Cross Recurrence Analysis as a Measure of Pilots' Coordination Strategy

Patrick Gontar[1] and Jeffrey B. Mulligan[2]

[1] Institute of Ergonomics, Technical University of Munich, Munich, Germany

[2]Human Systems Integration Division, NASA Ames Research Center, Moffett Field, USA

Abstract. When solving problems, multi-person airline crews can choose whether to work together, or to address different aspects of a situation with a *divide and conquer* strategy. Knowing which of these strategies is most effective may help airlines develop better procedures and training. This paper concentrates on joint attention as a measure of crew coordination. We report results obtained by applying cross recurrence analysis to eye movement data from two-person crews, collected in a flight simulator experiment. The analysis shows that crews exhibit coordinated gaze roughly 1/6th of the time, with a tendency for the captain to lead the first officer's visual attention. The degree to which crews coordinate their gaze is not significantly correlated with performance ratings assigned by instructors; further research questions and approaches are discussed.

Keywords: Coordination, Joint Attention, Cross Recurrence Analysis, Eye-Tracking

Introduction

With technology becoming increasingly complex, many tasks require teamwork to be completed. Organizations that require high reliability therefore rely on effective teaming (Baker, Day, & Salas, 2006). While pilots in an aircraft can share their total workload by working as a team, they also have opportunities to monitor and cross-check each other. To work together efficiently, teams have to coordinate their tasks accordingly.

In order to train pilots to work efficiently as a team, it is useful to measure team coordination, and give the pilots explicit advice on how to improve it. To facilitate this kind of training, reliable and valid rating tools are critical. In order to use such measures in training, they should be rapidly available (within minutes), to be used as part of debriefing. Unfortunately, current technology does not support real-time analysis of communication data; therefore, human factors specialists have developed several rating tools so that flight instructors can evaluate pilots' behaviour by means of observable markers. However, even experienced raters sometimes show insufficient inter-rater reliability when assessing pilots' non-technical skills, such as teamwork or communication. Gontar and Hoermann (2015b) showed that familiarity with the rating tool and the performance level of the crew influenced the ratings. They concluded that the current practice of using subjective expert ratings could have a negative impact on the quality of feedback, and might even have economic repercussions. In this paper, we present a method for evaluating pilots' coordination, based on cross recurrence quantification analysis of eye gaze data.

Eye gaze data can help to understand a participant's cognitive processes by identifying the spatial locus of visual attention. Eye-tracking has been used to evaluate single pilots' mode awareness, adherence to procedures, and distribution of attention (e.g., Haslbeck, Schubert, Gontar, & Bengler, 2012). This study found that only about half of the required cross-checks on the *flight mode annunciator* were performed by the pilots. Gaze data from a single pilot can tell us about his/her allocation of attention, but reveal little about crew coordination and team-work (except perhaps in relation to other data). Dual eye-tracking, on the other hand, can reveal the overlap of information acquisition, and provide insight into a crew's coordination strategy. Richardson, Dale, and Tomlinson (2009) refer to gaze alignment as one part of gaze coordination: "… to highlight the fact that it [gaze coordination] appears to be causally connected to what people know going into a conversation …" (p. 1470). They furthermore state that gaze coordination can be seen as a measure of *joint activity* (Clark, 2007; Richardson et al., 2009). In the context of an airline cockpit, however, we are of the opinion that gaze coordination also exists when pilots follow a different coordination strategy according to which the pilots split their capacity and explicitly work on different sub-tasks to share the overall task-load (divergent activity). In this mode they would consequently show little gaze alignment, but we think that this strategy is also a subset of coordinated gaze.

Thus, we expect to encounter two distinct coordination strategies, where pilots either: (1) work on a task together and closely monitor each other (joint activity); or (2) split their capacity, and explicitly work on different sub-tasks, to share the overall task-load (divergent activity). Strategy 1 would follow the approach of Richardson et al. (2009), who found a causal relation between the degree of joint activity and participants' shared knowledge, whereas Strategy 2 takes task-load sharing into account. To distinguish these two strategies, we previously suggested using a chi-squared approach and calculated receiver-operator-characteristics based on synthetic eye gaze data (Mulligan & Gontar, 2016).

In this paper, we address the following four research questions regarding Strategy 1:

1. Is there evidence for coordinated gaze behaviour between pilots in the cockpit?
2. Does the joint activity strategy change with different tasks?
3. Can one of the pilots be identified as the leader?
4. How does performance correlate with the amount of joint activity?

Previous Work

A large body of work exists on the comparison of fixation sequences, often referred to as *scan-paths*. Many of these methods work with coded sequences of fixation regions, rather than raw position data. Recent overviews can be found in Anderson, Anderson, Kingstone, and Bischof (2014) and Le Meur and Baccino (2013). Scan-path comparison methods may be classified according to whether or not the temporal order of fixations matters. For example, the Levenshtein (*string edit*) distance counts the swaps, insertions and deletions necessary to transform one sequence into another (Hacisalihzade, Stark, & Allen, 1992). We do not necessarily expect that crew members engaged in team work will exhibit the same sequence of fixations, but we do expect to see similar groups of fixation targets within a time window appropriate to the activity. This was the intuition underlying our initial approach (Gontar & Mulligan, 2015) for an environment with N areas of interest (AOIs), for which we constructed the temporal sequence of N-dimensional vectors representing the distribution of gaze. Initially, we had one vector per eye-tracker sample, with 1 in the component representing the fixated AOI, and 0 in all other components. By averaging each of the components over time, the vectors could be transformed to represent the fraction of time spent fixating each AOI within a temporal window about a given time. The vectors corresponding to each crew member can be correlated, providing a normalized measure of the degree to which the same regions were fixated.

One of the weaknesses of this approach is that it only detects correlated behaviour that is shifted in time when the temporal averaging window is larger than the time lag. We expect to find correlation with delay in the data surrounding verbal communication: if one crew member notices something unusual in the instruments, and tells his/her partner about it, then the partner is likely to look at the instrument to confirm what he/she has just been told. This shortcoming might be overcome by computing the temporal cross-correlation. In this paper, we explore an alternate method that similarly detects recurrent states with temporal lags, known as recurrence analysis.

