

Article

Extraction of CD&R Work Phases from Eye-Tracking and Simulator Logs: A Topic Modelling Approach

Aida Nordman ^{1,*}, Lothar Meyer ², Karl Johan Klang ¹, Jonas Lundberg ¹ and Katerina Vrotsou ¹

¹ Department of Science and Technology, Linköping University, 581 83 Linköping, Sweden; karl.johan.klang@liu.se (K.J.K.); jonas.lundberg@liu.se (J.L.); katerina.vrotsou@liu.se (K.V.)

² LFV Swedish Air Navigation Service, 601 79 Norrköping, Sweden; lothar.meyer@lfv.se

* Correspondence: aida.nordman@liu.se

Abstract: Automation in Air Traffic Control (ATC) is gaining an increasing interest. Possible relevant applications are in automated decision support tools leveraging the performance of the Air Traffic Controller (ATCO) when performing tasks such as Conflict Detection and Resolution (CD&R). Another important area of application is in ATCOs' training by aiding instructors to assess the trainees' strategies. From this perspective, models that capture the cognitive processes and reveal ATCOs' work strategies need to be built. In this work, we investigated a novel approach based on topic modelling to learn controllers' work patterns from temporal event sequences obtained by merging eye movement data with data from simulation logs. A comparison of the work phases exhibited by the topic models and the Conflict Life Cycle (CLC) reference model, derived from post-simulation interviews with the ATCOs, indicated that there was a correspondence between the phases captured by the proposed method and the CLC framework. Another contribution of this work is a method to assess similarities between ATCOs' work strategies. A first proof-of-concept application targeting the CD&R task is also presented.

Keywords: data science; complexity and machine learning in Air Traffic Management (ATM); human factors; situation awareness; conflict detection and resolution; en-route control; air traffic control; eye-tracking



Citation: Nordman, A.; Meyer, L.; Klang, K.J.; Lundberg, J.; Vrotsou, K. Extraction of CD&R Work Phases from Eye-Tracking and Simulator Logs: A Topic Modelling Approach. *Aerospace* **2023**, *10*, 595. <https://doi.org/10.3390/aerospace10070595>

Academic Editor: Álvaro Rodríguez-Sanz

Received: 28 April 2023

Revised: 11 June 2023

Accepted: 26 June 2023

Published: 29 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The creation of models that capture Air Traffic Controllers' (ATCOs) work patterns when solving particular problems, such as Conflict Detection and Resolution (CD&R) tasks, is gaining increasing attention from the research community. The motivation is manifold. Air Traffic Management (ATM) is currently progressing rapidly toward automation and digitalisation assistance to support the ATCO. Examples are AI-supported decision-making, which helps en-route ATCOs to solve conflicts or the Arrival Manager, proposing an inbound sequence of arriving movements, and many more that come along with remote tower technologies [1,2]. Automation has the purpose of relieving the ATCO from workload or decreasing uncertainty in the work quality by addressing a sub-task. Besides the desired effects, a side-effect of automation is that parts of the task spectrum that were not in the scope of the intended change may also be affected. Additional undesired effects include automation bias/complacency and out-of-the-loop effects [3]. This is critical, as ATCOs perform their work under efficiency and safety constraints and need to manage their attention and cognitive resources according to the traffic situation at hand.

Further research is needed to find methods that build models to capture controllers' activity patterns, with a focus on visual attention. Existing techniques still do not provide enough support to system designers and safety assessors in understanding the effects of automation in the ATCOs' work strategies. This concerns in particular the sequence of decision logic, involving the actions of information gathering, perception, and clearances.

Another aspect is that automatic activity pattern recognition, in the ATCOs' methods of working, can be used to create benchmarks and aid the assessment of training efforts. Alternatively, the evaluation of the influence of performance-shaping factors such as stress and fatigue can also benefit from such built knowledge. To build such models is, however, a non-trivial problem, since work patterns can involve many different activities (e.g., giving clearances, acquiring information about aircraft separation, etc.), and the events' composition can be significantly affected by time and other external factors. In addition, humans tend to solve problems by interweaving different work steps, leading often to concurrent and overlapping activities.

Though several authors have proposed models to capture the cognitive processes and problem-solving strategies of the controllers, those models either are not learned automatically from data collected in conjunction with the problem solving [4–7] or the models are mostly used to replicate controller strategies [8,9]. What we aimed to achieve is a method that reveals controllers' work strategies as patterns learned from raw data. More concretely, this work puts forward a method to extract ATCOs' work phases and their characteristic work patterns underlying the CD&R task, from collected eye-tracking data and simulator logs. To this end, an unsupervised learning technique was used based on topic modelling [10] applied on temporal sliding windows over event streams obtained by merging the eye-tracking data with the simulator logs. Combining the two datasets allowed us to identify look events as dynamic areas of interest and compare them with the information cues [11] collected in ATCO interviews. An advantage of the proposed method is that it caters to the possible concurrent and overlapping activities of ATCOs' work. In addition, we propose an approach to assess similarities between ATCOs in terms of the strategies used to solve a task. This aspect can be particularly relevant for ATCO training as their individual strategies can be contrasted with, e.g., best-practices or more-experienced ATCOs' problem-solving approaches. Evaluating the effect of automation on the ATCOs' strategies can be another potential application area of our method.

A statistical procedure revealing common aspects in a heterogeneous population of individual strategies is another contribution of this research. Finally, the learned work phases were then validated by comparison with the steps and information cues of the Conflict Life Cycle (CLC) proposed in [11].

This paper is structured as follows. In Section 2, we review the related work. The experimental setup used for the data collection is outlined in Section 3. Section 4 presents a topic-modelling-based method to reveal and characterise ATCOs' work phases from the data. The results are provided in Section 5, followed by a discussion in Section 6. The conclusions, limitations of the presented method, and future work are given in Section 7. Note that Section 3 only gives an overview of the practical experiment conducted for data collection, to aid the reader's understanding of the remaining sections of this paper. The experimental design utilised is thoroughly described in [11].

2. Related Work

In this work, we deal with data in the form of event sequences, i.e., sequences of discrete events, each of which is characterised by an event type, a timestamp, and a duration. In the current context of Air Traffic Control (ATC), an event is, for instance, looking at a specific waypoint or activating the aircraft separation tool at a certain point in time.

Several authors [12,13] have used techniques based on sequential pattern mining for pattern identification in event sequences. Perer and Wang, for example, proposed Frequence [14], an interactive tool built around the SPAM algorithm for detecting and visualising frequent patterns from event sequences. Vrotsou and Nordman [15] introduced Eloquence, a prototype system, based on an adapted pattern-growth approach, for interactively exploring patterns. Through a visual interface, the system allows a user to apply local constraints, grow patterns stepwise, and thus, steer the search according to their analysis interest. The approach was applied to ATC for exploring patterns in tower control in [16].

Overall, these approaches are highly sensitive to the order in which events appear when identifying patterns, and consequently, they are less-suitable to model activities where the events' order is less strict. Therefore, considering that the eye-tracking data we dealt with in this work are characterised by continuous back-and-forth shifts between elements on the screen, we did not pursue this approach.

Others have merely described raw sequences [17] in tower ATC; or used questionnaires for a self-assessment of sequences [18]; or used pre-defined sequences that are compared with pre-defined areas of interest [19]. Another simplified approach is to use dwell times on pre-defined screen positions [20] or search entropy (rather than specific patterns) [21] to establish the degree of task expertise. All of these approaches could benefit from some way of also measuring and estimating or establishing reference gaze patterns more automatically.

Approaches have been suggested based on regular expressions for exploring event patterns with more-relaxed conditions regarding the order of events. Zraggen et al., for example, proposed (s|q)ueries [22], a visual query interface for creating expressive queries on event sequence data. Cappers and van Wijk [23] introduced an approach based on regular expressions allowing the exploration of multivariate event sequences on both the multivariate data and sequential level. Such approaches, however, are less-suitable for our purpose since they generally assume that the user already knows what to query for.

Topic modelling [24,25] is an unsupervised machine learning technique that originated from the field of Natural Language Processing (NLP), and it has had its applications extended to diverse areas such as bioinformatics [26], computer vision [27], audio, and music [28]. Nguyen et al. [29] used topic models to identify human activities in event sequences obtained from server logs. Ozmen et al. applied topic modelling to event sequences of Electronic Health Records (EHRs). An important distinction is that our method, in conjunction with topic modelling, uses a sliding window over the sequences of events to be analysed. In the ATM field, to our knowledge, topic models have only been applied for automatic analysis of aviation safety incident reports [30]. We did not find any applications related to ATC.

