tapraid5/z4j-aviat/z4j-aviat/z4j00404/z4j1679d04g | gockleyj | S=4 | 1/29/04 | 6:00 | Art: ASEM1052

RESEARCH ARTICLE

Mental Workload in Air Traffic Control: An Index Constructed from Field Tests

Philippe Averty, Christian Collet, André Dittmar, Sylvie Athénes, and Evelyne Vernet-Maury

Averty P, Collet C, Dittmar A, Athenes S, Vernet-Maury E. *Mental workload in air traffic control: an index constructed from field tests.* **Aviat Space Environ Med 2004; 75:000–00.**

Background: Mental workload assessment is a recurrent issue in air traffic control (ATC). Studies of ATC have used either objective aspects, i.e., numbers and distribution of aircraft, or subjective factors, such as self-imposed performance and stress levels, with mixed results. This is partly due to the difficulty in bringing together comparable data pertaining to both air traffic, with its ever-changing distribution, and judgement or quickly fluctuating psychophysiological variables. ***Methods:*** We propose a method of mental load estimation devised to take into account both objective traffic variables and the additional load imposed by subjective effects, including the seriousness of conflicts and the time-pressure for their resolution. First, we developed a traffic load index (TLI) to identify time boundaries during which additional load may occur. Then we quantified the additional load according to the air traffic situation. ***Results:*** TLI was developed from analysis of 25 h of recordings of radar control sessions involving 25 professional air traffic controllers at a major airport. Results were then compared with a simple objective index (number of aircraft) and subjective workload ratings (NASA-TLX test). The whole population (intersubjects analysis) showed a better correlation between the TLI and the self-rated workload than for the number of aircraft alone. Among the controllers who rated more than one level of workload through the TLX-test, 77.8% showed better correlation between TLI and TLX than between N and TLX (intrasubjects analysis). ***Conclusion:*** Workload estimation should integrate both objective task variables and subjective evaluations associated with them.
Keywords: workload measurement, air traffic control.

MENTAL WORKLOAD is central to studies of air traffic control (ATC), whether the purpose is to elaborate a cognitive model of the task, create high performance tools, or evaluate traffic flows or sectors. There are three factors thought to have a strong influence on workload: a time-based factor, a task intensity-based factor (i.e., difficulty or complexity that make demands on attentional resources), and an operator's psychophysiological functional state (8). Workload is thus a multifaceted structure that cannot be studied directly, but may be inferred from different, quantifiable variables. A first aspect (stress) is related to measurable constraints pertaining to the task and its environment. Stress and mental load are two related concepts that originate from different theoretical frameworks (6). The second aspect (strain) is of a more subjective nature and represents the "cost" to the operator, i.e., the effects of previous constraints undergone. There is general agreement that strain is the actual workload that reflects the overall demand on resources.

Previous studies on ATC have generated a realistic workload index using the length or content of radio exchanges and judgements by experienced observers (13). The number of aircraft was the commonest index to be used. By contrast, due to safety constraints, subjective aspects of workload (strain) have never been assessed in real time during actual work sessions or even over short inervals. Studies have often been conducted during simulated scenarios where traffic complexity was considered a dependent variable and several factors could be manipulated, i.e., arrival-to-departure flight ratios, type of aircraft (2), and information displayed on the radar screen and on flight strips (11).

The number of aircraft (N) being simultaneously managed by a controller provides an easily recorded index. However, the observation of a controller managing a number of aircraft in a given ATC sector shows that N is not a perfect index (10), especially as it is heavily biased by the traffic situation, i.e., the way aircraft are spread over space and time. The main hypothesis of this paper is that N would provide a better basis for mental load measurement if the index integrated features of the interventions to be carried out by the controller. The relationships between workload and emotional states generated by the situation should be taken into account in assessing mental load in ATC. It will then require a description of the procedures leading to the computation of such an index.

METHODS

Theoretical Backgrounds

The activity of a radar controller involves the monitoring of traffic, diagnosis of conflicts between aircraft, then determining and implementing action to prevent these. A conflict may occur as soon as two aircraft are

From the Claude Bernard University, UFR STAPS, CRIS, Villeurbanne Cedex, France.

This manuscript was received for review in February 2002. It was revised in October 2002, and June, October, and December 2003. It was accepted for publication in December 2003.

Address reprint requests to: Christian Collet, Claude Bernard University, 27, 19Boulevard du 11 novembre 1918, F-69 622 Vileurbanne Cedex, France; christian.collet@univ-lyon1.fr.

expected to be separated by a distance shorter than minimal safety intervals. Devising action that eliminates a conflict does not represent the most demanding part of the job. Many acceptable solutions to any ATC problem may be found through action on a small number of variables (altitude, heading, airspeed, rate of climb/descent). Furthermore, repetitive and practical experience results in a highly automated set of solutions, requiring few cognitive resources. The key process of cognitive activity concerns the management of attentional resources to assess future conflicts between aircraft. ATC requires the processing of dynamically changing visual information which is largely automated in professional ATC but remains constantly activated through time. Thus, two factors are thought to influence workload: decisionmaking and using evolutionary data to forecast and diagnose conflicts. Data must be ascertained as the situation evolves, thus requiring attentional resources. This defines the seriousness of a conflict. Interaction between uncertainty and the real-time aspect of the task, therefore, leads to consideration of time pressure as the time interval available for solving the conflict. This will represent the urgency variable. Finally, the basic process of conflict diagnosis includes vagueness about future flight path data: controlling aircraft can be defined as managing uncertainty on the necessity to act, rather than applying specific rules. Thus, temporal allocation of resources must match the traffic situation, estimating the evolution of distance between aircraft and coping with unexpected events. In heavy traffic contexts, a significant disruption may induce a massive load increase when related to a loss of (subjective) control/command of the situation and/or of (objective) control of it. However, traffic patterns and conflicts do not impinge directly on mental workload.

