly necessary to understand the abilities of air traffic controllers and to identify the "safe" limits of workload [2]. In the literature on air traffic control complexity, surprisingly few definitions of "complexity" appear to have been given, presumably because the authors assume it is common knowledge. One of the authors defined complexity as a "…measure of the difficulty that a particular traffic situation will present to an air traffic controller…" [3] and went on to describe workload as "…a function of three elements, firstly, the geometrical nature of the air traffic; secondly, the operational procedures and practices used to handle the traffic and thirdly, the characteristics and behaviour of individual controllers (experience, orderliness etc.)….".

Measures of air traffic controller workload are typically based on subjective ratings made by controllers either while controlling air traffic or just afterwards [4]. It is clear that the relationship between air traffic control complexity and workload is an indirect one that is highly mediated by the influence of many individual characteristics, however, increase in complexity always means increased workload for the air traffic controller. A given level of traffic density and aircraft characteristics may create more or less complexity depending on the structure of the sector. Traffic density alone does not define air traffic control complexity, but it is one of the variables that influences complexity and so is a component of complexity. Its contribution to air traffic control complexity partially depends on the features of the sector. Sector and traffic complexity interact to produce air traffic complexity [5].

Although, measurable features of sectors and aircraft may be objective, the concept of air traffic control complexity is subjectively defined by the controller. It is developed from the controller's perception of and interaction with the sector and the air traffic within it [5], and therefore it can only be assessed by controllers using the subjective complexity assessment scores. Complexity is an acute problem in air traffic control and can ultimately limit the safety, capacity and efficiency of the system. The majority of research on air traffic control complexity has been concerned with examining the complexity imparted by the air traffic itself, and not on the overall complexity contributed by the human-machine interaction process [6].

There are several papers worth mention that addresses the same field of research that authors of this paper did. The human-machine interaction or as it is also referred in some other papers as human-computer interaction is very interesting in measuring air traffic control complexity. One of the papers addresses the testing of one of the methods to assess the complexity of air traffic control displays [7]. Also one of the studies worth mentioning is the development of TRACEr. The paper outlines a human error identification technique called TRACEr—technique for the retrospective and predictive analysis of cognitive errors in air traffic control [8]. TRACEr is a valuable aid to design, development and operations in United Kingdom air traffic control, but unfortunately it does not use human-machine interactions to predict errors. One research that uses human-machine interaction states that „...research in the field of human–computer interaction (HCI) has shown that early usability evaluation of human interfaces can reduce operator errors by optimizing functions for a specific population...“ [9]. The most related research paper to this one is by Christos George Tsonis (2006) who used the human-machine interaction with human-in-the-loop simulations [6].

Although very similar methods were used, authors of these paper research air traffic control complexity in a way not measured before. The main hypothesis of this paper is that human-machine interactions can be used to detect increase in air traffic complexity. And with that, a set of new future research can be made to further improve the air traffic systems safety. Authors thought that a new system could be created that records basic human-machine interactions and later on uses that data to detect increase in air traffic complexity and maybe detect if an individual air traffic controller is reaching his/her workload capacity. With that information a system could alert the shift supervisor if the air traffic controller is near his workload capacity and prevent any accidents that might have happened. For this paper, an experiment has been conducted with licensed air traffic controllers who assessed air traffic complexity in real-time human-in-the-loop simulations. Simultaneously, basic human-machine interactions were recorded. Analysis of the simulation data showed that the human-machine interactions can be used, with some limitations, to detect increases in air traffic complexity and situations where the controller's workload capacity is exhausted.

II. Methodology

For this experiment, real-time human-in-the-loop (HITL) simulations were chosen as a method for gathering data because simulations can be performed in a controlled environment which allows repeatable conditions for all participants. Simulations were performed with the ATC research simulator developed and validated at the Department of Aeronautics of the Faculty of Transport and Traffic Sciences, University of Zagreb. This study was part of a larger research project on the effects of trajectory-based operations on air traffic complexity. The scope of the study was narrowed down to only nominal area control operations (en-route airspace) to make it more manageable.