Several machine learning techniques have been used by researchers to address the problem of automation in ATC and to learn ATCOs' work tactics. Conflict detection is the focus of the work presented in [31], where a method based on classification and regression was used to predict separation infringements between aircraft. The methods presented in [8,9,32] focus on learning from ATCOs' conflict resolution strategies and analyse the ATCOs given commands collected through human-in-the-loop experiments conducted in a simulation environment. The framework proposed in [32] is based on an ensemble model of regressor and classifier chains, i.e., a supervised technique. However, it does not expose the features of the learned strategies. On the contrary, our method reveals the characteristics of the ATCOs' strategies in terms of the tools and elements (e.g., waypoints) looked at. The systems described in [8,9] use reinforcement learning and convolutional networks, respectively, and seek to mimic the controllers' decision logic. Unlike these frameworks, the approach presented here aims to model controllers' behaviour, in the form of work phases and event patterns, considering the CD&R steps of *conflict detection*, *conflict solution probing*, and *solution monitoring*. A distinctive aspect of our method is that eye-tracking data were also used, besides the data collected from the simulator logs originated from human-in-the-loop experiments.

3. Human-in-the-Loop Experimental Setup

The performed study aimed at identifying the work phases from ATCOs' eye-tracking look events during a human-in-the-loop en-route conflict scenario and to compare these work phases with results from follow-up interviews conducted with the ATCOs. The comparison of subjective interview data and empiric response data to a conflict scenario shall give proof to the validity of the proposed method.

Therefore, the study relied only on observations instead of variations that came from changing the input parameters to intentionally influence the work behaviour. We assumed that any observed variation of ATCOs' look events is the result of different backgrounds in terms of education, work experience, age, gender, and many other factors, which impose individual work behaviour on each ATCO. If the conflict scenario remains the same for all ATCOs, any inter-individual variation must be an effect of this individual work behaviour and, thus, supports the assumption. As we only compared observations, the results are descriptive in nature without drawing any conclusion about the cause of the observation. The study design complies as such to the "Simple ex-post facto design" [33], which allowed us to set the focus on a mere description of the difference and similarity of the work behaviour between ATCOs in response to the same conflict scenario.

Data were collected from 15 operative ATCOs (5 female, 10 male) with valid licences, who performed in total 6 working scenarios in randomised order in an en-route air traffic simulator, NARSIM. In comparison, Reference [34] relied on one ATC for the gaze samples in their study; Reference [21] had 18 domain experts (and a non-expert group); Reference [17] had 15 retired ATCOs. For further details about the data collection, see our previous paper with qualitative findings. A detailed description of how the practical experiment, leading to the data collection, was designed and conducted is given in [11].

In this paper, we continued analysing the collected data [11] from one scenario for the verification of the qualitative findings. From the 15 ATCOs, we received 14 complete datasets of eye-tracking data and simulator log data (5 female, 9 male). During this process, two datasets were created, for each ATCO: a simulator log and an eye-tracking dataset. Next, each pair of datasets was merged into one data stream (i.e., one data stream was created for each ATCO). The 14 data streams obtained could then be mined to discover work phases underlying the working process of the ATCOs.

3.1. The Chosen Scenario

The scenario used for validation was chosen because we had collected qualitative data from post-debriefings with the ATCOs. Of the 15 ATCOs who completed the simulations and debriefing, 13 ATCOs also completed an additional Retrospective Think-Aloud commenting session (RTA), while replaying this particular scenario for them. During this session, their own gaze point was shown in the video depicting where they had been looking during the scenario. Therefore, a comparison could be made between the merged datasets and the collected comments about the chosen scenario. For those 13 ATCOs, 4 were female and 9 were male.

Even though we collected more data in more complex traffic scenarios, we did not have qualitative comments from the ATCOs about those scenarios to compare with. More complex scenarios involving several conflicts at a time may appear more realistic to ATCOs. We decided, however, to choose a simple single conflict scenario because, so far, the CLC reference model (depicted in Figure 1) has been validated for this particular simple scenario. Therefore, the ATCOs' work behaviour and related cognitive modes can be unambiguously mapped to one of the steps described by the CLC. Using a more-complex scenario with multiple simultaneous conflicts, on the other hand, may raise the need for a more-complex cognitive model, where multiple instances of the CLC can interact. This will be explored in future work, where we intend to extend the cognitive model to include the capability to map multiple conflicts as well.

The chosen scenario contained four aircraft: an Airbus 320, a Boeing 737, and two Boeing 777 s. The scenario is shown in Figure 2. The sector was a square, 55×55 nm in size. The medium-sized Boeing 737 and heavy-sized Boeing 777 were in conflict as their flight-plans crossed at the same level (FL360) at a 90-degree angle. The other aircraft, on opposite courses with the ones in conflict, acted as constraints to solving the conflict. The other Boeing 777 was crossing the sector 1000 ft below the aircraft in conflict, at FL350. The Airbus 320 was crossing the sector 1000 ft above the aircraft in conflict, at FL370. The simple solution would be to climb or descend one of the aircraft. However, the other aircraft

restricted a climb solution or a descend solution. Unless the ATCO intervened, separation was lost at around 05:34 (5 nm between the two aircraft and closing). The Closest Point of Approach (CPA) was 0 nm and occurred 05:59 into the scenario. The reason for using a simple, generic, scenario with only one conflict situation was to test our method with data collected from a simple-enough, yet realistic and well-understood, scenario. The use of only one conflict situation means that the collected eye-tracking data mainly reflected this one situation, making it less prone to noise from the overlapping processes of, e.g., dealing with multiple conflict situations. Nevertheless, the conflict scenario was typical for the type of situations that ATCOs in training receive as an introduction to solving conflicts.

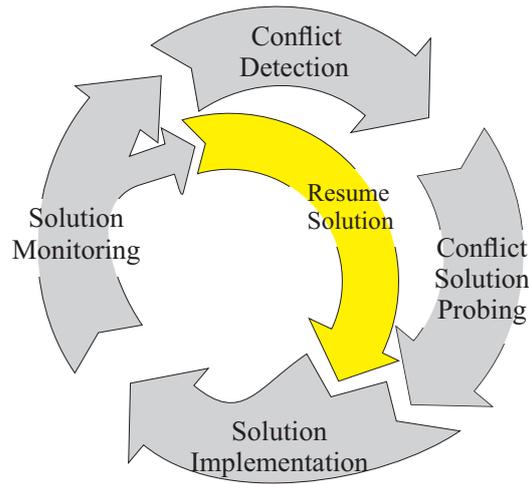


Figure 1. Conflict life cycle presented in [11].

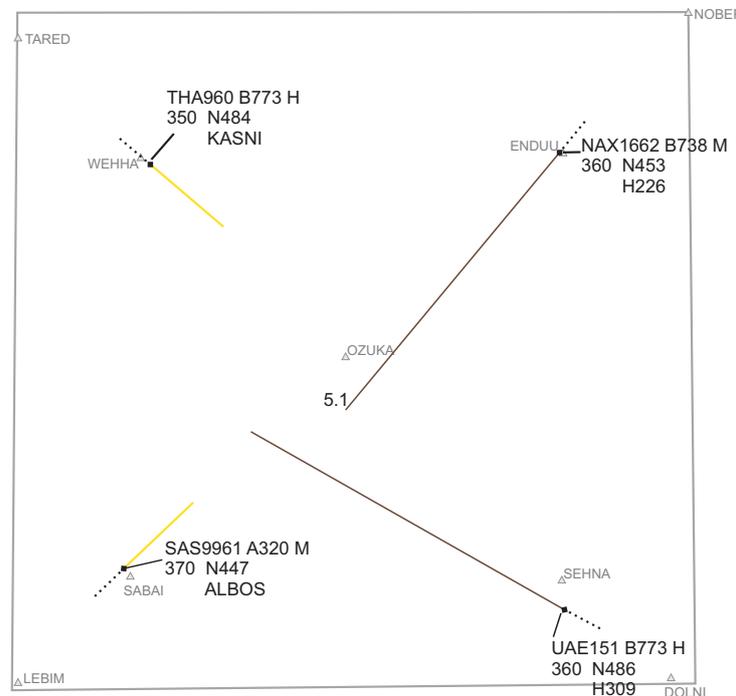


Figure 2. Scenario involving four movements with two of them in conflict and activated separation tool.

3.2. Simulator Data

The NARSIM simulation platform was used as the en-route ATC simulator to run the chosen scenario. The primary working instrument involved a 2D situation display (“radar”-like) of the respective sector, showing the aircraft as squares with trailing dots to indicate direction on the horizontal plane, sector borders, and waypoints. The interface

provided support tools in terms of speed vectors, a separation tool (*sepTool*), and a conflict display window. The activity of the controllers during the simulations was logged by the simulator, which included clearances given, aircraft conflicts selected, and the activation of *sepTool*. In addition, the screen positions of the graphical objects, such as the aircraft representations and the simulator conflict-detection tools, were also logged. Aircraft representations included the aircraft tracksymbol, clickable information label, and various movement-related graphical elements (which could be turned on and off) to symbolise direction and separation distance.

