Classification of flight phases based on pilots' visual scanning strategies

Vsevolod Peysakhovich
ISAE-SUPAERO, Université de Toulouse
France
vsevolod.peysakhovich@isae-supaero.fr

Wietse D. Ledegang
Human Performance Department, TNO
Soesterberg, The Netherlands
wietse.ledegang@tno.nl

Mark M.J. Houben
Human Performance Department, TNO
Soesterberg, The Netherlands
mark.houben@tno.nl

Eric L. Groen
Human Performance Department, TNO
Soesterberg, The Netherlands
eric.groen@tno.nl

ABSTRACT

Eye movements analysis has great potential for understanding operator behaviour in many safety-critical domains, including aviation. In addition to traditional eye-tracking measures on pilots' visual behavior, it seems promising to incorporate machine learning approaches to classify pilots' visual scanning patterns. However, given the multitude of pattern measures, it is unclear which are better suited as predictors. In this study we analyzed the visual behaviour of eight pilots, flying different flight phases in a moving-base flight simulator. With this limited dataset we present a methodological approach to train linear Support Vector Machine models, using different combinations of the attention ratio and scanning pattern features. The results show that the overall accuracy to classify the pilots' visual behaviour in different flight phases, improves from 51.6% up to 64.1% when combining the attention ratio and instrument scanning sequence in the classification model.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning approaches**; • **Human-centered computing** → **Human computer interaction (HCI)**.

KEYWORDS

aviation, scan patterns, classification, machine learning, viewing behaviour

ACM Reference Format:
Vsevolod Peysakhovich, Wietse D. Ledegang, Mark M.J. Houben, and Eric L. Groen. 2022. Classification of flight phases based on pilots' visual scanning strategies. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 8–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3517031.3529641

ETRA '22, June 8–11, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9252-5/22/06…$15.00
https://doi.org/10.1145/3517031.3529641

1 INTRODUCTION

The visual behavior of pilots is traditionally described with measures such as dwell time or "mean glance duration" on an Area-of-Interest (AOI), where a glance is defined as maintaining visual gaze within an AOI which may be comprised of multiple fixations. Based on the mean glance duration, the AOI attention ratio represents the AOI visual demand in terms of the percent of total glances. These measures are computed by most eye-tracking software and provide a global picture of visual behavior. However, the mean glance duration and attention ratio do not account for the order in which AOIs are scanned, thus neglecting temporal information. The scan patterns provide additional information on the distribution of pilots' attention while pilots are trained to exhibit a continuous, repetitive, scanning across the instruments [Ledegang and Groen 2018; Lounis et al. 2020a,b, 2021]. To illustrate the complementary aspects of traditional and temporal measures, Figure 1 shows two examples of viewing behavior with equal dwell time distributions, but different scan patterns.

Comparing scan patterns is complex, especially for long sequences like a continuous flight [Haslbeck and Zhang 2017]. In such situations, it seems sufficient to look at the transition matrix, or probability of switching between two AOIs (Link-Value-Probability) as an indication of the main scanning strategy [Harris Sr et al. 1986; Haslbeck and Zhang 2017]. More advanced methods to analyze scan patterns have been described. For example, the so-called Markov chains extend the transition matrix with the two preceding dwells [Glaholt 2014; Harris Sr et al. 1986], which may provide better insights into the predominant patterns. However, comparing these matrices with each other seems difficult [Harris Sr et al. 1986]. Other measures for spatial and/or temporal randomness in the pilot's scanning include Nearest Neighbor Index [Clark and Evans 1954; Di Nocera et al. 2006; Moore and Gugerty 2010], entropy [Diaz-Piedra et al. 2019; van de Merwe et al. 2012], or Lempel-Ziv Complexity [Lounis et al. 2020b]. Another promising approach is N-gram sequences [Lounis et al. 2021; Reani et al. 2018] that compute the number of occurrences of AOI sequences of length N. The transition matrix is an example of sequences of size $N = 2$. Compared to other measures of scan patterns, the N-gram sequence approach is the most transparent and easiest to interpret. As 10% attention ratio is interpreted as "the pilot spend 10% of time looking at this instrument", 10% of "Outside→Speed→Altitude" sequence