Air traffic controllers can modulate their own workload through cognitive strategies by learning to adjust it to an optimum level. Flight variables which are modified to prevent or solve a conflict may be combined with the moment of implementing this modification. Such self-regulation has been evidenced at high traffic levels but may also occur in a majority of actions. The interval between conflict diagnosis and the time when a decision is made is called maturing time (MT) and is delineated by time boundaries. MT follows the same course as workload: action undertaken at the outset of diagnosis holds workload at a low level, whereas extending MT increases attentional demands, thus increasing workload. However, increasing MT also decreases uncertainty: the later preventive action is undertaken, the more efficient it becomes. Better accuracy is then obtained for flightpath modification and minimum constraint is exerted on aircraft. Decisions have to be made by anticipating future positions and flight variables; the earlier they are made, the more they finally appear unnecessary (i.e., when separation is just above the minima). This leads to a trade-off between uncertainty and time pressure: elapsing time decreases vagueness but increases strain. Balance must be maintained according to objective and subjective contexts, i.e., traffic load and the current operator workload level, respectively. This uncertainty applies to any ATC issues: "Is it advisable to intervene on this flight path ?" In the affirmative, what would the right action be to guarantee safety with minimum flight path alteration and when would the right moment to take it be? Uncertainty may concern many aspects of flight data. The question on necessary and adequate action appears to be a major source of mental workload. Handling this issue is an important part of controllers' expertise. Thus, mental load in ATC should not be limited to N, even though the amount of information to be processed is closely linked to this variable. Enhancing a load index should, therefore, integrate time pressure and uncertainty, as these are closely connected to MT (1).

Workload and Emotion

ATC requirements confer an important role on emotional factors in decisionmaking, especially during high load situations. The effects of both cognitive and emotional aspects of workload must be considered simultaneously. Emotion is a part of subjective aspects and results mainly from possible consequences of errors in decisionmaking, but also depends on time constraints. The traffic situation and unexpected events are objective data that can induce such a threat. They temporarily focus the controller's attention, mobilizing considerable cognitive resources, possibly to such an extent that this could lead to excessive mental load. This must be distinguished from pathological stress manifestations as it contributes to decisionmaking and generally helps the controller to stay within his own limits to control the situation. The relevant situational features are mostly linked to conflicts that could occur, to MT associated with these, and to the contextual traffic load in which they have to be managed. Consequently, the number of aircraft would be a reliable basis for a mental load index if it took into consideration the additional load resulting from the threat of excessive arousal. The content of this additional load must be described to compute a traffic load index (TLI). The following sections will detail how the ATC data were processed and how the TLI was developed.

Method of TLI Computation

Enhancing the usual estimation of workload derived from aircraft positions requires integration of subjective data largely related to emotional correlates. The main difficulty is to quantify subjective processes attached to emotional events and thus to transform them into numerical data to be added to N. Accounting for additional load, TLI is expected to be more reliable in assessing ATC workload than the N variable alone. The experiment was carried out at Saint Exupéry International Airport (Lyon, France). Prior approval from the General Department of Civil Aviation (DGAC) was obtained. Subjects were 25 fully qualified ("radar approach" position) air traffic controllers from Saint Exupéry TRACON. Participants gave their informed consent after having been fully advised of the nature of the experiment. They were observed in field situations during sessions lasting about 1 h when working in the usual manner. Due to a greater local male population, only one female took part in the experiment. Heavy traffic periods were chosen to investigate medium and high levels of workload (between 6 and 9 pm). Over- and underload will not be discussed here. Such operator functional states can be identified, but it is assumed that this would require a specific study. Experimental material (e.g., strips, radar and audio communication recordings between pilots and controllers) was kept. Just after the work session, subjects were asked to take the NASA Task Load indeX test (TLX) (7). Each of them rated load through the five TLX scales (mental and temporal demands, performance, effort, and frustration). Subjects were briefed on the procedure and the scales they had to rate. None had taken the test before. They were told that it was possible to differentiate several periods, thus splitting the work session into distinct levels of workload.

ATC Data Processing

About 25 h of traffic (pilot/controller radio communications) were recorded, transcribed, and then verified. Empirical values propounded hereunder are closely linked to the ATC equipment (e.g., radar or no radar, number and characteristics of information displays, systems to help with decisions, etc.). Due to differences in maneuvering potential and traffic density, these values are also dependent on airspace categories (e.g., TRACON vs. Enroute). In other respects, problems resulting from Lyon traffic rarely need to be apprehended through clusters. Reasoning on the basis of pairs of aircraft appeared to restore features of almost all observable situations.

Operations on Aircraft: Monitoring/Vectoring/Conflicts

First, aircraft were categorized into two classes depending on whether they were simply monitored or involved in a control problem. This was carried out by an experimenter, who was also an air traffic controller instructor with ATC at Lyon St. Exupéry airport. Aircraft were considered as being simply monitored when they never needed any intervention from the controller (forecast separation with regard to differentials was always estimated as adequate). With regard to control problems, vectoring and conflict were differentiated. Vectoring was related to aircraft converging on the same airport and needing instructions on vectoring to provide the correct (forecast) spacing on the ILS axis. Conflict occurred as soon as two aircraft were expected to be separated by a distance lower than minimal safety intervals and thus required path alteration (noted from radio communications). When several aircraft are present in the same sector, the load will not be equivalent if each flight path remains independent from any other compared with the possibility of the flight paths interfering with one another. However, aircraft can be considered in conflict even if no action is taken by the controller. The relationship between two aircraft can potentially be as risky as any other, even without any action. The decision not to act was not established at the beginning of this situation, and lack of action must not necessarily be identified only with simply monitored aircraft. Thus, the experimenter had to identify such situations during the post-experimental analysis and take the affected pairs of aircraft into account along with those in conflict which required actual action.