3.3. Eye-Tracking Data

While the ATCOs' interactions using the mouse were logged through the simulator software, a SmartEye eye-tracking system was used to capture and record participants' visual activity during the simulation sessions. The equipment recorded eye-gaze movements at a sampling frequency of 60 Hz and calculated the screen coordinates from this. Thus, the eye-tracking data were measurements of the eye-gaze point coordinates on the radar screen.

3.4. Merging the Raw Data

The eye-tracking data and simulator log, collected for each ATCO, were then merged into a Human–Machine Interaction (HMI) stream of visual and mouse interaction information. The data-merging process checked for timewise intersections between the coordinates of the eye-gaze points and the positions of the graphical objects on the screen. Descriptive *look events* were created as intersections of gaze points and the graphical objects with similar timestamps, while *unknown* (no-match) events were created otherwise. *Interaction events* were added to the data stream, whenever the ATCO interacted with the simulator using the mouse. Such interactions corresponded to information querying (e.g., clicking on a label), switching graphical support tools on/off (e.g., aircraft separation information), and giving clearances (e.g., to change direction or flight level). Due to the outlined event-extraction process, *look events* always had a duration, while *interaction events* were treated as instantaneous and, therefore, were assigned a short default duration. For this reason, the interaction events were often very sparse and short.

We refer to some of the properties of the event streams created by the merge process described above:

- A large number of event types occurring in the merged stream (e.g., over 100).
- The streams are *noisy* due to a large amount of *unknown* events. These events were generated whenever it was not possible to determine the intersections of graphical objects on the radar with eye-gaze points or when the eye-gaze points fell outside the radar screen. Between 20% and 36% of the events in each of the generated event streams were *unknown* events. In addition, look events with a very short duration (e.g., below 100 ms) also added noise because one cannot assume in these cases that the subject could perceive any object on the radar screen [35].
- There was *high variability* in the data with respect to event duration and frequency. While some events were rare (some look waypoint events), others occurred hundreds of times. The standard deviation (sd) for the events' duration was also high. For instance, the duration of the *look aircraft's* trajectory event `look+MAX1662+TRAJ` for one of the ATCOs varied from 16 to 381 ms, $sd \approx 319$. We considered that look events with a very short duration originated from saccades. Therefore, these were eliminated from the data analysis by setting a minimum duration threshold for look events of 100 ms.

The points above made the search for work patterns from the data streams a challenging process.

4. Discovery of High-Level Phases for CD&R from HMI Streams Using Topic Modelling

In this section, we describe a method based on topic modelling to discover high-level work phases performed by ATCOs who solved the air traffic CD&R problem in the

selected scenario. The analysed data consisted of the HMI event streams (i.e., time series) obtained by merging eye-tracking data and simulator logs, for each ATCO, as described in Section 3.4.

The work reported in [11] had two important outcomes, which are used in this section. Firstly, information cues were identified such as *Predicted Top-of-Descent (ToD)* and *Predicted Separation Minimum Distance (PSMD)*. These information cues were used by the ATCOs to support them in the decision-making process of establishing a strategy for tackling a simulated aircraft conflict (depicted in Figure 2). As described in [11], the information cues were derived from ATCOs' statements obtained by post-simulation interviews, where the ATCOs described their strategies to tackle the conflict in the chosen scenario. Secondly, the information cues were then related to the work steps of the CD&R task modelled as a Conflict Life Cycle (CLC) inspired by the framework proposed by Pawlak [4]. These four work steps in the CLC presented in [11], and depicted by the four grey areas of Figure 1, were the main motivation for exploring our topic-model-based approach with four topics. Through topic modelling, we attempted to elicit the ATCOs' work phases from the HMI streams and characterise them in terms of events occurring in the streams. Moreover, the information cues played a key role in relating a topic model to the CLC. This connection was then used as part of our validation, described in Section 4.3.

4.1. Topic Models' Overview

We start by giving a brief overview of topic models. Topic modelling [24,25] is a machine learning technique with its roots in the field of Natural Language Processing (NLP). Its purpose is to automatically annotate large archives of documents with themes, usually called topics. Discovered topics are denoted as probability distributions over the words occurring in the analysed documents. Additionally, each document is also associated with a probability distribution over the extracted topics, reflecting the intuition that documents are usually composed of several topics, though topics may occur in different proportions in each document. For example, consider that the documents are newspaper articles. Then, a topic T_0 could be expressed as $0.3 \times \text{price} + 0.2 \times \text{market} + 0.28 \times \text{capital} + \dots$, where *price*, *market*, and *capital* are words in the articles. The distribution over topics (as retrieved by the model) for an article could be $0.6 \times T_0 + 0.35 \times T_1 + 0.05 \times T_2$, assuming the model is set to retrieve three topics T_0 , T_1 , and T_2 .

More formally, topic models reveal the hidden structure in an observed collection of documents. The hidden structure corresponds to the revealed topics, the per-topic word probability assignments $p(w_j|T_i)$, and per-document topic assignments $p(T_i|d_k)$. These probabilities can be interpreted as the importance of a word w_j in a topic T_i and the importance of a topic T_i for a document d_k , respectively. The relation between the probability of a word w_j to occur in a given document d_k is expressed as $p(w_j|d_k) = p(w_j|T_i) \times p(T_i|d_k)$, such that the probabilities $p(w_j|d_k)$ can be estimated directly from the observed collection of documents by counting the words in each document. However, the computation of the hidden structure, in the form of the probability distributions $p(w_j|T_i)$ and $p(T_i|d_k)$, is an intractable problem because the number of possible assignments of each observed word to topics is exponential. Consequently, existing topic-modelling algorithms can only compute approximations of these distributions by different methods. In our work, we used the well-known algorithm named *Latent Dirichlet Allocation* (LDA) [10], which is based on a sampling procedure. More concretely, the implementation of LDA offered by the open-source Python library Gensim was used in the presented work. LDA has two hyperparameters corresponding to Dirichlet distributions, α and β , which represent prior values for the document–topic and topic–word distributions, respectively. An empirical Bayes method can be used to estimate these distributions from the data [10]. We used this technique to automatically set the values of α and β . Table 1 shows a summary of the settings used for training in this study.

Table 1. Hyperparameters used for topic model training with LDA in Gensim library.

Parameter	Value
Number of topics (k)	4
α	estimated by LDA from data
β	estimated by LDA from data
Epochs	20
Iterations	400

Topic-modelling algorithms can also be seen as soft clustering methods, where discovered topics correspond to clusters and the probability distributions for per-document topic assignments reveal to which extent documents may belong to different clusters. Like popular clustering algorithms, several topic-modelling algorithms such as LDA require the user to pre-select the number of topics to be uncovered from the observed documents. Unlike clustering algorithms, there is no need to define a distance measure.

Next, we describe how topic modelling was used to analyse the event streams.

4.2. Phases' Estimation from HMI Event Streams

The HMI data streams, obtained by merging eye-tracking data with simulator logs, were time series of events. Each event was a tuple of three elements (id, t, d), where id is the event name, t is the start time of the event, and d is its dwell time measured in milliseconds (ms). As described in Section 3.4, there were two main types of events, look events and commands. For instance, `cmd+SAS9961+show_TRAJ` indicates that the ATCO, at some point in time, activated the tool that shows the flight leg for SAS9961. The data streams also contained noise in the form of unknown events.

Addressing realistic CD&R problems, as posed in the scenario prepared for this work, is a non-trivial task. Due to the complexity of the cognitive processes involved, it seems reasonable to assume that the procedure for solving a conflict between two aircraft involves several work phases (or steps). Our goal was to discover such possible work phases, when they were activated and deactivated, and reveal which event patterns occurred in the phases.

The method we propose here was based on topic modelling, and its main stages are illustrated in Figure 3. First, we needed to create “documents” from the event streams (Figure 3a). To this end, we used a sliding window of 30 s. The window slides over a stream by shifting five seconds each time, as illustrated in Figure 4. Then, the subsequence of events within each window W_i corresponds to a document. The size of the time window was chosen empirically. Considering the fast pace and quick changes of ATCOs' tasks and the high resolution of the event streams, 30 s was deemed a reasonable time that would capture potential work patterns.

To be able to obtain meaningful results, the data in each time window had to be cleaned from noisy events. Therefore, unknown events and look events with a dwell time of less than 100 ms were eliminated from each window. It is reasonable to assume that the participant could not have perceived an object on the simulator display if the event's duration was below 100 ms [35]. Using the HMI event streams extracted for the 14 participants in our study, we then obtained 1268 documents with an average size of ~166 words (events) per document.

The topics to be retrieved from the collection of documents can be seen as phases in the work performed by the ATCOs while solving the aircraft conflict in the scenario chosen for this study. It is then possible to discover when each of the phases was activated (including the “activation's level”), for each participant. Recall that each document was associated with a time window of 30 s on a participant's event stream. More concretely, the following procedure was performed (Figure 3b). After the analyst had chosen the number of work phases (topics) $k > 1$, a topic model was trained using the “documents” obtained from the event streams of all the other 13 participants, apart from a chosen participant P . Next, the model was applied to each of the “documents” obtained from the event stream of

participant P . In this way, the topic model maps each time window obtained from P 's data with a probability assignment of phases. In other words, by applying the model to P 's data, one obtains a time series of k -D vectors, where each vector represents the level of activation of each phase within a time window.