Recurrence Analysis

Recurrence plots were first introduced by Eckmann, Kamphorst, and Ruelle (1987), as an aid for visualizing the behaviour of dynamical systems; they have since found applications in numerous disciplines, including climate research (e.g., Proulx, Parrott, Fahrig, & Currie, 2015) and human-machine (e.g., Lehsing, Fleischer, & Bengler, 2016) or human-human interaction (e.g., Gontar, Fischer, & Bengler, 2016; Shockley & Riley, 2015). Following Marwan and Webber (2015), the recurrence function for a categorical signal $x(t)$ can be defined as:

$$R\ (t_1, t_2) = \begin{cases} 1, & x(t_1) = x(t_2) \\ 0, & otherwise \end{cases}, \qquad t_1, t_2 = 1 \dots N,$$

where N represents the number of distinct temporal samples. The quantity $x(t)$ is the integer index of the AOI fixated at time t. In the case of a continuously varying function, it would be necessary to include a threshold parameter ε, to describe the distance between $x(t_1)$ and $x(t_2)$ that is regarded as sufficiently close.

This results in the recurrence plot, which is an $N \times N$ array, such that a black dot is plotted at (t_1, t_2), whenever $x(t_1)$ is equal to $x(t_2)$. To calculate these recurrences, we used the CRP toolbox for MATLAB® provided by Marwan, Carmen Romano, Thiel, and Kurths (2007). Figure 1 shows the recurrence plot with 25 sec. of the gaze behaviour of a senior captain manually flying an aircraft on final approach.

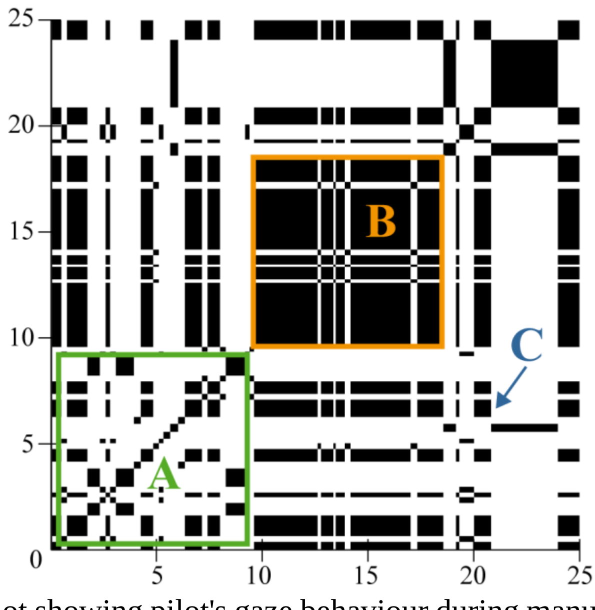

Figure 1: Recurrence plot showing pilot's gaze behaviour during manual flight on an approach.

In phase A, we can see that there are only short scattered recurrences, whereas by contrast, in phase B, the pilot focuses on fewer areas, with rather long fixations. The black dots further away from the diagonal (e.g., C) show that the pilot fixated the same areas as before. In case of C, this means (s)he fixated the same area at t = [6 7] as at t = [20 21]; however, it should be noted that recurrence plots do not directly tell us which areas were fixated nor how many areas were involved.

It can be seen that the recurrence plot is (by definition) symmetric with respect to the main diagonal. This approach allows us to analyse a single pilot's change in visual behaviour over time but must be extended to compare two pilots. Marwan and Kurths (2002) introduced the *cross recurrence plot* to analyse the dependencies between two dynamic systems. The cross recurrence function can be defined as follows:

$$C\,(t_1, t_2) = \begin{cases} 1, & x(t_1) = y(t_2) \\ 0, & otherwise \end{cases}, \qquad t_1, t_2 = 1 \dots N,$$

where again N represents the number of temporal samples. Analogously to the recurrence plot, the cross recurrence plot then marks each recurrent state with a black dot at (t_1, t_2). Figure 2 shows a generic cross recurrence plot on the left and real data of two pilots' gaze behaviour on the right. Black dots on the diagonal thereby represent recurrences that happen at exactly the same time in both time series; dots below the diagonal mark recurrences that appear later in the time of series x than on series y; dots above the diagonal indicate the reverse relationship.

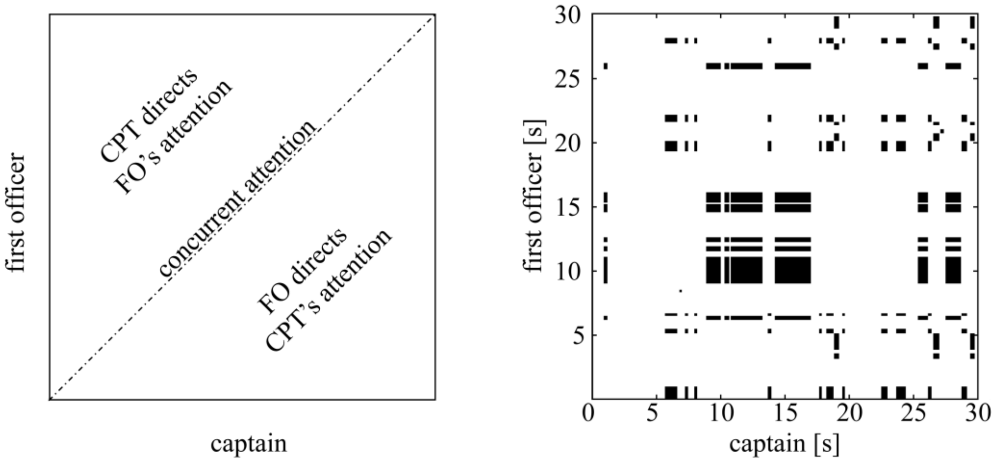

Figure 2. The plot on the left side shows a generic example of a cross recurrence plot comprising a crew's gaze behaviour. The plot on the right shows the relation between the gaze behaviour of two crewmembers.

In terms of gaze data, we can conclude that recurrence above the main diagonal belong to a fixation at an AOI that was visited first by the captain and afterwards by the first officer (see Figure 2 left). The reverse is true for points below the main diagonal. We can assess the degree to which one crew member directs the attention of the other by examining the balance of recurrences above and below the main diagonal. When looking at a randomly selected piece of real data (see Figure 2 right), we see that in the observed time frame, recurrences can be found at several times. The biggest cluster of recurrences is between t_{CPT} = [9 17] corresponding to the behaviour of the first officer at nearly the same time t_{FO} = [9 16].