A. *Participants*

Ten licensed air traffic controllers were recruited from the national air navigation service provider (ANSP). All were, at the time, working daily at the area controller positions. Participants were, on average, relatively young (mean age, 31; age range, 27-34) but with multiple years of experience working their positions (mean experience: 5 years; range, 2-9). Of the ten participants eight were male and two female.

All participants were briefed on the study protocol in broadest terms but no mention was made of the variables which were to be measured. Since there were some small differences between the professional workstations participants used daily during work and the ATC research simulator used in the study, participants were given three one-hour simulator sessions to make them accustomed to the differences. During or after these simulator sessions, all participants strongly affirmed that they thought the research simulator was representative of the actual system and that they felt unhindered in performing their routine tasks.

B. *Airspace*

To make the research environment as similar as possible to the actual work environment, local airspace was used (Croatia Upper North sector). All participants had multiple years of experience working with this airspace. Aeronautical Information Publications (AIP) were used to gather up-to-date data on local airspace and airspace of neighbouring countries.

Geographically, the sector consists of airspace over northern Croatia and north-western Bosnia and Herzegovina (Fig. 1.). Vertically, the sector, as used in this research, starts at FL 285 and ends at FL 660 (though no flights were flying that high). In reality, due to traffic demand, the sector is often vertically divided into several sub-sectors depending on the traffic loads and in that case 'Upper' is used to describe the sector from FL 325 – FL 355. For this research the complete vertical expanse was used.

The transfer of traffic between neighbouring Area Control Centres (ACC) and Zagreb ACC is regulated by Letters of Agreement (LoA). For this research the relevant parts of LoAs were Flight Level Allocation and Special Procedures sections which state the conditions that have to be met for all flights crossing the boundary of the CTA (called Flight Level Allocation Scheme - FLAS). The purpose of FLASes is to ensure that flights will cross the CTA boundary at required flight levels that enable them to land at the desired airport or to be seamlessly joined with existing traffic. It also states what are the coordination points (COP) or transfer of control (TOC) points. The participants were required to adhere to these procedures during the simulation runs.

C. *Traffic*

To ensure representativeness of the simulations (and validity of the results in extension), traffic sample needed to be as similar as possible to the real traffic flying through the selected airspace. For this purpose a detailed analysis of the traffic flows and patterns was performed. Historic traffic data was obtained from EUROCONTROL.

Since varying traffic levels were needed to measure HMI frequency at different levels of air traffic complexity, a summer day with high traffic variability was selected as a reference day

Fig. 1. Croatian Upper North Airspace Sector (as used in this research)

(Friday, 30 August 2013). Traffic from off-peak hours was used to create scenarios with low traffic levels, peak hours were used to create high traffic level scenarios, and peak hour traffic augmented with additionally generated traffic was used to create scenarios simulating future traffic levels (as explained in Section 2.D).

As expected, out of the 661 flights that flew through the Zagreb CTA Upper North sector most were commercial medium (approximately 70%) and long-range (approximately 9%) jets. Others were mainly regional turboprops and business jets.

Routes used most frequently during the selected day were those connecting South-East and North-West of Europe. 90% of the flights were in the general east-west direction, with remaining 10% in the north-south. More than 50% of all flights followed one of the five most frequently used routes (Fig. 2).

