Methods to Evaluate Pilots' Cockpit Communication: Cross-Recurrence Analyses vs. Speech Act–Based Analyses

Patrick Gontar, Technical University of Munich, **Ute Fischer**, Georgia Institute of Technology, and **Klaus Bengler**, Technical University of Munich

The training and evaluation of the crew resource management skills of pilots play an essential role in increasing flight safety, as they aim to reduce human error in aviation operations. Communication between pilots is a critical crew resource management skill, as flying an airplane requires coordinated action and collaboration by the flight deck crew. However, research that studied flight instructors' agreement in (and, thus, the accuracy of) their evaluation of pilots' communication behavior found little consistency in their judgments. As such, the present research explores the feasibility of a content-free approach—cross-recurrence analysis—to assess crew communication, in contrast to commonly employed content-based approaches that are grounded in speech act analysis. Results indicate that cross-recurrence analysis can identify communication patterns associated with high and low crew performance. We discuss the implications that these results may have for future research and communication assessment in pilot training.

Keywords: communication, topics, cross-recurrence analysis, aviation, domains, expert performance, topics

INTRODUCTION

Crew resource management (CRM; formerly, *cockpit resource management*) has been a component in the training of commercial pilots since the late 1970s. Research into the causes of aircraft accidents led to the surprising realization that, in most cases, shortcomings in pilots' nontechnical skills (e.g., leadership, crew communication, and coordination) were critical factors, rather than inadequate technical expertise (Cooper, White, & Lauber, 1980; Helmreich, Merritt, & Wilhelm, 1999). The aviation industry responded by institutionalizing courses in which pilots are trained in principles and practices for managing resources and working as a team to ensure shared and accurate situation awareness and risk assessment as well as sound decision making. Today, CRM is a mandatory component in pilot training and proficiency checks.

One essential CRM skill is team communication insofar as it is a critical team process that affects a variety of team performance aspects (Salas, Cooke, & Rosen, 2008; Salas, Sims, & Burke, 2005). "Communication is the glue that binds participants together in group interaction or team tasks" (Orasanu, Fischer, & Davison, 1997, p. 2), ensuring that team members have a common awareness and understanding of the situation, their required tasks, and their individual responsibilities (Serfaty, Entin, & Johnston, 1998). Communication is also the means by which team members provide performance-relevant feedback and coordinate adaptive responses to changing task conditions (DeChurch & Mesmer-Magnus, 2010; Kozlowski, Gully, Nason, & Smith, 1999; Orasanu, 1990; Orasanu & Fischer, 1992). The important role that team communication plays in crew performance heightens the need for measures that facilitate its objective and reliable assessment during training.

Cockpit Communication

Current commercial flight operations typically include a cockpit crew of two pilots—a captain (CPT) and a first officer (FO)—who differ in rank but are jointly responsible for flight safety and who share flying duties by alternating their roles of pilot flying or pilot monitoring after the completion of each flight segment. Flying an aircraft is a highly

Address correspondence to Patrick Gontar, Institute of Ergonomics, Technical University of Munich, Boltzmannstr. 15, 85748 Garching, Germany, gontar@tum.de.

Journal of Cognitive Engineering and Decision Making
2017, Volume 11, Number 4, December 2017, pp. 337–352
DOI: 10.1177/1555343417715161

rule-governed activity. Procedures determine how, when, and in which order tasks must be performed. Likewise, crew communication as part of the routine tasks is standardized and follows rules defined in standard operating procedures (SOPs) or flight operation manuals (Sassen, 2005). SOPs include callouts that crewmembers are required to make at specific points in time during a flight. For instance, they must announce when a prespecified altitude has been reached. SOPs also refer to checklists that detail crew actions in response to routine or abnormal events. During normal operations, one crewmember—typically, the pilot monitoring—reads aloud the checklist, and the pilot flying acknowledges each item. The sequential structure of checklists not only guides crewmembers through often complex tasks and ensures that they have a shared situation understanding but also supports the pilots in terms of anticipating each other's behavior.