While we can draw qualitative conclusions by visually inspecting the cross recurrence plots, we desire quantitative metrics that can be tested for statistical significance. A number of metrics have been proposed for the quantitative analysis of recurrence plots (Marwan et al., 2007; Marwan & Kurths, 2002; Marwan & Webber, 2015; Zbilut & Webber, 1992). Here we concentrate on two: the *recurrence rate* (RR), and the *lag-dependent recurrence rate*, or τ-recurrence.

The *recurrence rate* is the ratio of the number of observed recurrences, normalized by the maximum possible number of recurrences. It is defined as follows:

$$RR = \frac{1}{N^2} \sum_{t_1,t_2=1}^{N} C(t_1, t_2)$$

If most of the activity is joint, the RR will be high, with many recurrences near the main diagonal. Recurrences far from the main diagonal, on the other hand, are likely to result from coincidental overlap rather than coordinated activity. Therefore, it is useful to calculate the RR as a function of the time lag τ. This lag-dependent recurrence ratio can be defined as:

$$RR_\tau = \begin{cases} \sum_{t=1}^{N-\tau} C(t, t+\tau) \ , & \tau \geq 0 \\ \sum_{t=|\tau|}^{N} C(t, t+\tau) \, , & \tau < 0 \end{cases}$$

In the context of human interaction, this allows us to analyse the direction of interaction. In the case where the maximum of RR_τ lies at negative lag, it could be hypothesized that actor y follows the leadership of x. In the context of an aircraft cockpit, we would expect activities pertaining to the *aviate* task to be led by the *pilot flying* rather than by the *pilot monitoring*. In general, the captain may be expected to show the leading behaviour as he/she is likely more experienced and bears the ultimate responsibility for the aircraft and the passengers.

We are not simply interested in the degree of teamwork averaged over the entire flight scenario (which would result in one large cross recurrence plot), but also the switching between team and individual activities. Therefore, we compute the metrics using a sliding temporal window along the diagonal. This procedure is characterized by two parameters, the duration of the window (in seconds), and the step size by which the window is moved between calculations (which has a lower bound of 40 milliseconds, the sample period of the eye tracker).

To calculate confidence intervals, a bootstrapping approach has been suggested by several authors (e.g., Marwan, Schinkel, & Kurths, 2013). Bootstrapping refers to a class of methods for estimating an empirical distribution when a theoretical distribution is not known (Efron & Tibshirani, 1993). In the current context, we can obtain a noise distribution by randomly picking a pair of intervals, one from the pool of captains' data, and another from the pool of first officers' data. We do this rather than shuffling the data in order to maintain the underlying structure of the data set, so that we do not overestimate the significance of our real data. If the observed data distribution falls sufficiently far outside of the bootstrap distribution, we can conclude that there is evidence for statistically significant gaze coordination. The calculation of the recurrence rate, combined with the bootstrap approach, thus enables us to quantify the pilots' joint activity. To the best of our knowledge, this study is the first application of cross recurrence quantification analysis to pilot attention strategies in cockpit environments.

Method

Participants

We used eye-tracking data extracted from an experiment that was conducted in 2014 and was extensively reported by Gontar and Hoermann (2015b), Gontar, Porstner, Hoermann, and Bengler (2015), and Haslbeck and Gontar (2014). Sixty different two-person crews, made up from 120 current airline pilots executed a flight simulator mission. The participants were randomly selected from the roster of a partner airline. Participation in the experiment was not on a self-selection basis in order to avoid any bias (Rosenthal & Rosnow, 1975). The participants age was M = 39.9 years (SD = 8.65 years), and the pilots had M = 9249.36 hours (SD = 5139.8 hours) of flight experience.

Apparatus & Experimental Procedure

The flight simulator experiment took place in a certified level D (high-fidelity) flight simulator, over the course of 20 days at the flight training facility of the cooperating airline. Half of the crews flew an Airbus A320, while the other half flew an Airbus A340. Participants were informed before the experiment that they would be flying an approach and landing at the respective airport. The A340 crews flew approaches to New York City's John F. Kennedy airport (JFK), while the A320 crews flew approaches to Nice, France (NCE).

When the pilots entered the simulator, the head based eye-tracking system Dikablis® was attached and was subsequently calibrated using four points, all lying in the plane of the respective instrument panel. Markers were attached in the cockpit to provide a spatial reference for the AOIs later on. The markers did not obscure any instruments; Figure 3 shows the view of a first officer with the AOIs marked.

Figure 3. Cockpit of the full flight simulator (JAR STD 1A Level D) used in the experiment, with two markers and the defined AOIs (1 = Flight Control Unit; 2 = Engine Warnings and System Display; 3 = Secondary Engine Display; 4 = Multifunctional Control Display Unit; 5 = Navigation Display; 6 = Flight Mode Annunciator; 7 = Speed Indicator; 8 = Attitude Indicator; 9 = Altitude Indicator; 10 = Heading Indicator; 11 = Outside view).

In the flight scenario, the pilots experienced two malfunctions, caused by the same underlying problem: a leak of the green hydraulic system. On final approach, we induced the malfunction so that the landing gear was not able to fully extend (failure 1). As a consequence, the pilots had to perform a go-around, and complete several procedures before flying a second approach. When they then attempted to extend the flaps, the same malfunction prevented the flaps from moving to the correct position (failure 2). The flight scenario can thus be divided into four major phases, as shown in Figure 4.

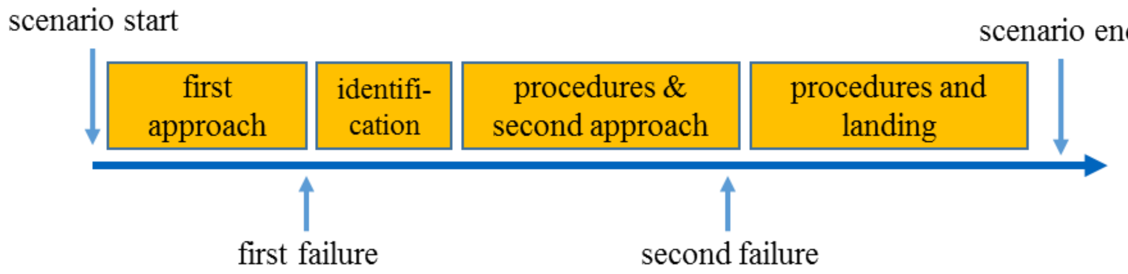

Figure 4: First rough definition of flight phases.