Defining Time Boundaries

The time interval during which a control problem could occur has to be determined. Time boundaries define the period during which additional load has to be considered. Extra specific moments within this period could be characterized according to the demand on resources and correspond to the few benchmarks which stake out decision processes. Time pressure and uncertainty during MT were quantified with regard to each control issue, i.e., transformed into numerical data to be added to N. **Fig. 1** shows how MT and corresponding F1 time boundaries vary following conflict diagnosis as they integrate the controller's action time. Setting the length of this period is the basic mechanism through which controllers partly regulate their workload. The time devoted to resolution implementation makes significant demands on the controller's attention and on his emotional load.

Identifying problems as conflicts or vectoring did not cause the experimenter any major difficulty. However, as far as time boundaries were concerned, evaluation could differ from one controller to another, showing individual responsiveness to a similar, continuous stimulus (visual perception of a radar picture) that evolved with time. But while time boundaries actually varied to a certain extent, an estimated order could be defined that made sense for each controller viewing the radar recording (1). With reference to MT, this additional load is not constant and evolves as long as the control problem lasts, which depends on the instructions given by the controller, or on any change in flight variables (without any intervention by the controller). Conversely, when an instruction has been given to avoid a conflict or as soon as the two aircraft involved are far enough from each other, they are then considered simply monitored and there is no longer uncertainty or time pressure. Conflicts and vectoring are two different classes of ATC issues and present different time boundaries (H_n), as shown in **Table I**. T1

*Definition of Traffic Load Index (*tli*)*

Fig. 2 shows that uncertainty and time pressure must, F2 therefore, be taken into account in additional load computation. Additional load applies to an aircraft relationship which leads the controller to actively vector traffic or to otherwise prevent an impending conflict, which focuses cognitive resources, thereby increasing mental load. The final index will thus result from the computation of all events occurring during the time lapse defined between time boundaries.

This index is defined as the overall load experienced by the controller while working. It is related to each aircraft, each generating an elementary traffic load index (*tli*). Conventionally, the *tli* corresponding to each aircraft being monitored is equal to 1. Thus, aircraft

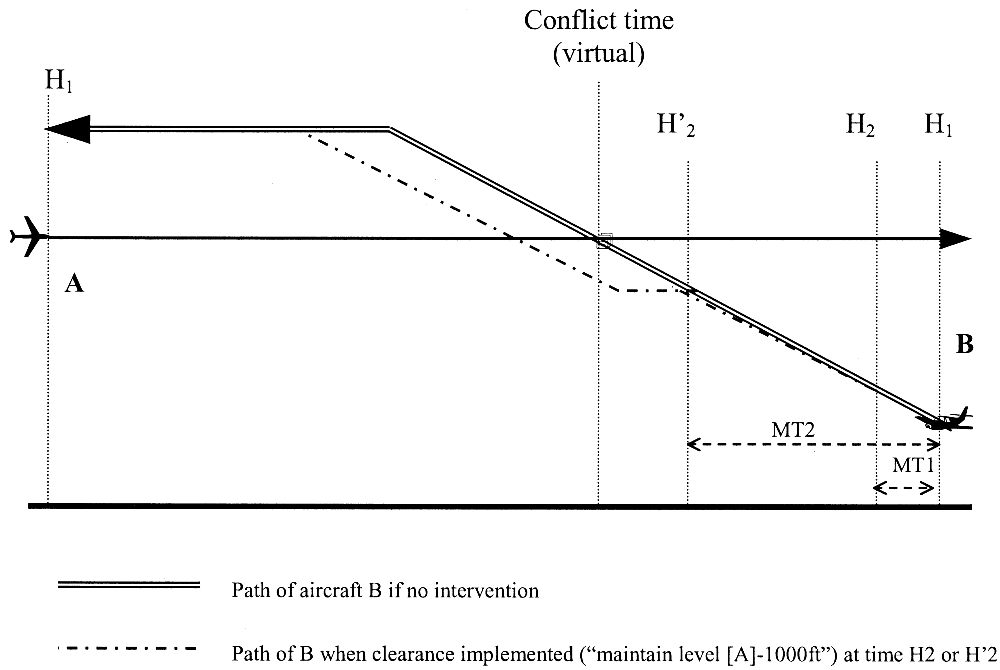

Fig. 1. Example of conflict diagnosis and maturing time. Aircraft B is climbing and runs the risk of insufficient separation from cruising aircraft A. H1 = time when conflict is identified, H2 = time when implementation is made, MT1 = maturing time related to resolution implemented at time H2 (short maturing time), H'2 = time when a later decision may be made, and MT2 = maturing time related to resolution implemented at time H'2 (long maturing time).

giving rise to a control problem will be given a *tli* greater than 1 (taking into account the additional load) within the time interval defined between boundaries as
F3 shown by **Fig. 3**. In most cases, resolution implementation takes place before H3. An increase in additional load (between H1 and H2) accounts for extra attention resource allocation; this is cancelled (H4) when separation becomes sufficient.

Seriousness

Differentiation of conflicts requires assessment of the seriousness of each conflict. Seriousness is defined by the distance between two aircraft and the intersection of flight paths. The controller evaluates the need to modify paths depending on whether the future distance between two aircraft is expected to be closer to or farther from minimum safety levels. This is a component of MT and contributes to the assessment of additional load through *tli.* The closer aircraft seem to be to each other around the intersection, the greater seriousness will be. Seriousness was estimated from flight data available before any avoidance implementation.