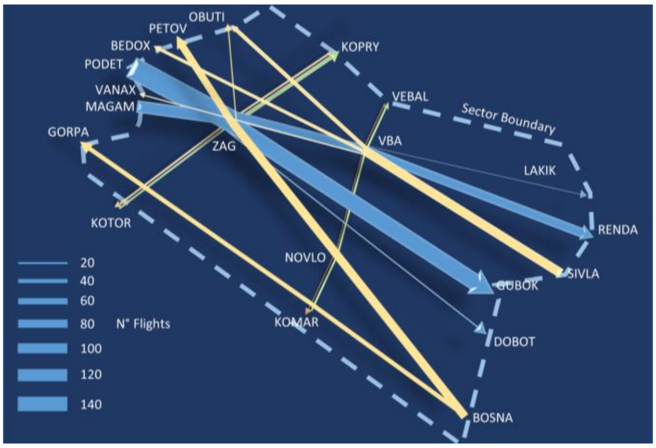

Fig. 2. Most Frequently Used Routes

D. *Scenarios*

Each participant did nine simulation runs (not counting the simulator preparation training). There were three categories of simulation scenarios according to traffic loads: *Low*, *High*, and *Future*. *Low* and *High* scenarios were developed on the basis of actual traffic data, while *Future* scenarios were based on the peak historic traffic loads which were taken as a starting point and then gradually increased beyond the expected controller's capacity levels. These scenarios had unrealistically high aircraft counts and, to increase air traffic complexity even further, the greater fraction of aircraft were climbing or descending compared to other scenarios. The goal was to increase complexity beyond the levels that could safely be reached with actual traffic and beyond what the controllers experienced during their careers.

To increase complexity above the levels of routine traffic, additional flights were generated in a semi-stochastic manner. Firstly, route was chosen randomly with probability of a given route being chosen equal to the frequency with which it was flown in reality. Secondly, aircraft type was randomly chosen from the actual aircraft distribution for that day. Thirdly, appropriate flight level for that route was chosen with regards to semi-circular system of cruising levels. Finally, time of entry into the sector for that flight was randomly generated until the flight was not in conflict with any other flight in the first few minutes since the entry into the sector (in reality those are solved through coordination between two sector controllers). This was checked by using fast-time simulations. Using this method it was ensured that the artificially generated flights had approximately the same distribution as real ones, without generating un-realistic traffic flow patterns.

Each scenario lasted between 50 and 55 minutes, however, data from the initial 5-10 minutes were discarded because scenarios started with no aircraft initially present in the airspace so there was no data to record. Starting with empty airspace also helped avoid the need to simulate the transfer of responsibility from one controller to another (like in shift changes) which would make the protocol unnecessarily more complicated.

Separation minima for these scenarios were 5 NM horizontally and 1000 ft vertically which was a familiar requirement for the participants because same minima was used during their everyday operations. However, in this experiment the simulation was stopped at the moment that the separation minima infringement occurred because that was, by definition, the maximum air traffic complexity for that controller.

E. *Data acquisition*

A wealth of information is generated by the simulator during the simulations. All data is stored in real-time to the hard drives at both the controller and pseudo-pilot stations. Stored data can be separated into three main categories:

- Aircraft state data. For each flight all variables pertaining to that aircraft's state are stored at one second interval. These include, but are not limited to: position, speed (TAS, CAS, Mach, GS), heading, track, mass, thrust setting, bank, pitch, drag, ESF, FMS variables, assigned level/heading/speed, etc.
- Human-machine interactions. All mouse events are stored (click, double-click, move, hover, scroll) and all keyboard inputs as well.
- Subjective complexity scores. Subjective complexity scores are collected during the simulation at two minute intervals, with complexity assessment panel sliding into controller's view at the right side of the radar screen, followed by aural notification (without stopping the simulation). They are time-stamped and stored on the hard drive.

To collect subjective complexity scores, modified Air Traffic Workload Input Technique was used (ATWIT) which was previously used to assess complexity by other authors [10][11]. In this study, score descriptors were modified to better reflect the goal of assessing complexity (Table 1).

Description of each subjective complexity level is mostly based on self-assessment of situational awareness which is additionally clarified using aircraft-aircraft or aircraft-airspace interactions. Before using this scale, controllers need to be briefed about the purpose of this technique and meaning behind the words *'complexity'*, *'interaction'*, and *'situational awareness'*.