Dependent measures

Joint activity was measured by calculating the recurrence rate of the pilots' gaze behaviour over time using a moving window approach with different window sizes, from 10 sec to 30 sec. We calculated the mean recurrence rate as well as the percent time that the recurrence rate exceeded the 95th percentile of the random distribution. In order to judge whether the observed recurrence rate was significantly different from what would be expected from chance, we used a bootstrap approach as described above. Both the average recurrence rate and the lag-dependent recurrence rate were calculated for the four different phases of flight we preliminarily defined.

To identify the leading pilot, we constructed an average cross recurrence plot for the entire flight and determined the recurrence rate diagonal-wise as a function of the distance to the main diagonal (lag τ). In our analysis, a negative lag indicated that the captain was leading, while a positive lag indicated a leading first officer (see Figure 2).

Performance was assessed by recently retired instructors using the Line Operations Safety Audit (LOSA) *Approach and Landing Sheet* (Klinect, Murray, Merritt, & Helmreich, 2003). The rating scale goes from 4 (outstanding), to 1 (poor). Assessments were made for flight phases 1-3 as a group, and separately for flight phase 4. Inter-rater reliability was determined by calculating intra-class correlation coefficients, using a two way random model *ICC(2)*. This had an average value of .39 for the two instructors who served as raters in this experiment showing only poor reliability on this scale (Gontar & Hoermann, 2015a).

Analysis

The eye-tracking data was filtered to only include pupil detection rates greater than 80% and accurate calibration. Data sets with too low marker recognition or too low pupil recognition (e.g. glasses, reflections on the pupil, or mascara) were excluded from the present analyses. After excluding the lower quality data, the sample size shrank to n = 26 crews (out of n = 60 participating crews), for which the quality of the eye-tracking data was sufficient for both pilots.

For this analysis, we used 11 AOIs relevant to the tasks the pilots had to handle. AOIs were then drawn using the software D-Lab 2.0® (see Figure 3). Blinks and cross-throughs (artifactual visits to an AOI caused when a saccade trajectory passed over it) were eliminated by applying a fixation duration threshold of 120 ms. The input data for the recurrence analysis was a time series of fixated AOIs for each pilot, sampled at the eye-tracker rate of 25 Hz. Each AOI was represented by its integer index.

To analyse the recurrence based metrics introduced previously, we used the CPR toolbox provided by Marwan et al. (2007). To reduce calculation time, we ran 32 instances of Matlab® on 10 separate machines. Sliding windows with sizes of 250, 500, and 750 frames (corresponding to 10, 20, and 30 seconds of simulation time) were applied, with a step size of 25 frames (1 second).

To estimate the significance of the experimental data, we bootstrapped the data using a Monte Carlo approach (n = 1,000,000 samples) to obtain a random distribution. Based on the bootstrapped random distribution, we calculated the 5th, 50th, and 95th percentiles, and used them to compare the experimental data.

Results

The crews' gaze behaviour was significantly coordinated (α-level = .05) about 17% of the time. The mean values for the recurrence rate as a function of the window size, which are shown in Table 1, were calculated by an analysis that applied a sliding window approach to the data set of 26 crews (scenario duration was ~ 30 minutes).

Table 1. Descriptive analysis of the recurrence rate (RR).

window size [frames (sec.)]	RR mean, max, min	Bootstrapping 5th, 50th, 95th percentile	% time RR higher than 95th percentile Mean
250 (10)	.11, 1, 0	.00; .02; .20	16.67 %
500 (20)	.10, .95, 0	.00; .03; .17	17.74 %
750 (30)	.10, .89, 0	.00; .03; .16	17.68 %

The step size between averaging windows was 25 frames (1 second) in each analysis; bootstrapping resulted in the percentiles displayed in Table 1. We did not see large differences between the random distributions using different window sizes. We found that the pilots' gaze behaviour was significantly coordinated about 17% of the time (observed recurrence rate larger than the 95th percentile of the bootstrap distribution).

The recurrence rate changes with task and window size. To develop a better understanding of the results, we will present all outcomes by illustratively visualizing three representative crews. Figure 5 shows the recurrence rate as it progresses over time for three example crews, one with low (LOW) recurrence on the top, one with medium (MEDIUM) recurrence in the middle, and one with high (HIGH) recurrence on the bottom.