Conflict differentiation depended on whether the expected spacing could be estimated as closer to or farther from standard distances. Conflicts were classified into four categories, from minimum (class A: expected spacing is inadequate to ensure total safety) to maximum (class D: maximal risk of conflict), class B and C being intermediate levels. Quantification of additional load was then carried out with individually monitored aircraft seriousness as the unit: a one-dimensioned variable is thus proposed. Maximum seriousness (tli_{max}) related to control issues is distributed among the four different classes: class A, $tli_{max} = 2$; class B $tli_{max} = 2.5$; class C $tli_{max} = 3$; and class D $tli_{max} = 3.5$. These values, ranging from 2 to 3.5, were empirically determined by a single subject matter expert (SME), and must, therefore, be considered as testing values. In order to make numerical values of conflict seriousness as reliable as possible, a complete review of the recordings was conducted and then verified, the mean being maintained when two weightings differed. Seriousness, therefore, accounts for the extent of conflicts, i.e., the expected distance of a future separation from the current data available in real time. Such quantification is thought to represent the attentional demand for each aircraft.

The reference value of seriousness is the basic value of *tli* for an ordinary, monitored aircraft (i.e., not in-

TABLE I. DESCRIPTION OF TIME BOUNDARIES.

	Meaning of H_n for:	
	conflict	vectoring
H_0	Time of first radio contact	Time of first radio contact
H_1	Time when prevention action is conceived	$= H_5 - 3$ min
H_2	Time when a resolution is undertaken	–
H_3	Time before which avoidance must be undertaken	–
H_4	End of conflict (preventive action completed)	$= H_5 - 30$ s
H_5	End of radio contact	End of radio contact/ Interception of ILS

The two classes of ATC issues (vectoring and conflict) are related to different boundaries. Every H_n can be identified directly on recordings, except H_1 and H_3 that had to be estimated (one by one for each conflict). H_1 and H_3 can be assessed even if no action was taken by the controller.

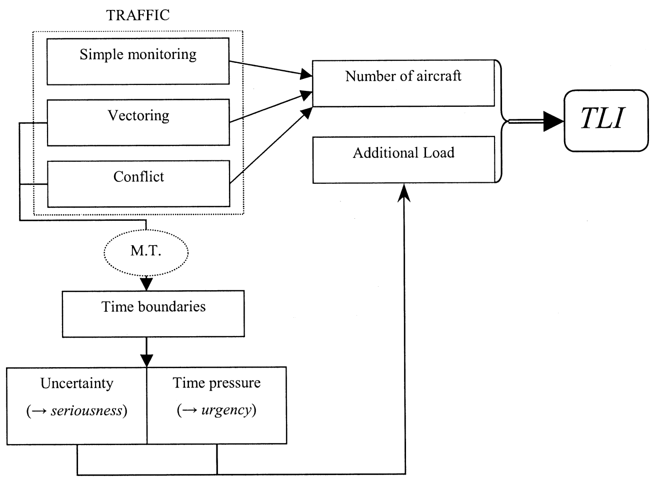

Fig. 2. Schematic diagram for the Traffic Load Index (TLI).

volved in any control issue), for which *tli* = constant = 1, from the time of first radio contact (H_0) until transfer (H_5) to another controller. According to this method, the load attached to two aircraft involved in a class D conflict could be equivalent to a maximum of seven aircraft being simply monitored simultaneously during the time interval H_3–H'_2 (Fig. 3).

Urgency

According to MT, additional load increases through time until reaching a maximum, tli_{max}. Such an increase accounts for time pressure and was assessed through the urgency variable. Here again, conflicts and vectoring need to be defined:

Several types of conflict are defined through seriousness, but present the same evolution pattern and are still related to time boundaries, as shown in Fig. 3. Generally, action is not appropriate as soon as conflicting aircraft come into radio contact, even if a future conflict appears to be possible. They are too far from each other and their future positions are still too doubtful for implementation of optimal avoidance measures or even for assertion that such are needed. Nevertheless, the period during which action has to be taken is roughly discernable (the time interval between boundaries H_1 and H_3). Thus, *tli* will be considered to increase between these two points, exerting an urgency effect. A time interval (H_3–H_4) is then necessary to ensure that avoidance action has been brought to a successful conclusion. Finally, *tli* = 1 again during the interval H_4–H_5. This is the generic frame for computing additional load due to a conflict (Fig. 3). Action is often undertaken at H_2, with $H_2 < H_3$. Thus, additional load will reach its maximum value at H_2, and will then decrease immediately down to H_4, when the problem has been solved. If preventive action is carried out later, at H'_2 (with $H'_2 > H_3$), *tli* will conventionally stay at the value of tli_{max} during the H'_2–H_3 interval.

Urgency in radar regulating is computed in a different way. When two (or more) aircraft are involved, detection may occur from the first radio contact. Thus, attentional demands increase slowly for nearly 3 min (12 to 15 nm) before ILS interception. At this time (H_1), aircraft are close enough to each other to undergo suitable action and more precise vectoring. The average length of this phase of maximum attentional demands (ending at H_4) was empirically set at 2.5 min for any vectoring. At H_4, the problem is considered to be solved (Fig. 3). Then, *tli* decreases down to H_5, when the aircraft intercepts the ILS and is transferred to the Tower. The urgency variable has no value of its own, but is part of the distribution of seriousness over time between H_1 and H_4. On account of the availability of a mathematical model, urgency variations will be encountered by linear functions as proposed in the above description.