However, not all pilot communication is covered by SOPs. Pilots work in a dynamic task environment, where changing conditions require that they continuously reassess their original plan and respond adaptively to evolving events. The non-SOP talk follows the norms and conventions of everyday discourse. It goes beyond the exchange of routine information and concerns flight safety–related events and pilots' problem-solving efforts (Orasanu & Fischer, 1992). Characterizations of non-SOP communication have sought to identify markers of effective performance and typically focused on the contents of pilots' contributions, specifically on their function regarding problem-solving efforts and intentions: Do pilots talk about the problem at hand? Do they collect information? Do they state their plans and intentions? Do they acknowledge each other's contribution? Agree or disagree with it? This approach is grounded in speech act theory (Austin, 1962; Searle, 1969)—that is, the idea that speakers "do things with words" (Austin, 1962) and that language itself can be seen as an action that is shaping reality (Lacity & Janson, 1994). For instance, speakers who state a fact intend to influence what others believe; they issue orders in an attempt to direct the behavior of others; or they ask questions to solicit input from them.

Speech Act Theory Approaches

Previous research in the aviation domain indicates that effective crew communication addresses critical components of a crew's task and teamwork and promotes shared situation models. Effective crews tend to talk more about the problem that they face and their response to it than do poorly performing crews (Bourgeon, Valot, & Navarro, 2013; Helmreich & Foushee, 2010; Krifka, Martens, & Schwarz, 2004; Mjos, 2001; Mosier & Fischer, 2010; Sexton & Helmreich, 2000). In particular, successful teams are more likely than unsuccessful teams to state their plans, refer to changes in task assignment, and articulate expectations and beliefs about future situations (Gillan, 2003; Mosier & Fischer, 2010; Orasanu & Fischer, 1992). Furthermore, Krifka et al. (2004) found that well-performing CPTs are more likely to encourage their FOs to speak up and that the FOs consequently contribute more than FOs in poorly performing crews. Compared to less effective teams, members of effective teams are also more likely to anticipate the information that their teammates need and to provide it to them, rather than waiting for them to request it. Team members' anticipatory behavior was operationalized by Serfaty, Entin, and Johnston (1998) as the ratio of information transfers and information requests. This so-called anticipation ratio provides a measure of communication efficiency (Nonose, Kanno, & Furuta, 2015), as it was shown to be associated with team effectiveness (Shah & Breazeal, 2010; Sperling, 2006).

Whereas speech act–based approaches of communication analyses have been successful in specifying characteristics of effective crew communication, they are not readily transferable to crew training. Proficiency in communication coding requires extensive training. Moreover, the classification of team members' communication is cognitively taxing and may place excessive workload demands on instructor pilots during a training session, as they need to perform additional tasks, such as operating the simulator, acting as air traffic controller, and evaluating pilots' procedural adherence and decisions.

Current Practice in Pilot Training

To keep their workload at a manageable level, instructors typically assess pilots' CRM skills by looking for the occurrence of specific behaviors, so-called behavioral markers, associated with specific performance dimensions, such as leadership and communication. For example, in current practice, instructor pilots evaluate trainees' communication skills by rating the extent (poor to outstanding) to which the crewmembers "announce ambiguities, clearly state plans and intentions, share information, and assure reception" (Brandt, 2010, p. 11). At first sight, the rating of behavioral markers seems to be a straightforward and objective approach to performance assessment. However, current research suggests otherwise. Recent analyses by Gontar and Hoermann (2015a) found significant disagreement among 37 highly experienced type-rating examiners. They judged pilots' CRM skills under the best of possible rating conditions (i.e., ratings were done from video recordings of flight deck crews, rather than in real time). In particular, raters showed their lowest interrater reliability when they assessed crew communication (intraclass correlations: ICC2 = .12, with a novel and thus unfamiliar rating tool; ICC2 = .22, with a familiar rating tool). Likewise, Gontar, Hoermann, Deischl, and Haslbeck (2014) as well as Yule and colleagues (2009) observed that team communication and teamwork were less reliably assessed compared to other nontechnical skills. These findings suggest that raters may focus on different behavioral aspects or use different approaches (e.g., instance based vs. holistic) to evaluate team members' communication behavior (for further discussion, see Weber, Mavin, Roth, Henriqson, & Dekker, 2014). Additionally, raters' judgments may be affected by their experience with, and social distance to, the individuals whose performance they assess. For instance, research by Gontar and Hoermann (2015a) and O'Connor, Hoermann, Flin, Lodge, and Goeters (2002) indicate that instructor pilots (who are most often CPTs themselves) tend to be more consistent in rating the performance of FOs than that of their own peer group, presumably because during their day-to-day professional experience, they need to size up the ability of their junior crewmembers. As discussed by Gontar and Hoermann (2015a) and Holt, Hansberger, and Boehm-Davis (2002), low rater reliability can lead to degrading training standards or to unnecessary additional training sessions. In any case, low interrater reliability results in inappropriate and inconsistent feedback to the trainee.