TABLE 1. AIR TRAFFIC COMPLEXITY RATING SCALE

Complexity Level	Description
1	No complexity – no traffic
2	Very low complexity – very little traffic, no interactions
3	Low complexity – situation and interactions obvious at a glance
4	Somewhat low complexity – firm grasp of the situation, interactions are anticipated and prepared for
5	Somewhat high complexity – aware of the situation, interactions are handled in time
6	High complexity – having trouble staying aware of all interactions, occasionally surprised by unnoticed interactions and conflict alerts
7	Very high complexity – losing situational awareness, unable to track all interactions, responding reactively

III. RESULTS

Before beginning the data analysis it was necessary to prepare the data. First, frequency of each human-machine interaction was calculated for each minute of the simulation. Then, since the subjective complexity scores were entered ideally at two-minute intervals, they needed to be interpolated at one-minute intervals. This issue was further made important by the fact that controllers did not enter their scores at exactly the same moment that the prompt appeared. In worst cases some controllers were late by more than a minute, probably due to heavy workload. For interpolation, the nearest-neighbour method was used. Finally, data from the beginning of each simulation run were discarded because there were no aircraft in the airspace at that time.

Early on, it was found that higher-level HMIs, such as ‘entering assigned altitude into stripless flight progress monitoring system’ or ‘activating range and bearing tool’, could not be used for analysis with any significant result because they occurred very infrequently and sporadically. Therefore, decision was made to analyse only low-level HMIs, such as ‘click’, ‘drag-and-drop’, and ‘hover’. These events occurred with much higher frequency and they also included all of the higher-level interactions which had low frequency by themselves.

Two types of data analysis were performed. First, correlational study was performed on individual and averaged data to detect possible correlation between subjective complexity scores and HMIs. Second, an analysis of predictive power was performed to determine under which conditions the HMIs could be used to infer the probable level of subjective air traffic complexity.

A. *Correlational study*

In this part of the analysis correlation between subjective complexity scores on the one hand and three types of human-machine interactions on the other was tested. Each pair of variables contained data for the whole of the experiment; data from all participants were combined into single variables. There were 515, 487, and 401 data samples for *Low*, *High*, and *Future* scenarios respectively. Lower number of data samples in scenarios with *High* and *Future* traffic levels were due to the fact that the simulation runs were stopped at the moment when separation minima were infringed which happened more often at higher traffic volumes. Sample Pearson’s correlation coefficient (r) for each pair of the variable can be seen in Table 2.

TABLE 2. SAMPLE PEARSON’S CORRELATION COEFFICIENT FOR COMBINED DATA (COMPLEXITY SCORES VS. HMIS)

	Scenario Type		
	Low	***High***	***Future***
Click	0.2224	0.3705	0.5086
Hover	0.0365	0.3585	0.5057
Drag	0.0736	0.3729	0.4223

As can be seen in Table 2, correlation is very weak in scenarios with low traffic volumes whereas it gets stronger with increased traffic volumes. This effect could be attributed to the very low variance of the subjective complexity scores in *Low* scenarios where some participants even assigned the same score (1) to all traffic situations throughout the scenario. Another cause of the low correlation coefficients could be due to the large variance in the complexity scores assigned to the same situation by different controllers. Therefore, another attempt at analysis was performed with data separated per participant. However, in this case data from the different scenario types were combined into single variables. Sample Pearson’s correlation coefficients per participant can be seen in Table 3.

TABLE 3. SAMPLE PEARSON’S CORRELATION COEFFICIENTS FOR PER-PARTICIPANT DATA (COMPLEXITY SCORES VS. HMIS)

	Participants									
	P1’	***P2***	***P3***	***P4***	***P5***	***P6***	***P7***	***P8***	***P9***	***P10***
Click	0.40	0.60	0.41	0.44	0.52	0.48	0.52	0.41	0.66	0.55
Hover	0.48	0.56	0.43	0.62	0.51	0.57	0.51	0.63	0.76	0.66
Drag	0.46	0.68	0.36	0.45	0.55	0.56	0.75	0.58	0.45	0.66

In Table 3, it is visible that the correlation between subjective air traffic complexity and HMIs is much more consistent when it is considered separately for each participant. Also, correlation coefficients hover around 0.5 which can be considered acceptable for some purposes.