Computing the Overall Traffic Load Index (TLI)

Overall traffic load index (TLI) is obtained by adding the *tli* computed for each aircraft at a given time (**Table II**). With reference to radar screen updating (from 4 to T2
8 s), *tli* was computed every 10 s. It is hypothesized that

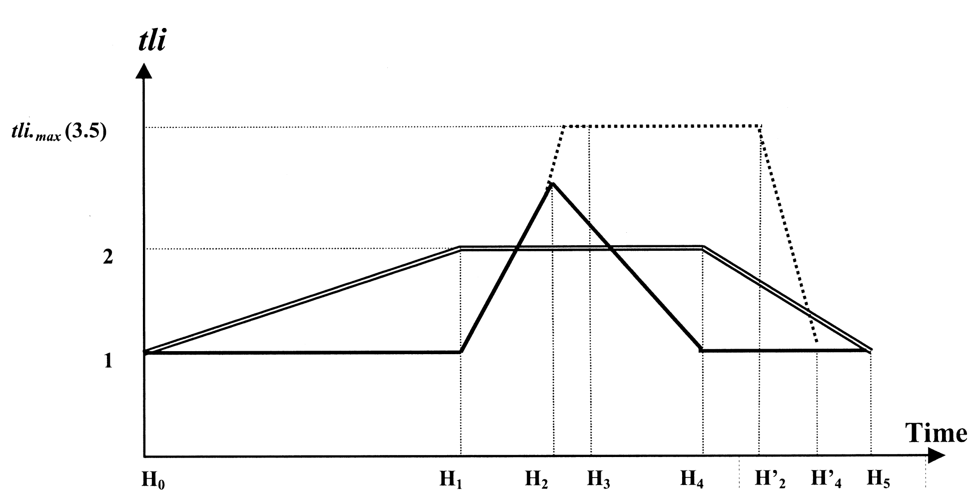

Fig. 3. Examples of fluctuation in Traffic Load Index (TLI) for vectoring (double line) and conflict (single line). H_0 = time of first contact; H_1 =
AQ: 1 conflict diagnosis time (conflict) or conventionally, H4–3 mn (vectoring); H_2 = resolution implementation time; H_3 = latest possibility of action
AQ: 2 implementation time; H_4 = end of conflict (conflict) or H_5–30 s (vectoring); H'_2= latest possibility of action implementation time (latest avoiding action); H'_4= end of conflict related to action at time H'_2; and H_5 = end of contact (conflict) or ILS interception time (vectoring). The dotted line shows TLI evolution for a late conflict avoidance.

TABLE II. EXAMPLE OF COMPUTING OVERALL TRAFFIC LOAD INDEX (TLI) THROUGH TIME.

TIME	C1	C2	C3	R1	R2	R3	N	T.L.I.
18:24:10	1.2			2.0		0.7	7	10.9
18:24:20	1.3			2.0	0.1	0.8	6	10.2
18:24:30	1.5	0.6	0.1	2.0	0.3	0.8	6	11.3
18:24:40	1.7	1.3	0.3	2.0	0.5	0.9	6	12.7
18:24:50	1.8	1.9	0.4	2.0	0.7	0.9	5	12.7

TLI is the sum of the number of aircraft (N) and the additional load computed from the relationship of each pair of aircraft. Additional load may originate from radar vectoring (Ri), conflict (Ci) or both. In this example, and according to this method of workload computing, TLI is higher in the 18:24:50 situation than in the 18:24:10 situation, whereas five and seven aircraft, respectively, have to be monitored. In this particular case, the TLI, therefore, shows a higher workload with five aircraft than with seven.

the amount of *tli* values gives a close estimation of a controller's strain: TLI = 3 tli_n (n = number of aircraft handled at a given time).

Statistical Analysis

Descriptive statistics are provided first. To test the reliability of TLI, the self-perceived workload (TLX), i.e., the Overall Weighted Workload (OWW), will be taken as a reference. By computing the correlation between TLI and TLX on the one hand, and TLX and N on the other, it is hypothesized that TLI would represent a better index than N if it is demonstrated to be better correlated to TLX than N. Thus, TLX was taken as a reference, e.g., a dependent variable (which may be predicted by N and TLI), whereas TLI and N were considered factors which may influence TLX, e.g., independent variables. The correlation between TLI and N was also computed even if TLI was constructed from N data. Intersubject and intrasubject analysis was carried out.

For intersubject analysis, since TLI and TLX were not computed from the same time scale, three different correlation tests were carried out. When the TLI time scale was considered, TLX values are repeated for the whole population to match the TLI values every 10 s. Thus, data were computed from original values. When the TLX time scale was considered, TLI and N values were averaged for the whole population to match each period where TLX remained constant. A correlation test, taking into account maximal individual values of TLI and TLX, was also carried out. For intrasubject analysis, a correlation between workload indices was carried out for each subject on the TLI time scale. The data from each individual were computed from original values.

To evaluate the weight of TLI in assessing workload, a comparison between two models is provided. The first is aimed at representing the perceived workload as a function of N: TLX = f(N). The second is designed to test the reliability of TLI, e.g., to represent the perceived workload as a function of N and TLI: TLX = f(N) + f(TLI). If the weight of TLI is real, the difference between the two models, when compared, was expected to be statistically significant.