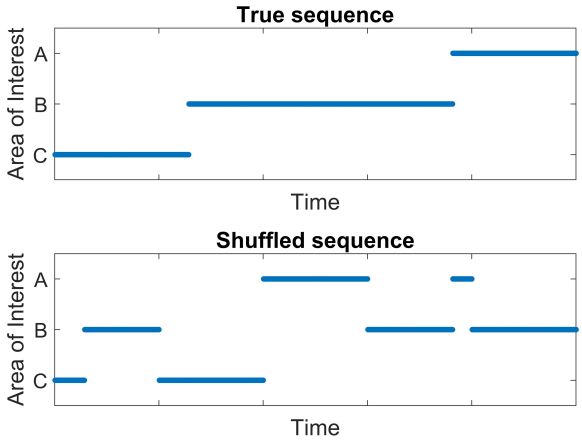

Figure 1: An example of different scan patterns with the same attention ratios (A = 25.8%, B = 50.6%, C = 23.7%). The top plot is a true simulated data, the bottom plot represents a shuffled sequence with the same attention ratio.

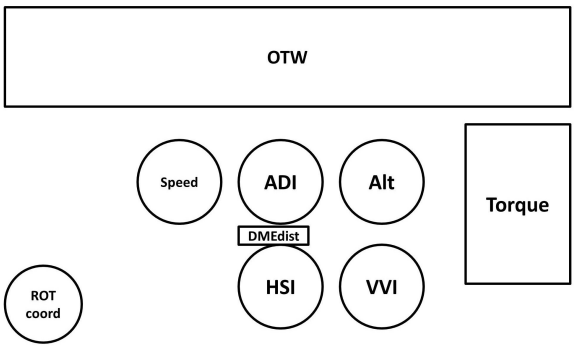

Figure 2: Schematic AOI layout. ADI = Attitude Director Indicator, Alt = Altimeter, HSI = Horizontal Situation Indicator, VVI = Vertical Velocity Indicator, DMEdist = DME distance information, ROTcoord = Rate-One-Turn coordinator, OTW = Out-The-Window.

occurrence is interpreted as "the pilot scanned these three instruments in that particular order 10% out of all performed transitions".

Eye movements analysis has great potential in understanding operator behavior in many critical domains including aviation [Peysakhovich et al. 2018]. Scan patterns of the experts can be used during training as an example of good instrument monitoring. It also can be used as objective data for observable behaviors within an evidence-based training approach. An important (and still unanswered) question is to what point eye movements are characteristic of different behaviors. For instance, if eye movements are recorded on a full flight simulator or an aircraft "black box", could we later understand whether the pilot scanning behavior was relevant to the situation or not? In this respect, machine learning approaches are promising where supervised algorithms can disentangle flight phase [Dehais et al. 2020; Scannella et al. 2018], expertise level [Castner et al. 2018; Lounis et al. 2021], or predict workload [Shojaeizadeh et al. 2019]. These supervised algorithms are only as efficient as the data we provide them. However, it is still unclear what are the best eye movements predictors (or features) given the numerous existing eye-tracking measures.

In this study, we investigated the problem of flight phase classification using eye movement recordings from pilots who performed flight circuits consisting of eight phases in a flight simulator. From this data, we selected the attention ratio (as most used measure) and AOI sequences as features to train classifiers with different feature combinations to compare their classification performance. In the following sections, we describe the setup and the flight scenario, the data processing, and eye movements' feature extraction, the machine learning approach, and classification results.

2 METHOD

2.1 Participants and apparatus

Ten Royal Netherlands Air Force (RNLAF) pilots participated in this study who had just finished their training for fixed-wing aircraft on a Pilatus PC-7. As this population did not include female pilots, all participating pilots were male, having an average ± standard deviation age of 24.8±2.2 years. On average they had 135±75 flight hours on powered aircraft, and 27±22 hours in a simulator. All pilots were proficient with a lookout and crosscheck when flying in Visual Flight Conditions (VFC), Instrument Meteorological Conditions (IMC) and Instrument Landing System (ILS) approach.

Before the experiment, all pilots signed informed consent, stating that the details of the experiment had been sufficiently explained and that they participated voluntarily. The experiment was conducted with the approval of the Institutional Review Board for research involving humans and was in accordance with the (revised) Helsinki Declaration.