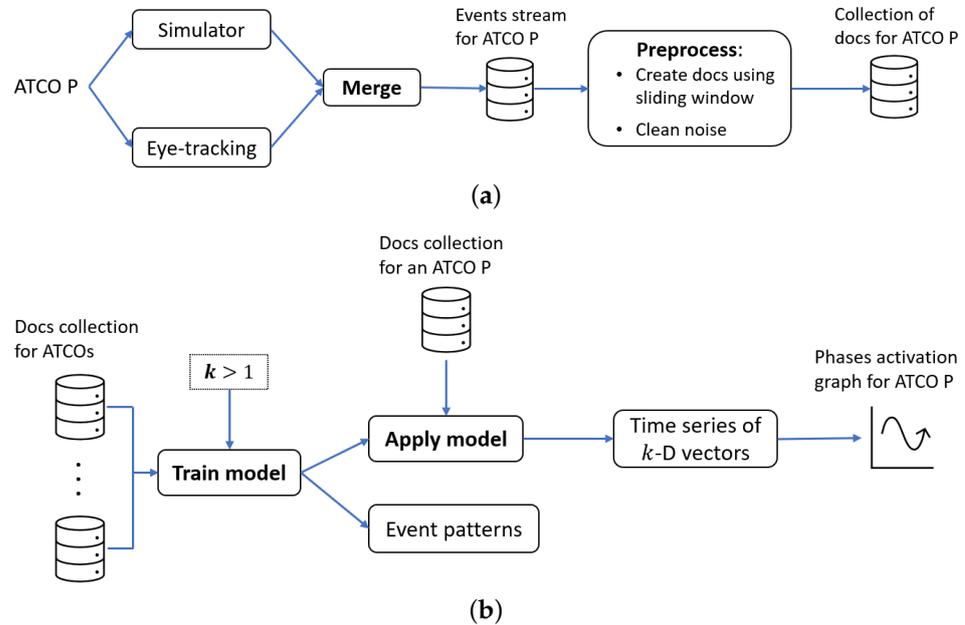


Figure 3. The pipeline for the proposed method, where k is the chosen number of work phases (topics). (a) Generation of a “collection of documents” for an ATCO. (b) Generating a phase activation graph for an ATCO. Figure 5b shows an example of event patterns characterising work phases.

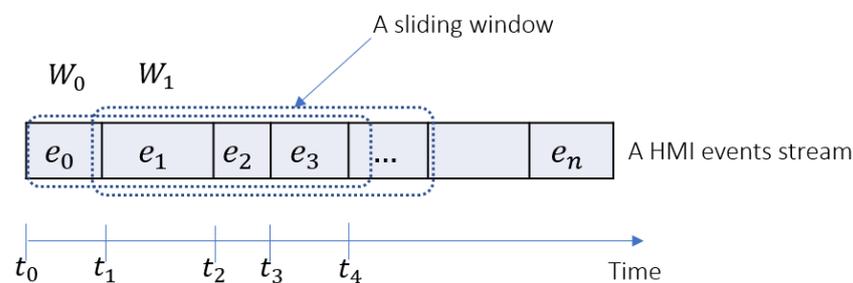


Figure 4. Sliding windows of 30 s over an HMI stream.

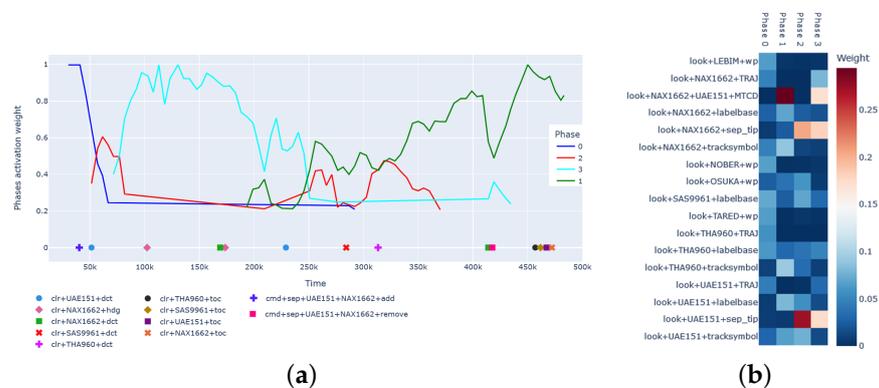


Figure 5. Activation of work phases for a participant. (a) The X-axis represents time in milliseconds, while the Y-axis shows the phases’ activation level. For simplicity, levels of activation below 0.2 are not shown. Clearance and open/close sepToo1 events are marked out in the figure to provide additional context to the phase activation levels. (b) Heat map characterising phases in terms of events.

Table 2 shows some of the sliding windows start and end times (first two columns), obtained from an HMI data stream. It is possible that the last event starting in a time window would not have its end time in the same time window. To deal with this situation, a time window's duration was stretched to completely accommodate all events starting in it. The last four columns (named "Phase 0", "Phase 1", "Phase 2", and "Phase 3") correspond to the 4D vectors associated with each time window.

Table 2. Sliding windows and corresponding 4D vectors, which are visualised in Figure 5.

Start Time	End Time	Phase 0	Phase 1	Phase 2	Phase 3
0	30,230	0.9970471	0.0012459053	0.0011321253	0.0005747694
5633	35,805	0.9971619	0.001195153	0.0010889024	0.00055401414
10,666	40,845	0.99707013	0.001232266	0.0011270973	0.000570499
16,048	46,128	0.8306506	0.0011075849	0.16772854	0.00051324465
21,048	51,385	0.6561338	0.001138541	0.3422003	0.0005273581
...					

4.3. Mapping the Topic Model to the CLC Work Steps

We assumed that each work phase represents a specific cognitive mode that follows a specific purpose and operator intention, thereby eliciting a significant composition of events. The probability distribution related to the events of the topic model provided a significant pattern that could give insight into this purpose and intention, thus giving the topic model an operationally relevant context. As part of our validation, we related the topic models determined by the HMI data streams to a reference model, the Conflict Life Cycle (CLC) model [11]. The work steps of this model represent basic cognitive modes during a simple CD&R task and 15 related information cues (listed in [11], Table 1) such as *Predicted Top-of-Descent (ToD)*, *Waypoints (WPYs)*, *Predicted Separation Minimum Distance (PSMD)*, and others shown in the first column of Table 3.

We followed a two-step procedure: (1) The probability vector representing the per-phase event probabilities' assignments, of the phase (topic) model, was mapped to the information cues (mapped information cues); (2) we determined the degree of matching of the mapped information cues to the CLC work step specification, as given in Table 2 of the work presented in [11], relating work steps to their characteristic information cues.

We assumed the 15 information cues ($n = 15$) to be linearly related to the look events of the HMI data stream ($m = 34$), using a transfer matrix $\mathbf{A}_{n \times m}$ involving transfer parameters a_{ij} at each linking position ij . This way, we mapped each look event's probability to the corresponding information cue. The model for this method is a "left stochastic matrix" (A left stochastic matrix represents a real square matrix, with each column summing to 1), which maps one probability vector \vec{x} to another probability vector \vec{y} , the mapped information cues. In this specific case, however, both probability vectors were unequally long, thus forming a non-square (pseudo) stochastic matrix. For a 1:1 relationship (one look event i matches one information cue j), we set the transfer parameter $a_{ij} = 1$ to maintain the sum of probabilities per column throughout the mapping. There were a few circumstances to consider for the case when no 1:1 relation could be determined:

1. If one look event mapped on several r information cues (underdetermined relation), then the transfer parameters were set for an equal share across all cues using $a = 1/r$.
2. Multiple look events may map to a single information cue. In this case, we chose to assign a transfer parameter of $a = 1$ to each individual relation. A particular cue, thus, involved several look events by the sum of their probabilities.
3. Look events may not relate to any of the information cues (overdetermined relation). For this case, we assumed a hidden row "Other", which was used to catch events that were not covered by the list of information cues.

A similar approach was used for further calculating the degree of matching between the matched information cues and the specifications of three of the CLC work steps: “conflict detection”, “conflict solution probing”, and “conflict monitoring”. The step “solution implementation” of the CLC was not included in the matching. The reason was that this step was characterised by activities using the voice radio captured by “menu” and “clearance” events, while our focus in this analysis was on look events. For the remaining three working steps, we determined a second transfer matrix $\mathbf{B}_{n \times q}$, where $q = 3$ (Table 3).