RESULTS

As the experiment took place in the field, the total amount of traffic processed by each controller varied. Thus the distribution of aircraft in time and space throughout the session was never identical; even with significant peaks of traffic, sessions had moderate mean workloads (**Table III**). By analyzing all radar record- T3
ings, time boundaries were identified for each of the 295 control issues. A total of 130 conflicts (from 1 to 9 for each subject) and 165 cases of vectoring (from 1 to 11 for each subject) were identified by the experimenter during the 25 h of recordings. Conflict seriousness related to each (i.e., tli_{max}), was distributed over the four classes as follows: class A: 12.3% (16 conflicts out of 130), class B: 28.5% (37 conflicts), class C: 16.9% (22 conflicts), and class D: 42.3% (55 conflicts).

Conflicts were not spread equally according to conflict seriousness (**Fig. 4**) and were managed in four main F4
ways: 1) By providing navigation or heading alterations, so as to separate aircraft on the horizontal plane: 46 resolutions out of 130 (36.51%); 2) By requesting altitude or flight level to be held, when horizontal separation was estimated as adequate: 31 resolutions (24.60%); 3) By requesting pilots to increase their rate of climb or descent, to create vertical separation at the intersection point of paths: 14 resolutions (11.11%); and 4) By not implementing any action: 35 resolutions (27.78%). As shown by Fig. 4, the number of actions on both altitude and direction (heading) increases with the initial conflict seriousness. Action on rate of climb or descent (expedite), and especially the choice of monitoring the flight path evolution only (no action), are the opposite ratio of seriousness. The four remaining conflicts were settled through traffic information and visual maneuvers.

Subjective rating of workload using the TLX-test resulted in the OWW (Table III, columns 5 and 6). The number of different workload level periods varied from one to five, each of these being rated independently; seven subjects who experienced the lowest traffic load sequences rated the whole sequence in an overall manner, with the exception of subject 10 (Table III, column 7).

Intersubjects Analysis

TLI time scale: TLI was better correlated with the self-rated workload through TLX (multiple r = 0.57, $p < 0.001$, Pearson correlation test) than with the mere N variable (multiple r = 0.47, $p < 0.001$). TLI remained highly correlated with N (multiple r = 0.88, $p < 0.001$).

TLX time scale: TLI and N values were averaged according to each series of self-rated workload. According to this method, the Pearson pairwise correlation test

TABLE III. WORKLOAD ASSESSMENT REGARDING N, TLI, AND TLX.

	1	2	3	4	5	6	7	8	9	10
Subject	Mean N (SD)	*N* max	Mean *TLI (SD)*	*TLI* max	Mean OWW	OWW max	Number of workload period	Correlation TLI/TLX	Correlation TLI/N	Correlation TLX/N
1	3.2 (0.9)	6	4.0 (1.4)	8.0	26.5	26.5	1	-	0.67	-
2	3.8 (1.6)	7	6.3 (4.0)	17.5	30.6	53.0	4	0.52	0.78	0.52
3	4.7 (1.0)	6	6.5 (2.1)	12.3	48.3	54.5	3	0.10	0.61	−0.21
4	3.2 (1.1)	6	4.5 (2.1)	11.0	48.2	53.0	2	0.11	0.76	0.10
5	2.6 (1.1)	5	3.7 (2.2)	9.6	33.0	33.0	1	-	0.83	-
6	3.6 (2.0)	9	5.8 (4.4)	15.0	46.8	72.0	2	0.68	0.93	0.63
7	3.2 (1.5)	6	4.4 (2.6)	10.3	33.5	33.5	1	-	0.81	-
8	3.0 (1.0)	6	4.6 (2.5)	12.6	36.9	68.5	2	0.64	0.52	0.31
9	3.6 (1.7)	8	4.8 (2.7)	11.1	34.0	34.0	1	-	0.91	-
10	5.1 (2.2)	11	6.8 (3.9)	17.3	36.0	36.0	1	-	0.95	-
11	2.8 (1.3)	7	4.0 (2.7)	14.2	36.2	45.0	2	0.65	0.87	0.60
12	4.1 (1.3)	7	6.9 (3.3)	14.9	55.7	58.0	2	0.60	0.78	0.53
13	4.7 (2.3)	10	6.5 (4.3)	19.4	34.6	80.0	4	0.60	0.89	0.55
14	2.9 (1.8)	7	4.0 (2.9)	12.3	18.1	32.0	2	0.59	0.83	0.50
15	4.0 (2.6)	10	6.6 (5.5)	18.4	42.4	71.0	2	0.90	0.96	0.81
16	4.7 (2.1)	10	6.0 (3.5)	16.3	16.4	49.0	2	0.83	0.90	0.71
17	3.4 (2.7)	11	4.3 (3.8)	14.2	42.9	77.0	3	0.58	0.95	0.70
18	4.4 (2.3)	9	7.2 (5.7)	20.1	27.2	62.0	2	0.87	0.93	0.75
19	3.2 (1.7)	8	3.8 (2.1)	9.6	35.5	35.5	1	-	0.92	-
20	3.9 (2.2)	9	6.0 (5.2)	23.1	41.0	67.5	2	0.87	0.93	0.82
21	3.7 (1.9)	7	6.0 (4.7)	17.0	40.2	60.0	2	0.78	0.93	0.79
22	3.9 (1.8)	8	5.9 (3.6)	15.1	27.0	42.0	2	0.72	0.88	0.68
23	3.0 (1.3)	6	3.2 (1.6)	8.0	20.0	20.0	1	-	0.94	-
24	3.6 (2.8)	9	6.3 (5.7)	17.3	34.8	79.0	2	0.87	0.97	0.89
25	3.8 (1.8)	8	5.7 (4.2)	16.6	49.2	67.0	3	0.82	0.92	0.73

Mean N and TLI (col 1 and 3) correspond to the traffic processed on average by each subject (standard deviation in brackets). Mean OWW (col. 5) results from the different periods (mean of associated workloads) identified by subjects using the NASA-TLX. "Max" values (columns 2, 4, 6) indicate the maximum value of load that each subject experienced during the sequence, respectively through N, TLI or OWW. Column 7 shows the number of periods of load which were identified by subjects through the NASA-TLX test during their work session. Individual values of TLI, N, and TLX, recorded every 10 s, were used to compute correlations (columns 8, 9, and 10).

showed that, again, TLX was better correlated with TLI (multiple $r = 0.74$, $p < 0.001$) than TLX with N (multiple $r = 0.71$, $p < 0.001$). Here again, TLI remained highly correlated with N (multiple $r = 0.93$, $p < 0.001$).