2.2 Equipment

The experiment was performed in a Spatial Disorientation simulator (Airfox ASD, manufactured by AMST, Austria), featuring a hexapod motion platform. The out-the-window visuals were projected on a double-curved screen, at about 1.5 m from the eye-reference point, with a 2048×1536 pixels projector at 60 Hz (a field of view of 120° horizontally and 90° vertically). The generic cockpit featured three flat monitors for the instruments, a central stick with force feedback, pedals, throttles, flaps, and gear levers. The Pilatus PC-7 simulation showed some fidelity differences with the real aircraft: a larger cockpit and a shift of instruments; lack of stand-by instruments; torque indication in percentages rather than psi; localizer and glideslope indications on the Horizontal Situation Indicator (HSI) rather than on the Attitude Director Indicator (ADI); had no force feedback on the pedals, and required slightly larger pitch attitudes during flight. However, in consultation with a PC-7 instructor pilot, these aspects were considered adequate for this study.

For the off-line analysis of flight behavior, the control inputs, flight performance, and simulator motion were logged. The pilots' visual behavior was recorded with an Ergoneers Dikablis Professional head-mounted binocular eye-tracking system, featuring a 30 Hz wide-angle scene camera (1920×1080 pixels, full HD) and

two 60 Hz infrared eye-camera's (384×288 pixels). Two markers were installed on the instrument panel to relate gaze direction to world-fixed coordinates of the following nine areas-of-interest (AOIs, see Fig. 2): (a) Attitude Director Indicator (ADI), (b) Air Speed Indicator (Speed), (c) Altimeter (Alt), (d) Horizontal Situation Indicator (HSI), (e) Vertical Velocity Indicator (VVI), (f) Engine torque indicator (Torque), (g) DME distance information (DMEdist), (h) Rate-One-Turn coordinator (ROTcoord), and (i) Out-The-Window (OTW). Conform ISO 15007 standards, blinks (artefactual glance split due to a closed eyelid), and fly-throughs (artefactual fixation which is part of a saccade when the eye is moving between AOIs) were removed.

2.3 Procedure and flight scenario

The pilots performed a common manual circuit three times, each lasting about 15 min. Each circuit consisted of a go-around, visual search, level turn, orbit, straight-and-level, descending turn, ILS intercept, and ILS approach phase. The first circuit served to familiarize the pilots with the aircraft, simulator environment, and procedures and was excluded from the analysis. Following the second go-around, all pilots performed a third circuit which included Spatial Disorientation (SD) events to study the effect of SD on scanning behavior. The results of this "SD-circuit" have been reported separately [Ledegang and Groen 2018]. In the current study, we use the data from the the second "Normal-circuit" only (from first go-around to the second final approach), i.e., without SD events.

For each pilot, the experiment lasted approximately 1.5 hr, including an introduction, signing the informed consent, calibration of measurement equipment, and flight time. The gaze tracking data sets of two pilots were incomplete due to recording issues so the analysis of the pilots' viewing behavior is based on eight pilots.

2.4 Data processing and feature extraction

For each participant, the eye tracker software computed a list of dwells with indicated AOI name, start and end times, as well as dwell duration. We excluded from the analyses all dwells shorter than 100 ms. Based on these dwell times, the attention ratio (in percentage) was calculated per flight phase and per AOI, as shown in Table 1. For further analysis, we excluded the following three AOIs with less than 5% dwell times: DMEdist, ROTcoord, and VVI.

We call *pattern of size n* a sequence of *n* consequent dwells on different AOIs. For instance, *ADI→Alt* is an example of a pattern of size 2, and *ADI→HSI→Alt* is an example of a pattern of size 3. While using dwells as features for classification is straightforward, it is yet unclear what features to choose when considering patterns of different sizes. Thus, with the six considered AOIs, there are 30 possibilities for the patterns of size 2, and 180 possibilities for the patterns of size 3. While it is possible to compute the frequency for all theoretical patterns and use principal component analysis to identify relevant features, it is time-consuming and this approach would "hide" the relevant patterns, i.e. the component explaining most of the data variance may not make sense in terms of flight instruments. Thus, to use patterns for classification, we defined base patterns like the following. For each subject and each flight phase, we computed the three most frequent patterns of sizes 2 and 3. We considered patterns to be equivalent if one can be obtained by a cyclic rotation of the other. For instance, *ADI→Speed* and *Speed→ADI* or *ADI→Alt→HSI* and *Alt→HSI→ADI* are considered to be the same patterns. Out of the resulting 192 patterns of each size, there were eight unique patterns of size 2, and 22 unique patterns of size 3. It means that most top patterns per flight phase per subject are the same.