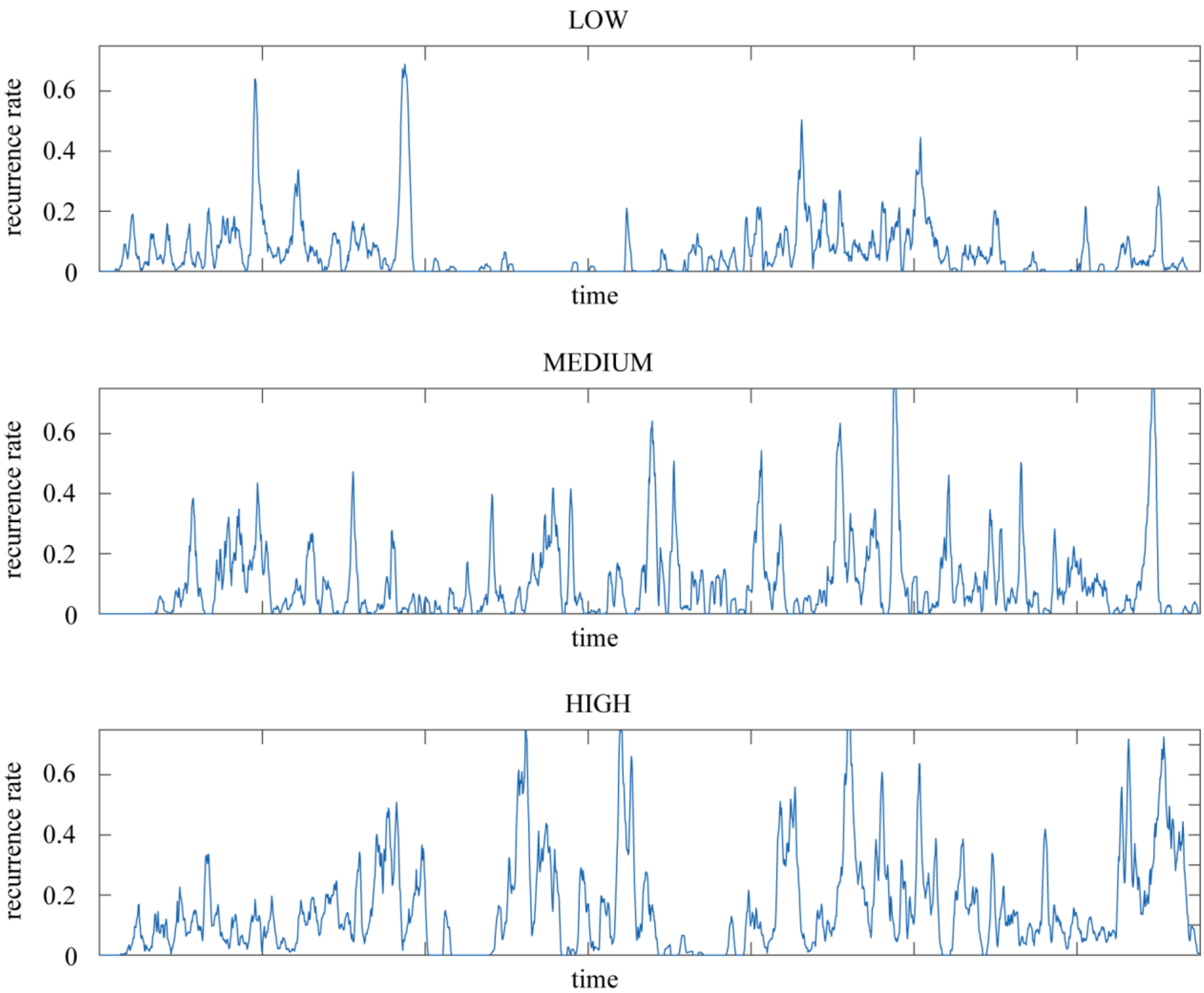

Figure 5. The recurrence rate changes over time as different tasks have to be managed. The top figure shows a crew with low recurrence, the middle figure shows medium recurrence and the bottom figure shows a crew with high recurrence. The size of the sliding window was 10 seconds; the shift between each calculation step was set to 1 second.

The results show that the pilots' joint visual attention did differ between crews, as measured by the recurrence rate, even though they were all flying the same task/scenario. A two-way ANOVA was run, taking the *flight phase* and *window size* as two repeated factors and mean *recurrence rate* as the dependent measure (see Figure 6). Mauchly's tests indicated a violation of sphericity on both factors, so we report the Greenhouse-Geisser corrected degrees of freedom.

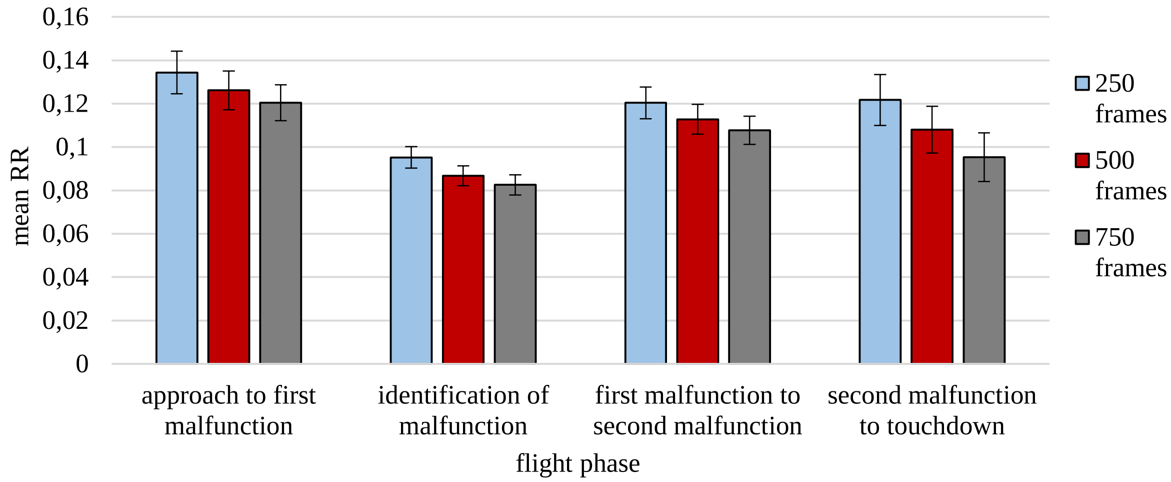

Figure 6: Recurrence rate as a function of the four flight phases and sliding window size. Error bars refer to the standard error.

The results show that the *flight phase* factor had a main effect, $F(2.18,34.88) = 3.40$, $p = .025$, on the dependent variable. The second factor, the *window size*, also had a main effect on the recurrence rate, $F(1.04,16.67) = 92.01$, $p < .001$. Pairwise comparison shows a significant difference between the mean values of the first ($M = .13$, $SD = .04$) and the second phases of flight ($M = .09$, $SD = .02$), $p = .009$.

We expect that some recurrences will occur even when the pilots are not engaged in co-ordinated activity; we estimated this by computing the bootstrap distribution as described above. Figure 7 shows the mean percentage of time in each flight phase where the recurrence rate exceeds the 95th percentile of the corresponding random distribution.

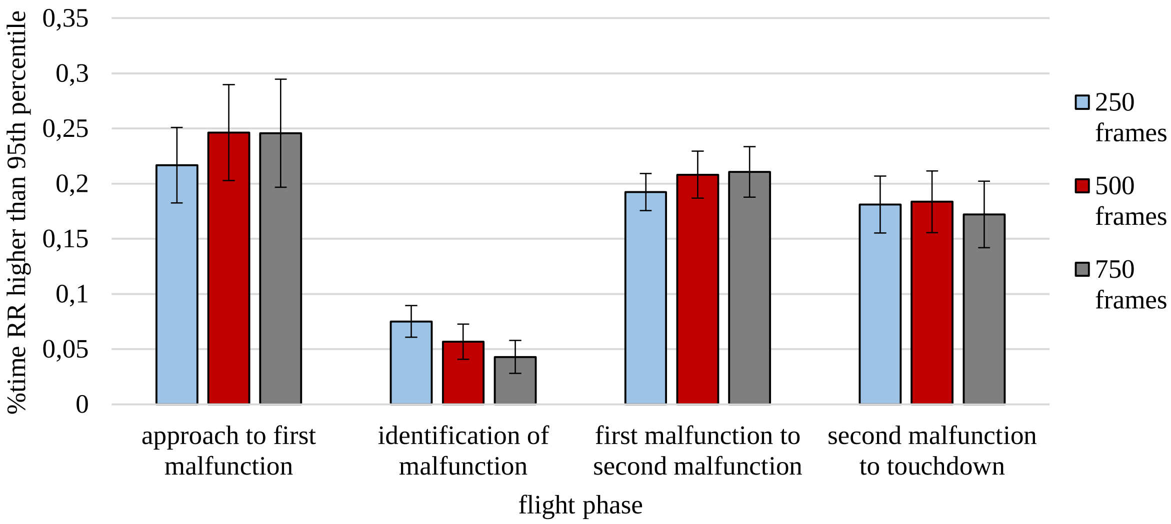

Figure 7. Percentage of time that recurrence rate is higher than the 95th percentile of the random distribution, as a function of the flight phase and the window size. Error bars refer to the standard error.