The calculation of the matrix followed a two-step approach starting with the frequency of information cues as mentioned by the ATCOs [11]. Starting with the work step specification, we assumed the frequency of information cues to indicate the significance of a specific cue for the respective work step, creating a $n = 15$ -long vector. Secondly, the vector was normalised to a probability vector with a sum of one. This normalisation was necessary to avoid bias effects resulting from differences in the amount of information cues involved in the steps. Emphasising the qualitative distribution of work step information cues for quantifying matches, rather than the frequency of cues alone, now allowed for a comparison of the assignment results. Third, adding each of them in a column, we obtained a three-column non-square (pseudo) left stochastic matrix \mathbf{B} , shown in Table 3. The complete mapping and matching operations were rendered as $\vec{y} = \mathbf{B}^T \cdot \mathbf{A} \cdot \vec{x}$, providing a vector of three scalar products, which represented the degree of matching between the mapped information cues and the CLC work step specification.

Table 3. Transposed transfer matrix \mathbf{B} , which maps information cues to the CLC work steps. Derived from [11], Table 2.

	Conflict Detection (%)	Conflict Solution (%) Probing	Solution Monitoring (%)
Flight Level	44.0	02.2	00.0
Destination	04.0	08.7	00.0
Flight Route	24.0	15.2	00.0
Approaching Traffic	12.0	00.0	00.0
Expected Climb	04.0	04.3	00.0
Flight Plan	00.0	06.5	00.0
Predicted ToD	00.0	15.2	00.0
WPYs	08.0	04.3	00.0
PSMD	00.0	26.1	90.0
Traffic Complexity	00.0	04.3	00.0
Expected Change Level Request	00.0	02.2	00.0
Wind	00.0	08.7	00.0
Rate of Descent	00.0	02.2	00.0
Speed	00.0	00.0	10.0
Conflict Window	04.0	00.0	00.0

5. Results: Discovered Work Phases and Event Patterns

For building a model, we had to decide the number of phases (topics) $k > 1$ that should be retrieved from the data. Inspired by existing research [4,11] and discussions with experts, we decided to investigate models with four phases, i.e., $k = 4$. We are, however, aware that it might also be reasonable to build models with a different number of phases, such as $k = 2$ or $k = 3$, since the interpretation of the discovered phases lies on the field experts. For instance, a phase in one model might appear decomposed into sub-phases in another model with a larger number of phases. The number of phases might also be decided as a function of the scenario’s complexity.

5.1. Individual Participant Analysis

Our first goal in the data analysis process was to investigate the “behaviour”, in terms of work phases, of individual participants. To this end, we built a model using the data streams of 13 participants and applied the obtained model to a participant P left out of the training set. Figure 5a shows the phases’ activation over time for a participant, named

here as P_{14} . The per-window phase assignments are called here the phases' activation weight (which correspond to the probabilities retrieved by the model). We observed that, during the first 57 s, Phase 0 was activated to be then replaced by a short time period where Phase 2 dominated (for ~20 s). Then, Phase 3 became the dominant phase (i.e., the activated phase with higher probability) for ~174 s. Finally, from approximately Second 251, Phase 1 became dominant, though around Second 320, both Phases 1 and 2 seemed to be equally dominant. The fact that several phases can be activated simultaneously can be explained by the probabilistic nature of topic models, more concretely LDA. In practice, the work steps shown in Figure 1 might occur as concurrent and overlapping activities, instead of well-separated steps. Humans often tend to address problems by executing one task, going on to another task, to later come back to a previously initiated task. In addition, the end of one phase and the start of another may also overlap. The fact that several phases can be activated simultaneously in a topic model, which can be explained by its probabilistic nature, is an advantage of topic modelling in this context.

Another advantage of using topic modelling in our proposed method is that it also caters to revealing possible phases' descriptions in terms of events, as shown in Figure 5b. In other words, this figure is a visual translation of the events' per-phase (or word per-topic) probability assignments discovered by the model. From this figure, we can conclude that looking at the SAS9961 and THA960 label bases, looking at the trajectories of NAX1662, UAE151, and THA960, and looking at the waypoints NOBER, TARED, and LEBIM are between the most-distinctive features of Work Phase 0. This can be interpreted as the ATCO gathering information about the current traffic situation by monitoring the radar screen. Looking at the *Medium-Term Conflict Detection* tool (MTCDD) is a very distinctive feature of the final Phase 1, together with looking at the NAX1662, THA960, and UAE151 tracksymbols (or label bases), indicating that the ATCO might be monitoring the aircraft after conflict resolution. We also observe in Figure 5b that looking at the separation tool tips of aircraft UAE151 and NAX1662 was between the events with larger probabilities in Phase 2. We hypothesised that the transition from Phase 1 to Phase 2 (around Second 57) was an indication that the ATCO identified a conflict between aircraft UAE151 and NAX1662. It is interesting that looking at the waypoint OSUKA was more relevant in Phase 2 than in Phase 1. This is compatible with our observations that the ATCO's gaze points tend to be very centred on the middle of the sector, close to the centre waypoint OSUKA, after the conflict has been detected. This waypoint is located near where the aircraft will be at their closest point of approach to each other. Finally, looking at the separation tool tips of aircraft UAE151 and NAX1662 and looking at the MTCDD between these two aircraft were the most-distinctive features of Phase 3.

5.2. Summarising Work Phases for All Participants

Attempting to draw conclusions directly from all participants' phase graphs (i.e., in our case, 14 graphs similar to the one shown in Figure 5a) was neither easy nor scalable. The large variability on the way the different phases overlapped during certain periods of time was possibly a consequence of the controllers' personal strategies and individual decision-making processes.

A statistical approach was then used to leverage the investigation of common aspects related to how participants tackled the CD&R problem in the chosen scenario. Figure 6 depicts four time-distributed boxplots, each involving the median, lower, and upper quartiles, as well as the lower and upper whiskers in one diagram. Each of the boxplot diagrams is associated with a work phase. This was calculated using the respective phases of 14 participants and indicated the cross-participant variance over the period of the trial. By comparing the variance across participants and time in this way, the systematic progression of the phases became clear, with periods alternating with each other and high- and low-amplitudes becoming apparent. This suggested that phases featured certain periods of dominance, indicating prevailing working patterns and related cognitive modes, as outlined by the CLC model.

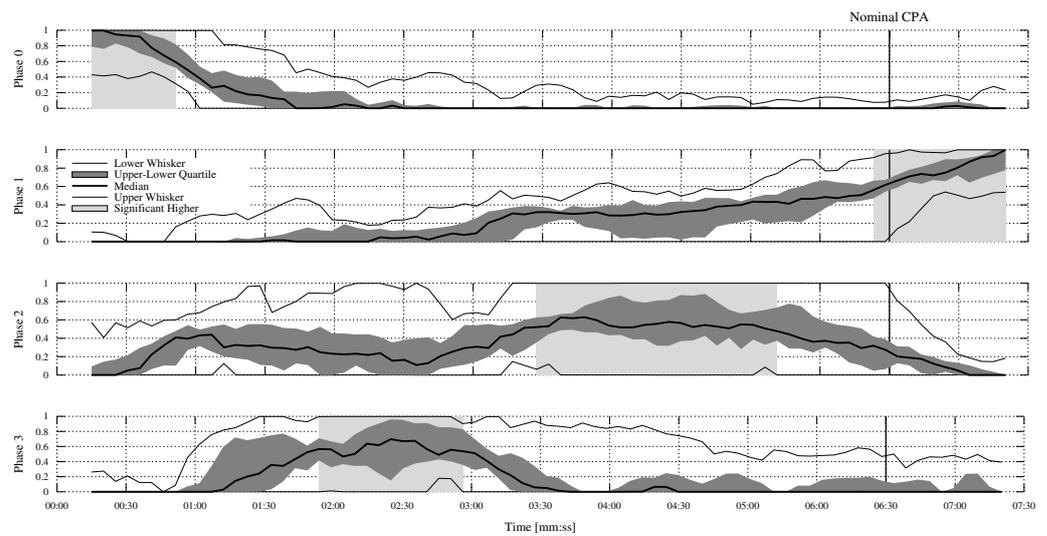


Figure 6. Time-distributed boxplot of phases for all participants.

For proving the order of dominance, we chose to use statistical tests that compared the variance of all phases (represented by the probabilities) using a non-parametric one-sided *Mann–Whitney U-test*. This tests the alternative hypotheses of whether a particular phase a is higher than another phase b ($H_1 : a > b$) using p -value ≤ 0.05 . Periods of statistical significance are highlighted by the light grey boxes in Figure 6. The temporal distribution of the significant periods followed the phase sequence 0-3-2-1 over the duration of the experiment.