Next, for each subject and each flight phase we computed 36 features: six dwell time percentages, eight frequencies for patterns of size 2, and 22 frequencies for patterns of size 3. Given the length difference in each flight phase, the frequencies were normalized by number of dwells per flight phase. For instance, 30 repetitions of the same pattern during a flight phase that resulted in 150 dwells gives a value of 0.2 or 20%. Given the relative rarity of some patterns (especially of size 3) among all subjects and all flight phases, we decided to reduce the base patterns and considered only those where the median value (out of 64 values = eight subjects × eight flight phases) was positive, meaning at least half of data points were non-zero. It resulted in fifteen base patterns: six patterns of size 2, and nine patterns of size 3. All data processing was performed using custom-made scripts in MATLAB R2019b (Mathworks Inc., MA, USA). The resulting fifteen base patterns are highlighted among the 30 initially considered base patterns in the following list:

- **ADI→Alt**
- **ADI→Torque**
- **ADI→HSI**
- **ADI→OTW**
- **ADI→Speed**
- **Alt→HSI**
- Alt→OTW
- HSI→OTW
- **ADI→Alt→ADI**
- **ADI→Alt→Torque**
- **ADI→Alt→HSI**
- ADI→Alt→OTW
- ADI→Alt→Speed
- ADI→Torque→HSI
- **ADI→HSI→ADI**
- **ADI→HSI→Alt**
- ADI→HSI→Torque
- ADI→HSI→OTW
- ADI→HSI→Speed
- ADI→OTW→Alt
- **ADI→Speed→ADI**
- ADI→Speed→Alt
- **ADI→Speed→HSI**
- ADI→Speed→OTW
- **Alt→ADI→Alt**
- Alt→ADI→HSI
- HSI→ADI→Alt
- **HSI→ADI→HSI**
- HSI→ADI→Speed
- Speed→ADI→Speed

2.5 Machine learning approach

We used the Classification Learner app in MATLAB R2019b (Mathworks Inc., Natick, MA, USA) to train models to classify flight phases based on eye-tracking data. The input was a 64×22 table corresponding to 64 lines of observations (eight subjects × eight flight phases) and 22 variables (21 predictors, and 1 classification class label). As previously described, the predictors were six features of attention ratio on each of six AOIs, six features based on patterns of size 2, and nine features based on patterns of size 3. The classification class label corresponded to flight phase. We used 5-folds cross-validation and a linear Support Vector Machine (SVM) classifier with standard app options for this classifier (linear kernel function, box constraint level: 1, auto kernel scale mode, multiclass method: one-vs-one, and standardized data). The results of the training of 21 individual models are reported in Table 3 with true positive rates and overall accuracy for each individual model. Given the low accuracy rates of individual models, as a next step,

we trained the following models using different combinations of predictors:

- Model based on attention ratio's only
- Model based on patterns of size 2 only
- Model based on patterns of size 3 only
- Model based on patterns of size 3 without *ADI→HSI→Alt* and *ADI→Speed→HSI* patterns, that had individual overall accuracy less than 10% (see Table 3)
- Previous model with added patterns of size 2
- Previous model with added attention ratio's
- Previous model without patterns *ADI→Speed*, *ADI→Alt→Torque*, *ADI→HSI→ADI*, *ADI→Speed→ADI*, that had individual overall accuracy less than 20% (see Table 3)

Besides, the Classification Learner app allows training optimizable models that iteratively optimize hyperparameters. We trained five following Optimizable classifiers using the predictors from the last model from the previous list: Tree, Discriminant, Naive Bayes, KNN, and SVM. For these last models, we report accuracy and the resulted training time, and given the limited paper length, we do not report any other metrics such as precision, ROC, etc.

3 RESULTS

3.1 Attention ratio

Table 1 shows the attention ratio per area-of-interest and flight phase, averaged over eight pilots. Relative high attention ratios are found for the ADI, ranging between 31.2 and 58.2%. During the Visual search phase, the pilot's information need is mostly out-the-window, which is reflected by a relatively high OTW attention ratio (33.1%) as compared to all other flight phases (between 0.2 and 1.0%). During the ILS intercept and ILS approach the pilots continuously monitor the glideslope and localizer on the HSI, which is reflected in high HSI attention ratios during these flight phases (29.6 and 32.5% respectively). Finally, the torque indication is not given much attention (between 1.4 and 4.8%), except during the Descending turn (9.1%) and ILS approach (9.9%) phases when power needs to be controlled more precisely.