Again, a two-way ANOVA was run with the *flight phase* and *window size* as two repeated factors and *percentage of time exceeding the 95th percentile* of the random distribution as the dependent measure. It shows that the flight phase again had a main effect on the recurrence rate, $F(1.86, 42.82) = 7.99$, $p < .001$, but that the window size did not, $F(1.41,32.43) = 1.81$, $p = 175$. Pair-wise comparison shows that flight phase 2 *identification of malfunction* had significantly lower mean recurrence rates than all the other phases, $p = .002$ (phase 1), $p < .000$ (phase 3), $p = .005$ (phase 4).

The captain leads the first officer. The analysis of the recurrence rate as a function of the lag τ (distance between the point of interest and the main diagonal) is illustratively shown in Figure 8 for the same crews as in Figure 5, showing low, medium and high average recurrence rates. We applied a moving window with a bin size of 250 frames (10 sec) and a step size of 25 frames (1 sec). We then averaged the single cross recurrence plots as shown below. The colour represents the recurrence rate, where red is the highest and blue the lowest recurrence rate of the respective crew. Note that the scale of the plots differ between crews.

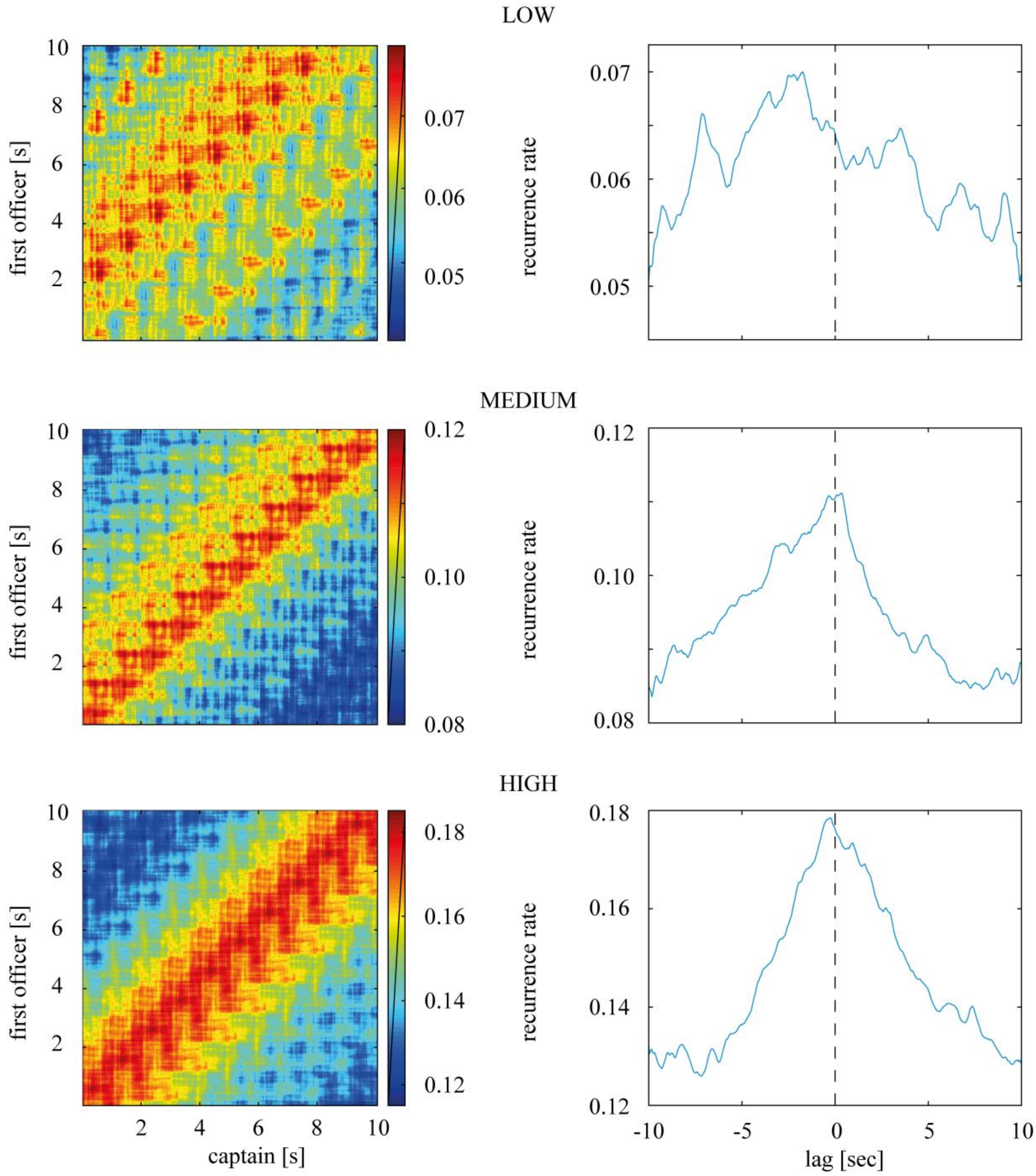

Figure 8. Average cross recurrence plots in the left column; corresponding lag diagrams in the right column. From top to bottom: one crew with low (LOW), one crew with medium (MEDIUM), and one crew with high (HIGH) overall recurrence rate.