Maximal values of TLI and TLX: Previous results were confirmed when maximal values were taken into account. Again, TLI was better correlated with TLX (multiple $r = 0.70$, $p < 0.001$) than N with TLX (multiple $r = 0.50$, $p = 0.03$). TLI and N remained highly correlated (multiple $r = 0.66$, $p = 0.001$), although the highest correlation obtained was between TLI and TLX.

Intrasubject Analysis

Table III shows intrasubject analysis (columns 8 to 10). Among the 18 aircraft controllers who rated more than 1 level of workload through the TLX test, 14 (77.8%) showed better correlation between TLI and TLX than between N and TLX. Perhaps with the exception of subject 10, those controllers who identified only one workload level were those who worked at the lowest values of TLI (Table III). All indices were below the average values of N (3.7) and TLI (5.4) except for controller 10 (mean N and TLI were 5.1 and 6.8, respectively).

Comparison Between Two Models

To explain the TLX variance, TLX may vary as a function of N [TLX = f(N)]. Secondly, TLX may vary as a function of N and TLI [TLX = f(N) + f(TLI)]. It is hypothesized that the weight of TLI is different from 0, e.g., the two models are different. The analysis of covariance implemented for the comparison of these two models showed that the sum of squares = 173085, $F = 35.7$, and $p < 0.0002$.

DISCUSSION

Subjective assessments of workload have the advantage of being related to the overall cost of a task to an operator. The main problems are to identify the factors which come in addition to objective data, and to propose a method of quantification with a view to assessing workload through numerical values. There were three main factors selected in ATC: monitoring, vectoring, and conflict solving. Such evaluation was carried out on the basis of objective data, i.e., aircraft positions on the radar screen and the evolution of paths. However, it was assumed that possible consequences of decisions would increase workload, especially in the event of conflicts. This additional load was quantified through rules described in the method section. Thus, TLI was computed by taking into account both objective data, i.e., the number of aircraft to be controlled (cognitive processes) and subjective data, i.e., the additional load attached to each situation (emotional processes, mainly). The TLI was computed after the sessions but integrated real-time conditions from current flight traf-

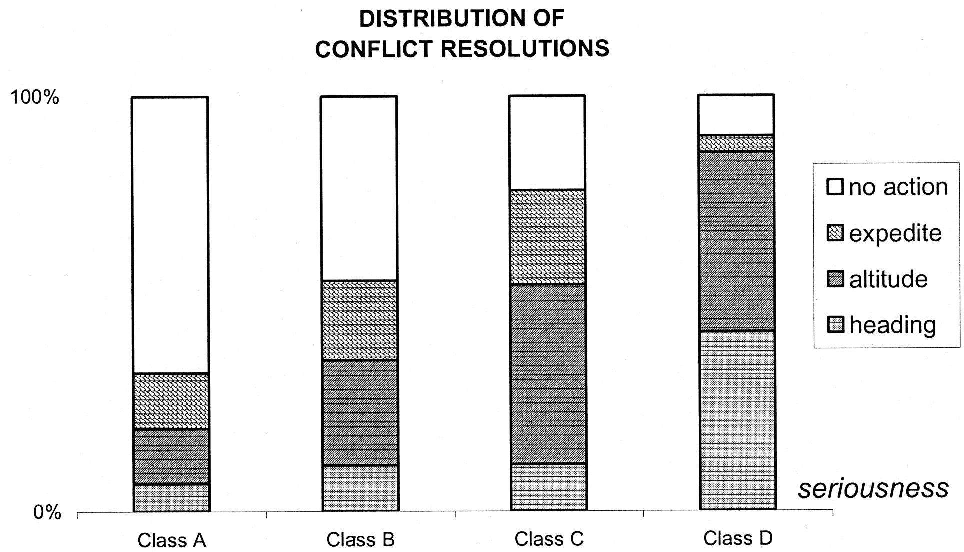

Fig. 4. Distribution of ATC solutions over the conflict classes. Class A = expected spacing is inadequate to ensure total safety, class D = maximal risk of conflict. Class B and class C = intermediate levels of separation loss probability.

fic data, and even by anticipating future potential situations. As TLI was computed every 10 s, a high-resolution time-scale was used, thus following radar screen updating (from 4 to 8 s). Task analysis, centered on the controller's behavior and elements of a theoretical cognitive model (MT, an assessment of temporal distribution of resources), supplied the framework for TLI computing.

The particular case of "no action" will be discussed first, followed by the study of resolutions related to conflicts. Then comparisons of the three indicators described in assessing aircraft controllers' mental load will be discussed. As assumed, "no action" should not be interpreted as a lack of decision. On the contrary, it should be considered a choice, a decision made as a result of situation analysis (expected evolution of flight variables). The advantage of making such a choice is a lessening of coercion on aircraft. This follows from the fact that the controller considers his/her job as that of supplying a service to aircraft on normal, regular flight paths. Here attention can be considered higher than during previous and tangible actions. This is obviously a costly strategy throughout the conflict with regard to cognitive resource allocation.