3.2 Patterns

Table 2 shows the pattern frequency per pattern per flight phase for 15 base patterns (6 of size 2, and 9 of size 3), averaged over eight pilots. As described in Sec. 2.4, the values are normalized. The flight phases of different lengths result in different dwell numbers, and, therefore, different numbers of possible patterns. To compensate for this aspect, the number of occurrences of each pattern was divided by the number of dwells for each flight phase and each participant.

3.3 Individual models

Table 3 shows the classification results of individual models. Each column corresponds to an individual feature — six based on attention ratios, six based on patterns of size 2, and nine based on patterns of size 3. Each column shows true positive rate for each flight phase and the overall accuracy. Given the eight subjects, 13% corresponds to one correctly classified pilot, 25% to two, 38% to three, 50% to four, 63% to five, 75% to six, and 88% to seven. The overall value gives the accuracy across all 64 points, for instance, model based on OTW attention ratio have overall accuracy of 18.8% because it classified correctly six out of eight visual search phases and six out of eight descending turns, i.e. 12 correct predictions out of 64. Note the low overall accuracy of individual models (highest score of 28.1% for pattern *ADI→HSI*).

3.4 Models with feature combinations

Table 4 shows the classification results of models with different feature combinations described in Sec. 2.5. As previously, each column corresponds to one given model with true positive rates for each flight phase and overall accuracy. Thus, the model using only six attention ratios features correctly predicted seven go-around phases, four visual search, four level turns, two orbits, one straight-and-level, five descending turns, four ILS intercept, and six ILS approaches. It results in 33 correct predictions out of 64, i.e. 51.6% overall accuracy. The first three models are based on same-type features, i.e. attention ratios only, patterns of size 2 only, and patterns of size 3 only. Given the relatively low overall accuracy for models using all base patterns, and the relatively low overall accuracy of some individual models, we decided to exclude some patterns from features based on their individual models' accuracy. Thus, first, we excluded patterns *ADI→HSI→Alt* and *ADI→Speed→HSI* with overall accuracy of the corresponding individual model less than 10% (see Table 3). The model of patterns of size 3 with these excluded low-score patterns improved the overall accuracy from 37.5% to 46.9% (fourth column). By further adding patterns of size 2 (fifth column), the accuracy increases to 50%, and by further adding attention ratios (sixth column), the overall accuracy increases up to 62.5%. By further excluding patterns with individual scores below 20% (i.e. *ADI→Speed, ADI→Alt→Torque, ADI→HSI→ADI, ADI→Speed→ADI*) from the last model, the overall model accuracy gained 64.1% but with a cost of some flight phases classification accuracy (for instance, straight-and-level true positive rate went from 63 to 25%).

3.5 Comparing classifier types

We used linear Support Vector Machines (SVM) classifier for our models. However, as described in Sec. 2.5, Classification Learner App allows optimizing some hyper-parameters (for example, maximum number of splits for a tree, or number of neighbors for k-nearest neighbors (KNN) classifier). Using the feature combination from the previous section (patterns with individual models scores above 20% and attention ratios), we trained five optimizable models. Tree model gave 48.4% accuracy and trained in 23.3 sec. Discriminant model gave 51.6% accuracy in 41.0 sec. Naive Bayes gave 53.1% accuracy in 111.7 sec. KNN gave 60.9% accuracy in 46.9 sec. Finally, SVM gave 8.8% accuracy in 126.8 sec. The optimized SVM had linear kernel (choice between Gaussian, linear, quadratic, and cubic), box constraint level of 0.40175 (in range of [0.001, 1000], standard value is 1), and multi-class method one-vs-all (standard value one-vs-one).

4 DISCUSSION AND CONCLUSION

In line with previous literature [Ledegang and Groen 2018], the attention was mainly shifted between the ADI, Alt, HSI, and Speed, with the ADI being the central node of attention, i.e. a typical

Table 1: Average attention ratio (in %) per flight phase and per area-of-interest. See Sec. 2.2 for instruments abbreviations.