From the cross recurrence plots, we can find the location of the maxima in relation to the main diagonal. The x-axis represents the timeline of the captain, and the y-axis gives the timeline of the first officer. By analysing the lag recurrence of these three crews, we can see what we already expected from the cross recurrence plots (Figure 8 left). The two crews with low and medium overall recurrence rates show more recurrences above the main diagonal, indicating leading behaviour by the captain. The crew with high overall recurrence, however, shows a more symmetrical distribution of recurrences (Figure 8 bottom). The average cross recurrence plot for the entire scenario and all crews is shown in Figure 9 on the left side; the corresponding lag-recurrence on the right side.

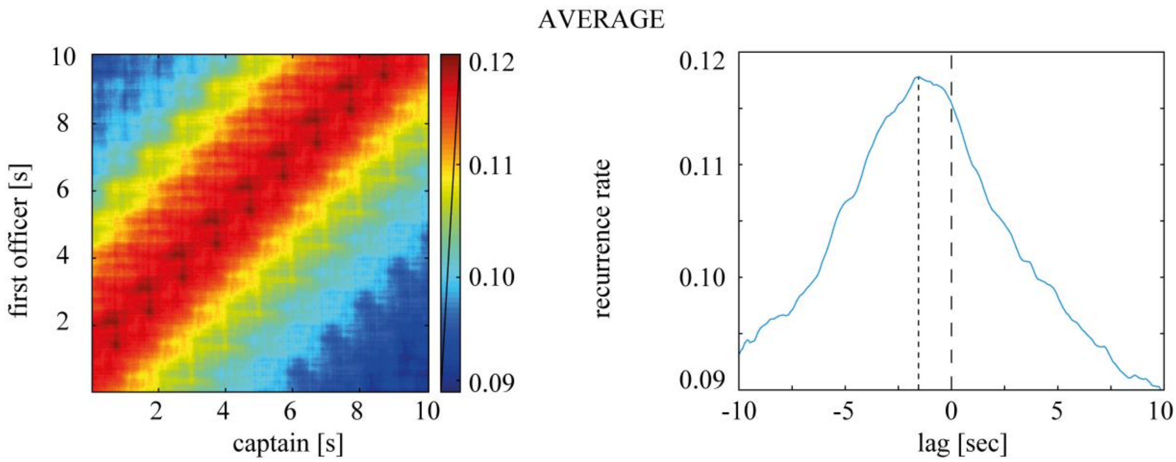

Figure 9. Left: average cross recurrence plot for all crews; Right: corresponding lag-dependent recurrence rates.

These overall estimates show a medium recurrence rate of M = .107, SD = .001 as well as a slight shift towards a negative lag. The maximum of τ-recurrences lays at τ = -1.52 sec. This leads to the conclusion that on average, the captain is leading the first officer, who is following his gaze; however, statistical tests have to be performed in order to back up these results.

The recurrence rate and performance data show weak positive correlation. The overall performance data show an average rating of M = 2.79 (SD = .65) for the sample we picked (n = 26) for our analysis. The overall performance (n = 60) was rated with M = 2.87 (SD = .70) on average for all crews. We conclude that there seems to be no selection bias in our sample in the sense that better eye-tracking data quality is not associated with either better or worse crew performance. Correlation analysis between recurrence rate in the last flight phase and the overall LOSA performance data showed a positive but weak correlation of $r(23)$ = .28, p = .09. The mean recurrence rates of the remaining flight phases (1 to 3) were not correlated with the performance measures.

Discussion

The results showed that the crews’ gaze was significantly coordinated around 17% of the time, as a function of time and the current task. This indicates that recurrence analysis is capable of revealing pilots' joint attention during a simulated flight scenario. There were, however, differences between individual crews (see Figure 5) raising the question of whether crews deliberately coordinate their behaviour via verbal communication, or whether their gaze is aligned as a result of the task they are working on. However, it has to be pointed out that the results we presented did not make any statement about strategy 2: *task-load sharing*. It would not be valid to assign the remaining 83% of time towards this strategy as we did not test for this hypothesis.

Our estimates of the percentage of time that crews showed coordinated gaze behaviour did not depend on the duration of the sliding window, for the range 10 to 30 seconds. However, the shorter the individual task time, the smaller the window size has to be in order to distinguish between the task and its corresponding recurrence. The high variance in the recurrence rate over time could correspond to adaptive coordination behaviour meaning that crews are well aware of their current situation and are able to adapt their workload management (choosing the strategy of *joint activity* or *task-load sharing*) to the dynamics of the system. Looking at the τ-recurrence, two of the example crews (Figure 8 top and middle) showed a shift of recurrence to a negative lag which indicates that, for these crews, the first officer was more likely to follow the captain's gaze than vice versa. The average lag over all crews also showed a shift of recurrences to the region above the diagonal, indicating that the captains are leading the first officers on average. These results seem plausible from the point of view that the captain is more experienced, holds the ultimate responsibility for the aircraft, and might therefore be expected to for example initiate decision making. Still it has to be kept in mind that these results only hold on a qualitative basis as we did not conduct any statistical test on the τ-recurrence. It might also be that leading a crew's gaze behaviour is rather a function of pilot role (pilot flying vs. pilot monitoring) than of crew position (captain vs. first officer). To answer this question, further research is necessary that includes statistical analyses.

The relationship between recurrence rate and instructor ratings of performance showed a weak positive correlation, but was not statistically significant. This may have been due to either a lack of reliability of the ratings, or the fact that a single overall performance rating is too coarse. A finer division into individual tasks with objective performance criteria for each may reveal more complex relationships between performance and gaze behaviour.

The present approach suffers from a number of limitations: first, we do not have an objective performance measure to use as a gold standard; second, the subjective performance measure had poor reliability and would therefore not be expected to produce high correlations; finally, our current definition of flight phases is rather crude. We expect that an analysis of specific tasks (or sub-tasks) may help to unravel the differences in crew coordination processes.

Conclusion

Cross recurrence analysis has been applied to identify pilots' coordination strategies during different tasks. The method is capable of identifying *leading* behaviour, and we find (as expected) more leading by captains than by first officers. Further research is needed to extend the present results, and to discover what triggers joint activity. Our current efforts lie in improving the eye-tracking quality (e.g., pupil detection and calibration), and improving data synchronization for the remaining crews. We expect that after additional data processing we will be able to analyse gaze data from at least 45 of the 60 crews. We also plan to integrate inter-pilot communication data extracted from voice recordings. Analysis of the communication data from this data set with cross recurrence analysis has been able to distinguish between poor and outstanding performance (Gontar et al., 2016). The combination of both gaze and communication data will help determine whether communication is the origin of gaze alignment, as opposed to specific tasks or environmental factors.