Content-Free Approaches

In contrast to speech act–based approaches, content-free approaches to the analysis of team communication could provide a less time-consuming and highly reliable alternative—perhaps even one that can be accomplished in real time. Instead of examining what team members talk about, content-free methods focus on such aspects as communication frequency and duration, to capture the degree of influence that one team member has over his or her mates (Cooke & Gorman, 2009). The analysis of communication flow (who is talking to whom and for how long) has been used to characterize team coordination and to identify shifts in team members' interaction patterns that are adaptive or poor responses to off-nominal events (Gorman & Cooke, 2011). For instance, Fischer, McDonnell, and Orasanu (2007) analyzed the communications among four members of search-and-rescue crews during simulated missions and noted that team interactions in successful teams were more balanced and inclusive than those in poorly performing teams. Members of successful teams participated equally in the team's discussion, whereas conversations in unsuccessful teams tended to occur in subgroups to the exclusion of others. Cooke and colleagues developed various techniques for automating measurements of communication flow patterns. For example, they employed computational tools, such as sequential analysis (ProNet) and clustering models (CHUMS), to identify communication sequences (i.e., who is talking after whom) and shifts in these sequences that are predictive of effective team performance in a dynamic task environment (Kiekel, Cooke, Foltz, Gorman, & Martin, 2002; Kiekel, Gorman, & Cooke, 2004). In more recent work, Gorman, Cooke, Amazeen, and Fouse (2012) applied recurrence analysis to characterize the extent to which team members adhere to rigid interactive patterns over time or adopt more flexible coordination patterns supportive of adaptive behavior. Other applications of recurrence analysis sought to isolate thematic patterns in team members' interactions (e.g., Topic A is followed by Topic B and Topic A again before discussing Topic C) rather than structural patterns (e.g., Agent A talks after Agents B and C). For instance, Angus, Smith, and Wiles (2012) adapted recurrence analysis to capture conversational strategies that the CPT of United Flight 232 used to manage the joint problem solving by the flight deck crew and ground support as they were trying to control their aircraft after having lost all hydraulic systems.

In the present study, we analyzed cockpit crew communication (two-person teams) using both content-free recurrence analysis and speech act–based analysis. The goal of this effort was to determine how well each approach can distinguish between poorly performing and outstanding teams and whether a content-free approach would provide a viable alternative to an approach based on the analysis of crewmembers' speech acts. We conclude with suggestions for further research on how to use content-free methods in the training environment.

METHOD

Participants

A total of 120 airline pilots participated in our experiment (Gontar & Hoermann, 2014). All pilots were from the same major European airline and held valid licenses for their aircraft types. Pilots were randomly selected from the airline's roster and scheduled for participation. The decision to recruit pilots from the same airline and to select pilots randomly was driven by an effort to minimize confounding factors such as familiarity and training influences. Participants were compensated by the airline as part of their duty time. There were 60 CPTs and 60 FOs, with an equal number of pilots in each group who were Airbus A340 and Airbus A320 type rated.

Materials and Apparatus

Flight simulator. The experiment took place at the training facility of the participating airline and used two of their full-motion flight simulators (*JAR STD* 1A, Level D): one configured as an Airbus A320, the other as an A340.

Scenario. The simulation presented the crew members with a leak in the hydraulic system as they approached their destination airport. The hydraulic leak led to an unsafe gear configuration (first failure) and subsequently to jammed slats/flaps (second failure). A340 crews experienced these issues on approach to John F. Kennedy Airport in New York; for A320 crews, these events were embedded in the approach to Nice Côte d'Azur Airport in Nice. The simulation started shortly before the initial approach fix and ended with stopping on the runway. To enhance the scenario's realism, recorded communications between air traffic control and pilots were played during the simulation, and participants had to wait for a break before they could talk to the air traffic controller. For further information on the scenario, see Gontar and Hoermann (2015a).

Flight information package. The flight information package provided to the crew included the approach plate showing the upcoming approach and all necessary navigational information, detailed data regarding the approach (i.e., current aircraft configuration, position, altitude, speed, zero fuel weight, and fuel on board), as well as weather and visibility data.