	ADI	Alt	DMEdist	Torque	HSI	OTW	ROTcoord	Speed	VVI
Go-around	55.0	9.8	1.9	3.3	14.7	0.8	0.1	13.4	1.0
Visual search	31.2	9.6	3.1	4.8	10.9	33.1	0.0	5.7	1.4
Level turn	58.2	12.7	3.0	1.4	16.0	0.8	1.3	3.7	2.8
Orbit	53.2	14.8	2.5	2.5	16.7	1.0	1.2	5.1	3.1
Straight-and-level	48.8	14.8	3.4	5.0	19.9	0.9	0.2	4.6	2.4
Descending turn	52.4	11.8	2.4	9.1	14.9	0.2	0.4	8.4	0.5
ILS intercept	46.9	11.2	0.6	3.7	29.6	0.9	0.1	5.8	1.1
ILS approach	40.3	7.0	1.0	9.9	32.5	0.8	0.0	6.8	1.8

Table 2: Average normalized pattern frequency (in %) per flight phase and per pattern. See Sec. 2.2 for instruments abbreviations.

	ADI→Alt	ADI→Engines	ADI→HSI	ADI→OTW	ADI→Speed	Alt→HSI	ADI→Alt→ADI	ADI→Alt→Engines	ADI→Alt→HSI	ADI→HSI→ADI	ADI→HSI→Alt	ADI→Speed→ADI	ADI→Speed→HSI	Alt→ADI→Alt	HSI→ADI→HSI
Go-around	5.7	6.6	6.2	8.7	5.1	6.3	1.9	4.3	1.2	0.9	1.3	3.2	3.7	4.6	2.7
Visual search	4.7	4.7	3.3	4.6	4.1	3.7	3.8	4.3	2.8	4.3	3.3	4.8	8.1	4.0	3.6
Level turn	8.2	4.7	4.0	6.0	8.0	6.0	3.1	2.3	4.1	2.9	1.3	4.5	5.4	3.2	3.3
Orbit	4.7	3.0	5.6	5.6	5.8	4.4	4.6	3.5	4.9	3.3	5.3	5.5	5.7	3.8	3.5
Straight-and-level	7.1	4.8	4.7	7.2	5.8	5.2	4.5	4.6	3.0	1.9	1.5	7.0	8.1	6.7	4.8
Descending turn	5.2	3.5	7.8	6.5	7.6	7.3	5.4	6.9	2.5	2.9	3.3	3.1	4.4	3.1	3.1
ILS intercept	5.5	5.3	3.7	6.7	8.3	6.8	2.7	2.2	4.2	1.8	2.2	3.4	3.9	2.6	2.5
ILS approach	5.7	6.4	3.8	6.0	7.6	5.0	3.9	3.0	3.1	1.7	1.7	4.4	4.9	4.5	3.3

Table 3: True positive rates for each flight phase for individual models that use only one feature. The last row indicates the overall accuracy for the given model. Values above chance level are indicated in bold.

	AOI Attention ratio						*Patterns of size 2*							*Patterns of size 3*							
	ADI	Alt	Torque	HSI	OTW	Speed	ADI→Alt	ADI→Torque	ADI→HSI	ADI→OTW	ADI→Speed	Alt→HSI	ADI→Alt→ADI	ADI→Alt→Torque	ADI→Alt→HSI	ADI→HSI→ADI	ADI→HSI→Alt	ADI→Speed→ADI	ADI→Speed→HSI	Alt→ADI→Alt	HSI→ADI→HSI
Go-around	0	13	0	13	0	**63**	0	**25**	**25**	0	**75**	**88**	0	0	**88**	13	0	**63**	0	0	0
Visual search	**38**	**25**	0	**38**	**75**	0	0	13	**50**	**63**	13	0	0	0	0	**38**	0	13	13	0	**38**
Level turn	**50**	0	13	**25**	0	**38**	**25**	13	**38**	13	**25**	0	**25**	**38**	0	**25**	13	13	0	13	13
Orbit	0	13	**25**	0	0	0	**38**	**25**	**38**	**25**	13	**25**	**38**	**25**	13	13	13	13	13	**38**	**38**
Straight-and-level	13	13	0	0	0	13	13	**25**	13	0	13	13	**38**	0	0	13	13	13	0	**63**	**38**
Descending turn	**25**	13	13	13	**75**	**25**	13	13	0	**50**	0	13	0	**38**	0	0	0	0	13	13	0
ILS intercept	13	0	13	0	0	0	**25**	0	13	0	13	0	**38**	0	**50**	13	13	13	13	13	**25**
ILS approach	**50**	**63**	**38**	**50**	0	0	**50**	**63**	**50**	**25**	0	**25**	**38**	**25**	**38**	**38**	**25**	0	**25**	**38**	**50**
Overall	23.4	17.2	12.5	17.2	18.8	17.2	20.3	21.9	28.1	21.9	18.8	20.3	21.9	15.6	23.4	18.8	9.4	15.6	9.4	21.9	25