5.3. Assessment of Phases’ Distinctiveness

We also investigated how well-“separated” (distinct) the phases uncovered by the model were. This question can be answered by associating a vector with each phase representing the per-phase event probability assignments $\langle w_0, w_1, \dots, w_m \rangle$, where $m > 0$ is the total number of events (features) used. The cosine similarity between the vectors representing two phases, f_1 and f_2 , was used to quantify the similarity between f_1 and f_2 , i.e., how distinct they were in terms of their events’ characterisation. Figure 7 depicts the results obtained in this study. It shows that Phases 2 and 3 were the most-similar (lowest distance), while Phase 0 was quite distinct from all other phases. We could say that Phase 3 shared characteristics with Phases 1 (e.g., look at the MTCD) and 2 (e.g., look at the sepToo1 tips), which is also reflected by the phases’ description shown in the heat map of Figure 5b.

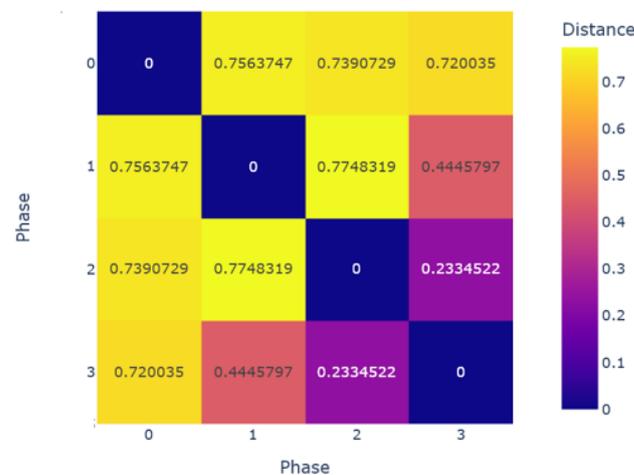


Figure 7. Cosine pairwise distance for the model’s 4 phases.

5.4. Phases' Validation Using the CLC Model

The degree of matching between the mapped information cues and the CLC work steps was calculated based on the assignments of the events per-phase. Applied to all phases, four vectors \vec{y} could be determined, which are shown in Table 4. This result is discussed in Section 6.

Table 4. Percentage matching between the mapped information cues and the CLC work step specification.

	Phase 0	Phase 1	Phase 2	Phase 3
Conflict Detection	11.9	09.9	06.1	06.4
Conflict Solution Probing	07.8	11.3	16.9	15.4
Solution Monitoring	01.6	18.3	44.8	40.0

5.5. Assessment of Similarities between Participants

Participants may differ in the progress of the phases over time. This is an indicator of differences in their work methods and the cognitive modes applied by the ATCO. In this context, an important question is how ATCOs' work strategies, as reflected by a model, can be compared. Answering this question gives a basis to find (dis)similarities between participants in terms of their work phases.

To answer this question, we looked at the time series of the 4D vectors, for each participant. Recall that each vector represents the phases' activation, on a sliding window, over a participant's event stream. Assume that the sequence of time windows W_0^P, W_1^P, \dots , for a participant P , are sorted in chronological order and that $\#W$ denotes the number of sliding windows over P 's event stream. Without loss of generality, we can assume that $\#W$ is equal across participants. Based on the representation of time windows as vectors, we propose the following parameterizable function to assess similarity between two participants P and P' :

$$\bigoplus_{i=0}^{\#W} (\vec{W}_i^P \approx \vec{W}_i^{P'}), \quad (1)$$

where \approx is a function to quantify the similarity between the 4D vectors associated with two time windows (W_i^P and $W_i^{P'}$, over the event streams of two participants P and P') and \bigoplus is an aggregation function over the pairwise similarities.

In our study, we experimented with the cosine similarity for the sliding windows' similarity function \approx and averaged as aggregation function \bigoplus .

To demonstrate the proposed concept, we look into the concrete question of how similar (or dissimilar) other participants were compared to participant P_{14} in Figure 5. Figure 8 depicts the answer to this question by ordering the participants based on their similarity to P_{14} . We can see that Participants 3, 4, 1, 11, and 7 were the most-similar to participant P_{14} , while Participant 12 seemed to be the most-dissimilar.

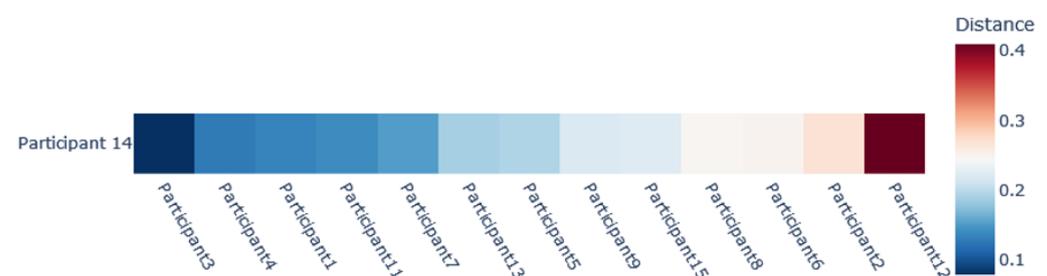


Figure 8. Heat map showing the similarity of P_{14} to all other participants. Distance was obtained as one minus the similarity.

Figure 9 shows clearly noticeable differences in the activation of the work phases for Participants 14 and 12. The initial Phase 0 was activated for almost double the time for Participant 12 compared to Participant 14. Moreover, during the period of time from

~200 ms to ~400 ms, Phase 2 was the dominant phase for Participant 12, while Phase 1 dominated during this period of time for Participant 14. Phase 3 was quite short for Participant 12 compared to Participant 14.

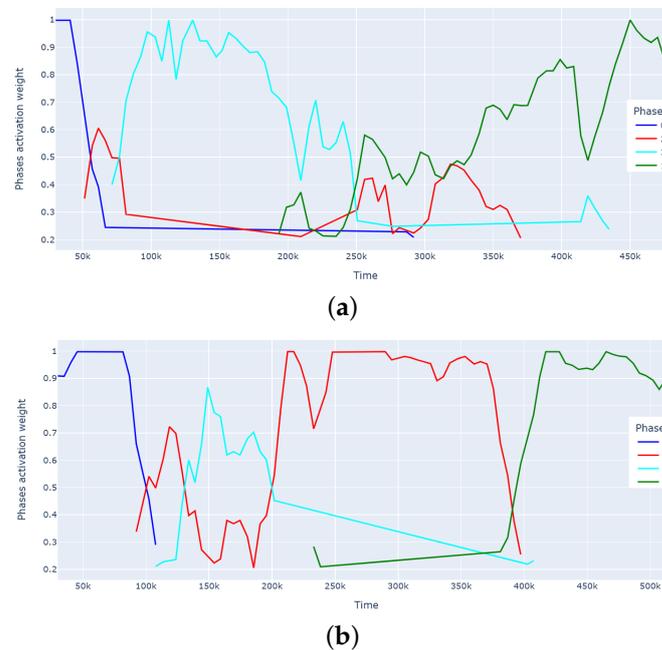


Figure 9. Phase activation graphs for two participants illustrating possible differences between individuals' strategies. (a) Phase activation graph for the participant of Figure 5. (b) Phase activation graph for Participant 12.

6. Discussion

Event streams, collected from 14 ATCOs, for the life cycle of the CD&R task were analysed in this work. We explored four phases in the CD&R tasks accomplished by the participants in our experiment. In addition, our method sought to identify the main features, i.e., events, related to each phase (see Figure 5b).

Though performance measures have been developed for topic models [36,37], we did not search for an optimal hyperparameter combination (e.g., via grid search) to generate the topic models, which could optimise some of these measures. The main reason was that, in this study, we focused on introducing our work as a novel ATC approach. Moreover, we had only access to data for 14 participants who performed a scenario of about eight minutes long once. As such, it was not feasible to divide the data into training, test, and validation datasets. Therefore, the results presented here should be seen as the first promising steps in our proposed approach, and further refinement is left as future work.

As one can see in the example in Figure 5b, the event patterns featuring the discovered work phases tended to be described in terms of look events. The main reason for this was that clearances and other types of commands given by the controller are infrequent when compared to look events, e.g., an ATCO may give four or five clearances, while she/he may have looked dozens of times at a waypoint. This drawback of our current method can be addressed by using a measure based on term frequency–inverse document frequency (tf-idf) when training a topic model, instead of the term frequency used now. The measure tf-idf captures rare words in a corpus and, consequently, facilitates the integration of less-frequent events, such as clearances, in the patterns. Moreover, boosting the possibility of including rare events in the patterns' description also makes it easier to investigate the correlation of commands with the phases, e.g., whether certain commands act as triggers for the end of one work phase and the start of another. For instance, we visually inspected the phases' activation graphs, for the 14 controllers in our study, in conjunction with commands to add (remove) `sepTool`, for aircraft in conflict (Figure 5a shows the graph for P_{14}). We

then concluded that adding sepTool marked the end of the dominance of Phase 0 for all participants, but three, while removing sepTool coincided with the start of the dominance of Phase 1 in all cases.

Though the model built extracted four work phases (i.e., parameter $k = 4$), the assessment of the phases' distinctiveness shown in Figure 7 pointed out that two of the phases, Phases 2 and 3, were rather similar. A possible explanation is that the "solution implementation" step of the CLC was mostly associated with non-look events (such as clearances) and, therefore, could not be detected by the built model.