One more interesting effect was noticed when mean complexity scores from all participants were used. In this test, for each traffic situation, subjective complexity scores from all participants were used to calculate the mean complexity score for that traffic situation and then the correlation with HMIs was tested. Results can be seen in Table 4.

TABLE 4. COEFFICIENTS OF CORRELATION BETWEEN MEAN AIR TRAFFIC COMPLEXITY AND HMIS

	Scenario Type		
	Low	***High***	***Future***
Click	0.420	0.577	0.735
Hover	0.682	0.883	0.704
Drag	0.832	0.905	0.874

With mean complexity scores, the sample Pearson’s correlation coefficients show much stronger correlation between complexity and frequency of HMIs. This is especially true for *Drag* events which correlate extremely well with the complexity. The source of the increased correlation performance when using mean complexity data might be the well-known 'wisdom of the crowds' effect where mean of scores from independent assessors more accurately predicts some value than the individual scores [12][13]. Possible implications of these results are explored in later section (Discussion).

B. *Predictive power of HMIs*

Linear regression model was created with HMIs as predictors for subjective air traffic complexity. At first, combined data from all participants was used, however regression performance was relatively poor as can be seen in Table 5.

TABLE 5. RESULTS OF LINEAR REGRESSION FOR COMBINED DATA

Scenario Type	Dependent Variable	Predictors	R	R^2	R^2 - adjusted	Std. Error of the Estimate
Low	Complexity Scores	Click, Hover, Drag	0.228	0.052	0.047	0.748
High	Complexity Scores	Click, Hover, Drag	0.497	0.247	0.243	1.088
Future	Complexity Scores	Click, Hover, Drag	0.599	0.358	0.353	1.529

Coefficient of determination (R^2) is lower for scenarios with *Low* traffic volumes which can, again, be attributed to the low variance in subjective complexity scores for these scenarios (many controllers gave lowest score throughout the whole scenario). R^2 increases somewhat with the increase in traffic volume but so does the standard error of the estimate. Similarly to the correlational analysis, regression was attempted again with data separated per participant but combined for all scenario types. Results from these linear regression analyses can be seen in Table 6.

TABLE 6. RESULTS OF LINEAR REGRESSION FOR PER-PARTICIPANT DATA

Par.	Scenario Type	Dependent Variable	Predictors	R	R^2	R^2 - adjusted	Std. Error of the Estimate
1	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.575	0.330	0.315	0.7809
2	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.771	0.595	0.586	0.9314
3	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.587	0.345	0.330	0.8698
4	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.647	0.418	0.405	1.0441
5	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.741	0.549	0.539	0.9215
6	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.718	0.515	0.504	0.6217
7	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.784	0.614	0.605	0.9904
8	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.699	0.488	0.476	1.0385
9	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.800	0.640	0.633	1.0226
10	Low, High, and Future	Complexity Scores	Click, Hover, Drag	0.770	0.593	0.584	0.6272

With the exception of participants 1 and 3, coefficient of multiple correlation (*R*) is between 0.65 and 0.80, which is quite satisfactory when human factors and subjective assessment are involved. Further improvement could be achieved by selecting subset of data with no *Low* scenarios because those scenarios showed poor regression performance in the first analysis, however that was avoided because in real operations controllers do have periods of time with low traffic volumes.

Although the linear regression gave meaningful results, another method of detecting high air traffic complexity was tested. A threshold was set at subjective complexity score of '4'. Those traffic situations which had score below or equal to threshold were considered to be of low or medium complexity and thus not inherently difficult or unsafe, whereas the rest of the traffic situations (with complexity scores above the threshold) were considered very complex and therefore potentially unsafe.

Means of the three HMI indicators were calculated separately for those traffic situations below the threshold and those above it. Results are in Table 7.