radial crosscheck strategy. In specific situations, such as looking out-the-window during the visual search phase, or when accurately changing the power settings during the descending turn and ILS approach, attention is shifted towards other areas of interest as well. This reflects the pilots' ability to apply a flexible crosscheck strategy to cope with a shift of information need and to adapt for (dynamic) deviations from normal flight. Furthermore, there are differences in the pilots' viewing behavior which we attribute to the specific information need per flight phase. For example, during the ILS intercept and approach, the pilots continuously monitor the glideslope and localizer that are indicated on the HSI, resulting in higher attention ratios for the HSI during these flight phases (29.6 and 32.5% respectively).

In this work, we investigated the characteristics and classification of pilots' eye movements in different flight phases. Although the dataset is rather limited, it serves well to establish the presented methodological approach. Further research with a larger dataset is, of course, required. We underlined that the traditional analysis

Table 4: True positive rates for each flight phase for models that use different combinations of features. The last row indicates the overall accuracy for the given model. Patt. = Patterns; * = This model excludes patterns *ADI→HSI→Alt* and *ADI→Speed→HSI*. ** = This model further excludes patterns *ADI→Speed, ADI→Alt→Torque, ADI→HSI→ADI, ADI→Speed→ADI*.

	Attention ratios	Patt. of size 2	Patt. of size 3	Reduced patt. of size 3*	Patt. of size 2 & reduced patt. of size 3	Patt. of size 2 & reduced patt. of size 3 & attention ratios	Reduced patt.** and attention ratios
Go-around	88	38	38	63	38	75	75
Visual search	50	63	38	50	38	50	75
Level turn	50	50	38	50	75	50	75
Orbit	25	63	50	38	63	50	38
Straight-and-level	13	13	25	50	38	63	25
Descending turn	63	38	0	0	25	50	63
ILS intercept	50	38	50	63	63	88	88
ILS approach	75	63	63	63	63	75	75
Overall	51.6	45.3	37.5	46.9	50	62.5	64.1

methods are not telling the whole story and taking into account the temporal nature of scanning patterns is important. To that end, we presented a methodology of computing base patterns — most encountered sequences (of sizes 2 and 3) of AOIs across pilots and flight phases. It allowed reducing the considered patterns from possible two hundred and fifteen combinations to only 15 that we used as model features. The resulting patterns make sense from the piloting point of view and reaffirm the ADI as the central node of pilots' monitoring strategy.

Individual models using one feature at a time showed that a single feature cannot suffice to separate different flight phases. Different features are predictive of different flight phases. For instance, OTW attention ratio allows to identify visual search and descending turn phases, while *Alt→HSI* and *ADI→Speed* patterns are indicative of go-around phase. Thus, we need to construct models with different combinations of features to separate correctly all flight phases.

The rightful question is, therefore, what features to consider and whether the standard attention ratios are enough. Models using only patterns of size 2 and 3 gave overall accuracy lower than the model based on attention ratios (45.3% and 37.5% compared to 51.6%, respectively). It shows that, attention ratios are a good tool to classify different flight phases, and patterns frequency alone isn't enough to correctly separate them.

However, we showed that, while using pattern frequencies alone is not sufficient to correctly discriminate the flight phases, by selecting the most characteristic patterns (based on individual models classification accuracy), we can improve the performance of the classifier and increase the overall accuracy up to 64.1%. Thus, combining AOI attention ratios and scanning sequences improves the overall accuracy of the classification model. It shows that the AOI order is important and considering different monitoring patterns is a relevant approach.

Comparing different classifier types showed that the chosen linear SVM model performs well at a cost of training time. Considering different classifiers can be interesting (such as KNN, for instance) when the data volume increases.

To conclude, in this study, we propose a novel methodology of selecting typical scan patterns ("base patterns") to improve the classification of different flight phases based on the pilot's distribution of visual attention. Adding pattern-based features allows increasing the overall classification accuracy of the model based on only attention ratios. It is a promising method of identifying characteristic monitoring strategies of different flight phases as the features are easy to interpret. As perspective, this approach can be used to get supportive data for the flight instructors to provide feedback to the ab initio students and reflect the way pilots are trained to scan their areas of interest.