Comparing Extracted Work Phases with the CLC Steps

The statistical analysis started with time-distributed boxplots, which visualised the variance across all 14 participants. Periods of significant dominance were identified that indicated the prevailing composition of events in the stream. The order of when phases experienced dominant time periods revealed the participant's approach to dividing a comprehensive task into a sequence of sub-steps with dedicated activities, evoking each significant event pattern. A possible explanation was provided by the four CLC work steps outlined in [11], which was used to determine which subtasks were obligatory for CD&R and were likely to be related to a change in the work pattern. A second approach was based on the mapping of probability vectors from the phase model to the CLC work step specification. The advantage was to find a relation of the phase model with the qualitative characteristics, as described in Section 4.3.

The fact that periods of dominance followed the phase sequence 0-3-2-1, as shown in Figure 6, provided a first indication of the associated work steps. Conflict identification was a step that the participant strove for at the beginning, indicating Phase 0. This was supported by the model, where Phase 0 relied specifically on events with the NAX and UAE labels (the conflict pair), as well as en-route symbols, which contain information about the flight level and heading toward the CPA. This is backed by the results shown in Table 4, assigning the highest probability to Phase 0 for *conflict detection*.

Phase 1 was difficult to assign to a work step because the period of dominance was situated after the conflict pair had passed the CPA, as indicated in Figure 6. At this point, the CLC did not provide sufficient information on appropriate work steps that adequately describe this particular period. The logic of the CLC would have the participant seek and detect the next conflict, which was not supported by the simple conflict scenario. From the phase model, it can be seen that the "label base" and the "track symbol" were characteristic look events that had a broad qualitative match through the *solution-monitoring* step, as indicated by Table 4. It appeared that Phase 1 indicated more-passive monitoring at the end of the scenario, including the post-conflict phase when movements exceeded the CPA, and the participant activity transitioned to a subsequent phase-out of the scenario with a final set of look events. Interestingly, the MTCD received considerable attention during this phase, which was confirmed also by the participants during the interviews [11]. The reason was that the conflict did not disappear from the MTCD window after it was resolved. This is a simulator glitch that attracted the participants' attention.

Phases 2 and 3 showed the highest degrees of activity in the period between conflict detection and the conflict pair passing the CPA. Following the sequence logic of the CLC work steps, these phases could include several work steps: *conflict solution probing*, *solution implementation*, and *solution monitoring*. From Figure 6, it seems that, besides the periods of dominance, Phases 2 and 3 featured a particular overlap in the time frame from 01:00 till 03:00. This overlap did not allow a clear assignment of the phases without further hints. The results shown in Figure 7 also indicated that Phases 2 and 3 were not so distinct from each other. A further look into the matching results (Table 4) revealed Phases 2 and 3 to have a predominant match with the *solution-monitoring* step. On the other side, *conflict solution probing* matched also in large part Phases 2 (16.9%) and 3 (15.4%), which emphasises the focus on the separation tool vector tip and the MTCD window.

A possible explanation for the ambiguity was provided by the CLC, where work steps *solution implementation* and *solution monitoring* may iterate several times within a short time period. This was to implement an initial solution plan based on a series of predefined time-distributed clearances (work step “resume solution”). Even corrective clearances along the way are conceivable, if separation margins turn out to be insufficient. This loop could have a considerably high switching frequency, which is probably too high for the selected time window 30 s to distinguish the change between these work steps.

In contrast, in Phase 2, clearances were usually given that focused on the period immediately preceding the CPA and showed corresponding visual activity on the part of the participant. Both Phases 1 and 2 showed a large overlap, not giving the same amount of separation sharpness as, e.g., Phases 0 and 3. A possible explanation is a highly individual variation of the transition time of switching, where some participants chose to switch over to a passive final scan pattern early, whereas others remained in a more active work step until the CPA and then switched. Allocating the phases with the corresponding work steps, the CLC suggested the order of appearance accordingly as 0-3-2-1.

To conclude this section, we refer to another possible validation method based on a temporal comparison of the distribution of subjective CLC work steps with the corresponding work phases. This presumes the availability of the CLC work steps mapped to “timelines”, which define periods of time when certain cognitive modes appear to be predominant. Such timelines represent ATCOs’ estimations when they believed they performed conflict detection, solution probing, solution implementation, or solution monitoring. These timelines of the CLC work steps would then provide a ground truth against which the results of the proposed method could be compared based on temporal distribution and consistency. The similarity function (1) proposed in Section 5.5 could be used to assess the similarity between the ground truth and the timeline of work phases inferred by the model for an ATCO. This evaluation procedure could be repeated in a leave-one-participant-out fashion. The retrospective think-aloud session supported this approach, as demonstrated in [11], by using the recordings of the simulation in playback mode, allowing the ATCO to identify specific steps and transitions between the steps of the CLC themselves and, thus, to produce these timelines.

7. Conclusions and Future Work

The work presented proposed a method based on topic models to reveal ATCOs’ work phases and their characteristic work patterns during CD&R tasks. The topic models were built from data in the form of event sequences, which were obtained by merging the eye movement data with the data from the simulation logs.

To identify and analyse the work patterns, our method puts focus on how ATCOs divide their attention across individual areas of interest while performing a task. In our case, attention to areas of interest corresponded to attention to the interface features of aircraft involved in a conflict, as well as to interactions with the interface objects (e.g., clicking on menus). By studying the interactions of the ATCOs with the simulator’s interface, we extracted information about the work actions being performed, while attending to objects relevant to a conflict.

Applying the proposed topic-modelling-based approach proved highly promising as it made it possible to extract work phases directly from the data with associated event patterns. These work phases can then be used as a basis for the comparison of the behaviour of individual ATCOs, to assess how similar their work strategies are. Aggregate group patterns can be compared using time-distributed boxplots, and individuals’ detailed patterns can be compared using a similarity function. Moreover, the proposed approach allowed for cross-comparisons to be made between subjective statements by ATCOs concerning their work patterns with the identified work patterns from the data. Finally, our approach made it also possible to determine what features were most important to be able to ascertain that a specific pattern is active.

To investigate the support for our proposed approach and its identified work phases, we related it to the work steps in the CLC model. We found correspondences between the phases and the CLC work steps. Here, Phase 0 best matched *conflict detection*, while Phase 1 best matched *solution monitoring*. From the previous discussion and data interpretation, we assumed that *solution monitoring* and *solution implementation* are strongly alternating activities covered by Phases 2 and 3. These processes cannot be sharply separated with the built model because of the high frequency of changes between them.

To conclude, this paper introduced a novel ATC approach for the identification of ATCOs' work phases. While the initial feedback received was encouraging, there are a number of interesting extensions that could be explored as future work. Firstly, upcoming extensions of this work should also adopt the validation method referred to at the end of Section 6. Secondly, this study focused on a single conflict being present on the display; future work should address situations with more conflicts. Thirdly, an interesting question to explore is the effect of interface adjustments that move information in the interface, which are associated with specific work phases, further apart. Further, in more applied studies, the sensitivity of this approach to expertise could also be explored, by evaluating, for example, how well it can show differences between experienced and competent controllers versus novices who require more training. Additionally, it would be interesting to use the method presented here (in particular, the possibility to assess work strategies' similarities) to study the effect of automation on the individual work approaches of controllers. Finally, we plan also to find a rational process to help the analyst select the number of phases in the mining activity.