TABLE 7. MEAN VALUES OF HMI INDICATORS FOR SITUATIONS BELOW AND ABOVE THE THRESHOLD

HMI Indicator		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Click	*>Thr.*	13.2	14.4	13.6	11.6	12.3	11.7	13.4	10.1	13.8	16.1
	<=Thr.	7.7	6.2	7.3	8.4	6.6	6.2	7.5	6.3	5.9	8.3
Hover	*>Thr.*	384	346	218	397	300	192	208	396	427	373
	<=Thr.	355	211	203	262	249	237	166	249	197	211
Drag	*>Thr.*	18.8	18.6	4.3	9.7	14.3	21.8	10.2	17.1	15.4	20.5
	<=Thr.	17.8	9.2	5.8	9.1	7.8	12.8	4.4	10.5	9.7	11.3

The first noticeable result is that the frequency of HMIs well increases when the complexity is above the threshold compared to less complex situations. It seems possible to choose a threshold of interactions frequency which can be calculated by the computer. If the frequency is above this threshold, the system could react by alerting shift supervisor in order to inform them that the particular controller is currently experiencing heavy workload.

The next step is to evaluate the value of this threshold. However there is another significant tendency in this table. It is noticeable that the ATCOs have different interactions frequencies. Some ATCOs interact much more than others, sometimes three times more faced to the same situation. The HMI interaction rate appears to be dependent on the ATCO. For the same situation, for the same air traffic they do not have the same need to interact with equipment. Due to this fact, the threshold has to be specific to each ATCO and evaluated with a simulation run such as one of scenarios which were used in this research. The threshold has to be evaluated in relative terms, regarding to the lowest number of interactions calculated from the less complex situations, in order to be more adaptive to every ATCO. Each ATCO could have his/her own interactions threshold saved in the system in order to assess whether they are experiencing high workload.

Calculation was performed in order to find the best threshold for this detection of high workload and also to optimize the false alarm rate and the detection probability. Here, the first one means that the ATCO performs a high value of HMI interactions but he/she ranks the complexity below the threshold. This rate has to be as low as possible. The second one means the rate of complex situations (complexity score above the threshold) associated with a number of HMI interactions which is above the threshold. This metric should be high.

Mean values of the below-threshold HMI frequency was selected as a baseline (100%). The threshold was then increased by 10% and false alarm rate and detection probability calculated. Initial results were poor due to high variance in the frequency of the interactions, however once data was smoothed by averaging past three minutes of the simulation, results significantly improved. The following table shows different values of these three figures (Table 8). The statistical rates were calculated by averaging the results from the ten controllers.

TABLE 8. DETECTION AND FALSE ALARM RATES AS A FUNCTION OF THRESHOLD VALUE

Threshold (%) of interaction frequency	Detection rate	False alarm rate
110	0.954	0.323
120	0.920	0.276
130	0.890	0.230
140	0.829	0.179
150	0.774	0.142
160	0.681	0.105
170	0.609	0.080

Setting a threshold depends on the purpose of the system but some trade-off between detection rate and false alarm rate will always be present.

IV. DISCUSSION

There are several lessons to be learned from this research. Generally speaking, it is possible to use human-machine interaction frequency as a proxy measure of subjective air traffic complexity. This conclusion, however, comes with several caveats. Firstly, low-level interaction events are more common and therefore more useful as a measure of HMI frequency. Secondly, humans differ very much in the way they use equipment. Some participants displayed much higher frequency of HMIs than others. Because of this, any attempt of analysis that uses combined data from a number of participants is bound to fail. Thirdly, controllers tend to bunch the complexity scores at the lower end of the scale which makes it difficult to correlate data from scenarios with low traffic volumes with any set of data.

Linear regression can be a useful tool to make a model for predicting air traffic complexity based on HMI frequency. This model, however, needs to be created for each controller individually which makes it somewhat impractical because every controller should do at least three hours of simulator runs while giving complexity scores in order to gather enough data to create a model. Also, such model should probably have to be updated once equipment, airspace or procedures change.

Somewhat simplified method for detecting high complexity by continuously analysing HMI frequency and setting a frequency threshold was presented as well. With this method, a controller workstation could automatically detect peaks in HMI frequency and inform the shift supervisor or store them for later analysis. It could also be used for forensic analysis in the aftermath of an incident or accident.