ACKNOWLEDGMENTS

This research is supported by the Dutch Defence Research and Development Programme V1917 "5thGenStressors", and is a cooperation between TNO, the Centre for Man and Aviation (aeromedical institute of the Royal Netherlands Air Force) and ISAE-SUPAERO, Université de Toulouse, France. We thank J. Verhoeven for his contribution during the execution of the experiment, and M.C. Drost, M. Van Gaalen and J. Meesters for their support and operational knowledge in the preparation of the experiment. Finally, we thank all pilots for their enthusiastic participation.

REFERENCES

Nora Castner, Enkelejda Kasneci, Thomas Kübler, Katharina Scheiter, Juliane Richter, Thérése Eder, Fabian Hüttig, and Constanze Keutel. 2018. Scanpath comparison in medical image reading skills of dental students: distinguishing stages of expertise development. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 1–9.

Philip J Clark and Francis C Evans. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35, 4 (1954), 445–453.

Frédéric Dehais, Stéphane Juaneda, and Vsevolod Peysakhovich. 2020. Monitoring eye movements in real flight conditions for flight training purpose. (2020).

Francesco Di Nocera, Michela Terenzi, and Marco Camilli. 2006. Another look at scanpath: distance to nearest neighbour as a measure of mental workload. *Developments in human factors in transportation, design, and evaluation* (2006), 295–303.

Carolina Diaz-Piedra, Hector Rieiro, Alberto Cherino, Luis J Fuentes, Andres Catena, and Leandro L Di Stasi. 2019. The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. *Applied ergonomics* 77 (2019), 92–99.

Mackenzie G Glaholt. 2014. Eye tracking in the cockpit: a review of the relationships between eye movements and the aviators cognitive state. (2014).

Randall L Harris Sr, Bobby J Glover, and Amos A Spady Jr. 1986. Analytical techniques of pilot scanning behavior and their application. (1986).

Andreas Haslbeck and Bo Zhang. 2017. I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. *Applied ergonomics* 63 (2017), 62–71.

Wietse D Ledegang and Eric L Groen. 2018. Spatial disorientation influences on pilots' visual scanning and flight performance. *Aerospace medicine and human performance* 89, 10 (2018), 873–882.

Christophe Lounis, Almoctar Hassoumi, Olivier Lefrancois, Vsevolod Peysakhovich, and Mickaël Causse. 2020a. Detecting ambient/focal visual attention in professional airline pilots with a modified Coefficient K: a full flight simulator study. In *ACM Symposium on Eye Tracking Research and Applications*. 1–6.

Christophe Lounis, Vsevolod Peysakhovich, and Mickaël Causse. 2020b. Lempel-Ziv Complexity of dwell sequences: visual scanning pattern differences between novice and expert aircraft pilots. In *Eye-Tracking in Aviation. Proceedings of the 1st International Workshop (ETAVI 2020)*. ISAE-SUPAERO & ETH Zürich, 61–68.

Christophe Lounis, Vsevolod Peysakhovich, and Mickaël Causse. 2021. Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study. *PLoS one* 16, 2 (2021), e0247061.

Kristin Moore and Leo Gugerty. 2010. Development of a novel measure of situation awareness: The case for eye movement analysis. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 54. SAGE Publications Sage CA: Los Angeles, CA, 1650–1654.

Vsevolod Peysakhovich, Olivier Lefrançois, Frédéric Dehais, and Mickaël Causse. 2018. The neuroergonomics of aircraft cockpits: the four stages of eye-tracking integration to enhance flight safety. *Safety* 4, 1 (2018), 8.

Manuele Reani, Niels Peek, and Caroline Jay. 2018. An investigation of the effects of n-gram length in scanpath analysis for eye-tracking research. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 1–8.

Sébastien Scannella, Vsevolod Peysakhovich, Florian Ehrig, Evelyne Lepron, and Frédéric Dehais. 2018. Assessment of ocular and physiological metrics to discriminate flight phases in real light aircraft. *Human factors* 60, 7 (2018), 922–935.

Mina Shojaeizadeh, Soussan Djamasbi, Randy C Paffenroth, and Andrew C Trapp. 2019. Detecting task demand via an eye tracking machine learning system. *Decision Support Systems* 116 (2019), 91–101.

Koen van de Merwe, Henk van Dijk, and Rolf Zon. 2012. Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology* 22, 1 (2012), 78–95.