Author Contributions: Conceptualization, A.N.; Methodology, A.N., L.M. and K.V.; Validation, L.M. and K.J.K.; Resources, J.L.; Data curation, A.N.; Writing—original draft, A.N., L.M. and K.V.; Writing—review & editing, K.J.K. and J.L.; Project administration, K.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swedish Transport Administration (Trafikverket) as part of the project 'Objective Verification of Competence wrt Advanced Automation', and by the Swedish Research Council as part of the project 'Progressive Interactive Event Sequence Analytics: Combining Explorative Mining and Visual Reasoning', grant number 2020-05000.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jun, L.Z.; Alam, S.; Dhief, I.; Schultz, M. Towards a greener Extended-Arrival Manager in air traffic control: A heuristic approach for dynamic speed control using machine-learned delay prediction model. *J. Air Transp. Manag.* **2022**, *103*, 102250. [[CrossRef](#)]
2. Degas, A.; Islam, M.R.; Hurter, C.; Barua, S.; Rahman, H.; Poudel, M.; Ruscio, D.; Ahmed, M.U.; Begum, S.; Rahman, M.A.; et al. A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory. *Appl. Sci.* **2022**, *12*, 1295. [[CrossRef](#)]
3. Meyer, L.; Carlsson, C.B.; Svensson, Å.; Peukert, M.; Danielson, L.; Josefsson, B. Stressing Safety Assessment Methods by Higher Levels of Automation. In Proceedings of the 33rd Congress of the International Council of the Aeronautical Sciences (ICAS2022), Stockholm, Sweden, 4–9 September 2022.
4. Pawlak, W.; Brinton, C.; Crouch, K.; Lancaster, K. A framework for the evaluation of air traffic control complexity. In Proceedings of the Guidance, Navigation, and Control Conference, San Diego, CA, USA, 29–31 July 1996. [[CrossRef](#)]
5. Inoue, S.; Furuta, K.; Nakata, K.; Kanno, T.; Aoyama, H.; Brown, M. Cognitive process modelling of controllers in Enroute air traffic control. *Ergonomics* **2012**, *55*, 450–464. [[CrossRef](#)] [[PubMed](#)]
6. Lundberg, J.; Svensson, Å.; Johansson, J.; Josefsson, B. Human-automation Collaboration Strategies. In *SESAR Innovation Days*; Schaefer, D., Ed.; University of Bologna: Bologna, Italy, 2015.
7. Palma Fraga, R.; Kang, Z.; Crutchfield, J.M.; Mandal, S. Visual Search and Conflict Mitigation Strategies Used by Expert en Route Air Traffic Controllers. *Aerospace* **2021**, *8*, 170. [[CrossRef](#)]
8. Regtuit, R.; Borst, C.; Van Kampen, E.J. Building Strategic Conformal Automation for Air Traffic Control Using Machine Learning. In Proceedings of the 2018 AIAA Information Systems-AIAA Infotech @ Aerospace, Kissimmee, FL, USA, 8–12 January 2018. [[CrossRef](#)]

9. van Rooijen, S.; Ellerbroek, J.; Borst, C.; van Kampen, E. Toward individual-sensitive automation for air traffic control using convolutional neural networks. *J. Air Transp.* **2020**, *28*, 105–113. [[CrossRef](#)]
10. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
11. Meyer, L.; Klang, K.J.; Boonsong, S.; Westin, C.; Nordman, A.; Lundberg, J.; Josefsson, B.; Vrotsou, K. Mapping the Decision-Making Process of Conflict Detection and Resolution in En-Route Control: An Eye-tracking based approach. In *12th SESAR Innovation Days*; SESAR JU: Brussels, Belgium, 2022.
12. Guo, Y.; Guo, S.; Jin, Z.; Kaul, S.; Gotz, D.; Cao, N. Survey on Visual Analysis of Event Sequence Data. *IEEE Trans. Vis. Comput. Graph.* **2022**, *28*, 5091–5112. [[CrossRef](#)] [[PubMed](#)]
13. Fournier-Viger, P.; Gan, W.; Wu, Y.; Nouioua, M.; Song, W.; Truong, T.; Duong, H. Pattern mining: Current challenges and opportunities. In *Proceedings of the Database Systems for Advanced Applications, DASFAA 2022 International Workshops: BDMS, BDQM, GDMA, IWBT, MAQTDS, and PMBD, Virtual Event, 11–14 April 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 34–49.
14. Perer, A.; Wang, F. Frequency: Interactive Mining and Visualization of Temporal Frequent Event Sequences. In *Proceedings of the International Conference on Intelligent User Interfaces, Haifa, Israel, 24–27 February 2014*; pp. 153–162.
15. Vrotsou, K.; Nordman, A. Exploratory Visual Sequence Mining Based on Pattern-Growth. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 2597–2610. [[CrossRef](#)] [[PubMed](#)]
16. Westin, C.; Vrotsou, K.; Nordman, A.; Lundberg, J.; Meyer, L. Visual Scan Patterns in Tower Control: Foundations for an Instructor Support Tool. *SESAR Innovation Days*. 2019. Available online: https://www.sesarju.eu/sites/default/files/documents/sid/2019/papers/SIDs_2019_paper_42.pdf (accessed on 10 June 2023).
17. Crutchfield, J.; Kang, Z.; Palma Fraga, R.; Lee, J. Identification of Expert Tower Controller Visual Scanning Patterns in Support of the Development of Automated Training Tools. In *Proceedings of the Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry: 14th International Conference, VAMR 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, 26 June–1 July 2022*; Chen, J.Y.C., Fragomeni, G., Eds.; Springer: Cham, Switzerland, 2022; pp. 183–195.
18. Rataj, J.; Ohneiser, O.; Marin, G.; Postaru, R. Attention: Target and actual-the controller focus. In *Proceedings of the 32nd Congress of the International Council of the Aeronautical Sciences, ICAS 2021, Shanghai, China, 6–10 September 2021*.
19. Bruder, C.; Hasse, C. What the eyes reveal: Investigating the detection of automation failures. *Appl. Ergon.* **2020**, *82*, 102967. [[CrossRef](#)] [[PubMed](#)]
20. Causse, M.; Lancelot, F.; Maillant, J.; Behrend, J.; Cousy, M.; Schneider, N. Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation. *Int. J. Hum.-Comput. Stud.* **2019**, *125*, 55–65. [[CrossRef](#)]
21. Lanini-Maggi, S.; Ruginski, I.T.; Shipley, T.F.; Hurter, C.; Duchowski, A.T.; Briesemeister, B.B.; Lee, J.; Fabrikant, S.I. Assessing how visual search entropy and engagement predict performance in a multiple-objects tracking air traffic control task. *Comput. Hum. Behav. Rep.* **2021**, *4*, 100127. [[CrossRef](#)]
22. Zraggen, E.; Drucker, S.M.; Fisher, D.; DeLine, R. (s| qu) eries: Visual regular expressions for querying and exploring event sequences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea, 18–23 April 2015*; pp. 2683–2692.
23. Cappers, B.C.; van Wijk, J.J. Exploring multivariate event sequences using rules, aggregations, and selections. *IEEE Trans. Vis. Comput. Graph.* **2017**, *24*, 532–541. [[CrossRef](#)] [[PubMed](#)]
24. Papadimitriou, C.H.; Raghavan, P.; Tamaki, H.; Vempala, S. Latent Semantic Indexing: A Probabilistic Analysis. *J. Comput. Syst. Sci.* **2000**, *61*, 217–235. [[CrossRef](#)]
25. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [[CrossRef](#)]
26. Liu, L.; Tang, L.; Dong, W.; Yao, S.; Zhou, W. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus* **2016**, *5*, 1–22. [[CrossRef](#)] [[PubMed](#)]
27. Kalmbach, A.; Hoeberechts, M.; Albu, A.B.; Glotin, H.; Paris, S.; Girdhar, Y. Learning deep-sea substrate types with visual topic models. In *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016*; pp. 1–9. [[CrossRef](#)]
28. Boyd-Graber, J.; Hu, Y.; Mimno, D. Applications of topic models. *Found. Trends[®] Inf. Retr.* **2017**, *11*, 143–296. [[CrossRef](#)]
29. Nguyen, P.H.; Henkin, R.; Chen, S.; Andrienko, N.; Andrienko, G.; Thonnard, O.; Turkay, C. VASABI: Hierarchical User Profiles for Interactive Visual User Behaviour Analytics. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 77–86. [[CrossRef](#)] [[PubMed](#)]
30. Kuhn, K.D. Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transp. Res. Part Emerg. Technol.* **2018**, *87*, 105–122. [[CrossRef](#)]
31. Pérez-Castán, J.A.; Pérez-Sanz, L.; Serrano-Mira, L.; Saéz-Hernando, F.J.; Rodríguez Gauxachs, I.; Gómez-Comendador, V.F. Design of an ATC Tool for Conflict Detection Based on Machine Learning Techniques. *Aerospace* **2022**, *9*, 67. [[CrossRef](#)]
32. Guleria, Y.; Tran, P.; Pham, D.T.; Alam, S.; Durand, N. A machine learning framework for predicting ATC conflict resolution strategies for conformal automation. In *Proceedings of the 11th SESAR Innovation Days, Virtual Event, 7–9 December 2021*.
33. Leedy, P.D.; Ormrod, J.E. *Practical Research Planning and Design*; Pearson Education: London, UK, 2019.
34. Li, W.C.; Kearney, P.; Braithwaite, G.; Lin, J.J.H. How much is too much on monitoring tasks? Visual scan patterns of single air traffic controller performing multiple remote tower operations. *Int. J. Ind. Ergon.* **2018**, *67*, 135–144. [[CrossRef](#)]

35. Miller, R.B. Response Time in Man-Computer Conversational Transactions. In Proceedings of the Fall Joint Computer Conference, Part I, New York, NY, USA, 9–11 December 1968; AFIPS '68 (Fall, part I), pp. 267–277. [[CrossRef](#)]
36. Stevens, K.; Kegelmeyer, P.; Andrzejewski, D.; Buttler, D. Exploring topic coherence over many models and many topics. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju-si, Republic of Korea, 12–14 July 2012; pp. 952–961.
37. Röder, M.; Both, A.; Hinneburg, A. Exploring the space of topic coherence measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015; pp. 399–408.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.