Acknowledgements

This work was partially funded by TUM Graduate School (PG) and the Technologies for Airplane State Awareness (TASA) project of NASA's Airspace Operations and Safety Program (JBM). The authors are grateful to Sonja Schneider for her support in the project. This research used data from a previously published experiment (Gontar et al., 2015; Haslbeck & Gontar, 2014; Haslbeck et al., 2012) but does not overlap with any prior finding. The original experiment was partially funded by the German Federal Ministry for Economic Affairs and Energy via the Federal Aeronautical Research Program LuFo IV-2 (20V0803).

References

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2014). A comparison of scanpath comparison methods. *Behavior research methods.* doi:10.3758/s13428-014-0550-3

Baker, D. P., Day, R., & Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health services research, 41*(4-2), 1576–1598. doi:10.1111/j.1475-6773.2006.00566.x

Clark, H. H. (2007). *Using language* (7. print). Cambridge: Cambridge Univ. Press.

Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence Plots of Dynamical Systems. *Europhysics Letters (EPL), 4*(9), 973–977. doi:10.1209/0295-5075/4/9/004

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.

Gontar, P., Fischer, U., & Bengler, K. (2016). Methods to Evaluate Crew Communication in a Training Environment: Speech Act Based Analyses vs. Cross Recurrence Analysis. In A. Droog, M. Schwarz, & R. Schmidt (Eds.), *Proceedings of the 32nd Conference of the European Association for Aviation Psychology*. Groningen, NL.

Gontar, P., & Hoermann, H.-J. (2015b). Interrater Reliability at the Top End: Measures of Pilots' Nontechnical Performance. *The International Journal of Aviation Psychology, 25*(3-4), 171–190. doi:10.1080/10508414.2015.1162636

Gontar, P., & Hoermann, H.-J. (2015a). Reliability of Instructor Pilots' Non-Technical Skills Ratings. In *Proceedings of the 18th International Symposium on Aviation Psychology* (pp. 366–371). Dayton, OH: Wright State University.

Gontar, P., & Mulligan, J. B. (2015). A Metric to Quantify Shared Visual Attention in Two-Person Teams. In T. Pfeiffer & K. Essig (Eds.), *Proceedings of the 2nd International Workshop on Solutions for Automatic Gaze Data Analysis 2015 (SAGA 2015)* (pp. 37–38). Bielefeld.

Gontar, P., Porstner, V., Hoermann, H.-J., & Bengler, K. (2015). Pilots' Decision-Making under High Workload: Recognition-Primed or Not – An Engineering Point of View. In G. Lindgaard & D. Moore (Eds.), *Proceedings of the 19th Triennial Congress of the International Ergonomics Association* . Melbourne.

Hacisalihzade, S. S., Stark, L. W., & Allen, J. S. (1992). Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(3), 474–481. doi:10.1109/21.155948

Haslbeck, A., & Gontar, P. (2014). How Pilots Believe to Act - and how They Really Do: Insights from Flight Simulator Studies. In A. Droog (Ed.), *Aviation Psychology: facilitating change(s): Proceedings of the 31st EAAP Conference* (pp. 317–324). Malta.

Haslbeck, A., Schubert, E., Gontar, P., & Bengler, K. (2012). The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In S. Laundry, G. Salvendy, & W. Karwowski (Eds.): *Advances in Human Factors and Ergonomics, Advances in Human Aspects of Aviation* (pp. 561–568). Boca Raton: CRC Press.

Klinect, J. R., Murray, P., Merritt, A. C., & Helmreich, R. L. (2003). Line Operations Safety Audit (LOSA) - Definition and operating characteristics. In *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 663–668). Dayton, OH.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods, 45*(1), 251–266. doi:10.3758/s13428-012-0226-9

Lehsing, C., Fleischer, M., & Bengler, K. (2016). On the Track of Social Interaction - Using a Non-Linear Method to Visualize and Quantify Behavior. In *Proceedings of the 19th IEEE International Conference on Intelligent Transportation Systems*. Rio de Janeiro, Brazil.

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5-6), 237–329. doi:10.1016/j.physrep.2006.11.001

Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A, 302*(5-6), 299–307. doi:10.1016/S0375-9601(02)01170-2

Marwan, N., Schinkel, S., & Kurths, J. (2013). Recurrence plots 25 years later—Gaining confidence in dynamical transitions. *EPL (Europhysics Letters), 101*(2), 20007. doi:10.1209/0295-5075/101/20007

Marwan, N., & Webber, C. L. (2015). Mathematical and Computational Foundations of Recurrence Quantifications. In C. L. Webber & N. Marwan (Eds.), *Recurrence Quantification Analysis. Theory and Best Practices* (pp. 3–43). Cham: Springer International Publishing.

Mulligan, J. B., & Gontar, P. (2016). Measuring and Modeling Shared Visual Attention. In *Proceedings of the Computational and Mathematical Models in Vision Workshop (MODVIS)*. West Lafayette, IN. Retrieved from http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160008390.pdf

Proulx, R., Parrott, L., Fahrig, L., & Currie, D. J. (2015). Long Time-Scale Recurrences in Ecology: Detecting Relationships Between Climate Dynamics and Biodiversity Along a Latitudinal Gradient. In C. L. Webber & N. Marwan (Eds.), *Understanding Complex Systems. Recurrence Quantification Analysis* (pp. 335–347). Cham: Springer International Publishing.

Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, Gaze Coordination, and Beliefs About Visual Context. *Cognitive Science, 33*(8), 1468–1482. doi:10.1111/j.1551-6709.2009.01057.x

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject*. Oxford, England: John Wiley & Sons.

Shockley, K., & Riley, M. A. (2015). Interpersonal Couplings in Human Interactions: Recurrence Quantification Analysis: Theory and Best Practices. In C. L. Webber & N. Marwan (Eds.), *Recurrence Quantification Analysis. Theory and Best Practices* (pp. 399–421). Cham: Springer International Publishing.

Zbilut, J. P., & Webber, C. L. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A, 171*(3-4), 199–203. doi:10.1016/0375-9601(92)90426-M

Contact Information

Patrick Gontar
Institute of Ergonomics
Technical University of Munich
Boltzmannstr. 15
85747 Garching, Germany
Tel. +49.89.28915428
E-mail. gontar@tum.de