Perhaps the most unexpected result of the research was very large improvement in correlation between complexity scores and HMI frequency when mean complexity scores were used. Obviously, this type of group judgement can provide new insights into the air traffic complexity. It might be possible to use mean complexity scores from a number of controllers to create a universal model for calculation of baseline complexity score for any given traffic situation. This could then, in turn, be used instead of aircraft count as a measure of air traffic controller workload. Also, prior to the experiment, participants had the concept of complexity explained to them, however, they were not given any pre-scored traffic situations or guidance as to how to determine the complexity based on the features of the traffic situation. This was a conscious choice to avoid influencing controllers in any way but maybe some training with reference scores could be used to give controllers at least some sense of the range of complexity assessment scale that they will use in the experiment.

V. CONCLUSION

This paper presented methodology and results of the research on the correlation between subjective air traffic complexity and human-machine interaction frequency. Through real-time human-in-the-loop simulations ten controllers assessed air traffic complexity. Their scores were then compared with the HMIs gathered by the ATC simulator. Hypothesis that the HMIs can be used to infer air traffic complexity (and workload, by proxy) was confirmed. However, this method comes with several limitations which severely reduce its practical application.

One unexpected and interesting finding of this research was the fact that all HMI indicators correlated very well with mean of complexity scores of all participants (as opposed to the individual participant's scores) which is an interesting target for future research. Similar findings were already published in other fields but authors believe this is the first time such a phenomenon was detected in relation to air traffic complexity.

REFERENCES

[1] G. McAuley, J. M. Lenzi, and C. Lepadatu, "The factors affecting airspace capacity in Europe: A Framework Methodology based on Cross Sectional Time-Series Analysis using Simulated Controller Workload Data," ATM seminar, 2005, Baltimore, MD, USA.

[2] B. Hilburn (STASYS), "Cognitive Complexity In Air Traffic Control – A Literature Review," EEC Technical/Scientific Report No. 2004-007.

[3] C. Meckiff, R. Chone, and J.-P. Nicolaon, "The Tactical Load Smoother

for multi-sector planning," Proceedings of the 2nd FAA/Eurocontrol ATM R&D Seminar, Orlando, Florida., Dec. 1998.

[4] S. Athènes, P. Averty, S. Puechmorel, D. Delahaye, C. Collet, "ATC Complexity and Controller Workload: Trying to Bridge the Gap," HCI-02 Proceedings, 2002.

[5] R. H. Mogford, J. A. Guttman, S. L. Morrow, P. Kopardekar, "The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature," Atlantic City International Airport: Federal Aviation Administration, 1995.

[6] C. G. Tsonis, "An Analysis of Information Complexity in Air Traffic Control Human Machine Interaction," 2006, Massachusetts Institute of Technology.

[7] J. Xing, C.A. Manning, "Complexity and Automation Displays of Air Traffic Control: Literature Review and Analysis," 2005.

[8] S. T. Shorrocka, B. Kirwanb, "Development and application of a human error identification tool for air traffic control," Elsevier, Applied Ergonomies 33 (2002) 319-336.

[9] M. Clamann, B. A. Hamilton, M. Virginia, D. B. Kaber, "Applicability of Usability Evaluation Techniques to Aviation Systems," the International Journal of Aviation Psychology, 14(4), 395–420, 2004.

[10] P. Kopardekar, A. Schwartz, S. Magyarits, and J. Rhodes, "Airspace Complexity Measurement: An Air Traffic Control Simulation Analysis," in *US/Europe 7th Air Traffic Management Seminar*, Barcelona, 2007.

[11] A. Masalonis, M. Callaham, and C. Wanke, "Dynamic Density and Complexity Metrics for Real-Time Traffic Flow Management," in *Proceedings of the 5th USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary, 2003.

[12] F. Galton, "Vox populi," Nature 75:7, 1907.

[13] I. Lorge, D. Fox, J. Davitz, M. Brenner, "A survey of studies contrasting the quality of group performance and individual performance," 1920-1957. Psychol Bull 55:337–372, 1958.