On the other hand, action on "altitude" and even on "heading" can give early safety against loss of spacing. Although an increase in these safety resolutions at the same time as the decrease in "no action" and "expedite" decisions increases the seriousness of the conflict, they are easily understandable if one takes the initially assumed MT process into account. Thus, as seriousness increases, "heading" and "altitude" decisions increase, whereas "no action" and "expedite" decisions decrease simultaneously. This is consistent with the basic characteristic of these four types of solution; action on "altitude" and "heading" present the advantage of settling the conflict immediately. This is not the case with the two other solutions, which become less relevant when conflicts seem to be sure to occur. This result shows that the empirical classification of conflicts through seriousness can be quite close to a classification proposed by a significant number of SMEs.

The NASA-TLX rating scale has often been shown to be an appropriate tool in self-estimation of workload. It was carried out after the work session and was based on recall of events which occurred during the 1-h work session. The TLI was computed on the basis of data available online through radar recordings. Integrating such objective data may explain why the correlation between TLI and N was high. Despite disparity between TLI and TLX time scales, TLI is shown to be more highly correlated to TLX than N to TLX. Conversely, the correlation between TLX and N was lower, meaning that N does not exactly reflect the perceived workload. This conclusion is emphasized when the correlation between maximal values of N, TLI, and TLX is considered. Taking into account the periods of highest load, the correlation clearly indicates that N and TLX were poorly associated. This means that the mere number of aircraft does not match the workload perceived by controllers at best, particularly when numerous conflicts occur, leading the controller to make decisions under high strain conditions. Finally, the comparison between two models explaining TLX variance showed that the weight of TLI in workload perceived by controllers was very different from 0. This speaks for the integration of subjective data into workload evaluation, as with TLI.

At this stage of research, it should be noted that numerical values evaluating seriousness and time boundaries were set up from radar recordings and computed by a single SME. However, the TLI is based on simple rules which could be similar despite possible particularities related to airports and procedures. Possible approximations related to numerical quantification do not impair the ability of this index to represent the greater part of the effect of objective traffic data on mental load. To enhance TLI reliability, the validity of seriousness and time boundaries must be generalized. The next step is to demonstrate the validity of TLI by asking a set of SMEs to compute it from different air-

tapraid5/z4j-aviat/z4j-aviat/z4j00404/z4j1679d04g | gockleyj | S=4 | 1/29/04 | 6:00 | Art: ASEM1052

traffic situations using the decision rules which are described here.

Building a general work model is required: algorithms should be capable of integrating the relevant traffic features (expected spacing, conflict geometry and attitude, etc.). Furthermore, human factors still have to be investigated and related to air-traffic data by taking emotional reactivity through conflicts into account. Future experimental approaches to this issue should involve neurophysiologists and engineers, the aim being to build up an expert system capable of computing TLI in real time. Such studies are currently being carried out. A correlation with psychophysiological variables works toward validation of TLI, which is shown to give a close estimation of the final cost of the task to the controller (3). These results should contribute to a better understanding of actual workload in ATC.

Emotional processes have for a very long time been considered to impair factors in human performance. Many examples could show that excessive emotional reactivity affects performance dramatically. However, this dated view should now be reconsidered in view of recent findings related to the role of emotion in decisionmaking (5). Emotional factors, which are essential to the mental processes, are integrated at each step of information-processing and influence cognitive activity.

REFERENCES

1. Averty P. Les effets de la charge de trafic sur le niveau d'activation psychophysiologique du contrôleur aérien [dissertation]. (Effects of workload on psychophysiological activation level in air-traffic controllers). Lyon, France: Université Lumière Lyon II; 1998.
2. Brookings JB, Wilson GF, Swain CR. Psychophysiological responses to changes in workload during simulated air-traffic control. Biol Psychol 1996; 42:361–77.
3. Collet C, Averty P, Delhomme G, et al. Subjective aspects of mental load in air-traffic control. In: Hockey GRJ, Gaillard AWK, Burov A, eds. Operator functional state: The assessment and prediction of human performance degradation in complex tasks. NATO ASI series, Series A: Life Sciences. New York: Plenum Press; 2003:291–302.
4. Costa G. Evaluation of workload in air traffic controllers. Ergonomics 1993; 36:1111–20.
5. Damasio AR. Emotion and the perspective of an integrated nervous system. Brain Res Rev 1998; 26:83–6.
6. Gaillard AW. Comparing the concepts of mental load and stress. Ergonomics 1993; 36:991–1005.
7. Hart SG, Staveland LE. Development of NASA-TLX: Results of empirical and theoretical research. In: Hancock PA, Meshkati N, eds. Human mental workload. Amsterdam: Elsevier; 1987: 139–83.
8. Hendy KC, Liao J, Milgram P. Combining time and intensity effects in assessing operator information-processing load. Hum Factors 1997; 39:30–47.
9. Kahneman D. Attention and effort. New York: Englewood Cliffs Prentice Hall; 1973.
10. Mogford RH, Guttman JA, Morrow SL, Kopardekar P. The complexity construct in ATC: A review and synthesis of the literature. Atlantic City, NJ: Federal Administration; 1995. DOT/FAA/CT-NT95/22.
11. Niessen C, Eyferth K, Bierwagen T. Modelling cognitive processes of experienced air traffic controllers. Ergonomics 1999; 42:1507–20.
12. Stein ES. Human operator workload in air traffic control. In: Smolensky MW, Stein ES, eds. Human factors in air traffic control. New York: Academic Press; 1998:155–83.
13. Wickens CD. Processing resources in attention. In: Parasuraman R, Davies DR, eds. Varieties of attention. Orlando, FL: Academic Press; 1984:62–103.