Procedure

We tested six pilots (three crews) per day, either in the late afternoon or during the night. Participants received an email providing them with the flight information package and experiment schedule. Before the simulation session, the crew was given time to conduct its approach briefing. The pilot role during the scenario—that is, who is flying the aircraft and who is pilot monitoring—was not fixed a priori, but the decision was left to the crew. After the crew had completed the approach briefing, the simulation began.

The flight simulation lasted on average 28.2 minutes (*SD* = 5.02). Two flight instructors who recently had retired from the airline were trained to assist with the simulation, with experimental runs evenly divided between them. However, both instructors were present during six runs, for assessment of interrater reliability in terms of their performance ratings. The flight instructor assigned to an experimental run was in the simulator with the crew and had several responsibilities. He operated and monitored the simulator and acted as air traffic controller. Simulator flight data and eye-tracking data on the pilots were recorded during the simulation; however, analyses of these measures are not included in this paper but are reported elsewhere (see Gontar & Mulligan, 2016; Haslbeck & Hoermann, 2016). Each crewmember and the flight instructor wore individual microphones to record their communications.

After the simulation, pilots were asked to rate their own and their crew partners' nontechnical skills using the rating tool commonly used by the airline. This tool involves 40 items tapping four CRM dimensions: communication, leadership and teamwork, workload management, and situation awareness and decision making (Brandt, 2010). Ratings are given on a 5-point scale ranging from *poor* to *outstanding*. Ratings were provided individually (for results and further discussion, see Gontar et al., 2014; Gontar & Hoermann, 2015b). Subsequently, during a joint debriefing session, pilots were asked to review the scenario and explain their decisions. These findings are not reported in this paper but are discussed by Gontar, Porstner, Hoermann, and Bengler (2015).

Measures

Crew performance. The instructors assisting with the simulation also served as raters of crewmembers' performance. The pilot's performance was assessed by means of the Line Operations Safety Audit's "Approach and Landing Sheet" (Klinect, Murray, Merritt, & Helmreich, 2003). The tool comprises 13 behavioral markers relating to four dimensions of a crew's performance: planning, plan execution, review and modification, as well as overall performance. Ratings were given on a 4-point scale ranging from *poor* to *outstanding*. To derive a global performance grade for each crew, we calculated the mean of the four performance dimensions. The communication of a crew was not directly measured, although some markers (e.g., SOP briefing, workload assignment, and contingency management) were based on the contents of crewmembers' communications. The raters were with the crew in the flight simulator and gave their performance ratings while they observed a crew during the simulation.

Interrater reliability of crew performance judgments. To assess interrater reliability, six crews were assessed by both instructors. We calculated intraclass correlation coefficients for every performance dimension using two-way models (ICC2) and found an average reliability of .39 (based on Fisher z transformation; Fisher, 1925). Test-retest reliability was found to be .69 on average (Gontar & Hoermann, 2015b). In light of these findings and previous research showing considerable disagreement between instructors in their ratings of crew behavior, we decided to adopt an extreme-group design. That is, the present analysis includes only data from the top six crews, as well as the data from the six crews with the lowest performance scores (from a total of 60 crews). This design assumes that extreme behaviors can be assessed with sufficient precision (Yule et al., 2009)—an assumption that has been confirmed at least with respect to the performance of outstanding pilots (Gontar & Hoermann, 2015a).

Content-based communication coding: Speech acts. Crewmembers' communications were coded with STACK (a German acronym for Sprechakt-Typeninventar zur Analyse von Cockpit-Kommunikation), a speech act inventory developed by Krifka et al. (2004) to analyze cockpit communication. The inventory includes 50 speech acts that are clustered in seven categories: information sharing (n = 13), request (n = 5), agreement/negotiation (n = 5), dissent (n = 7), question (n = 5), expressive (n = 10), and interaction marker (n = 5). To account for communications in which crewmembers read from checklists and procedures, we added the category *procedure related*. The communication coding is illustrated in Table 1. The excerpt provides the communications between the CPT and FO right after they realized that the landing gear was not down and locked.

Communication coding was done directly from video recordings of crewmembers' interactions during a given scenario. Videos were loaded into INTERACT, a video-editing software that allows human raters to segment a video stream into

TABLE 1: Sample Communication From One of the Crews With Corresponding STACK Categorization

		Speech act	
Crewmember	Utterance	Type	Category
CPT	Hmm, ich würd sagen, du machst Funk und fliegst. "Hmm, I suggest you take the radio and fly the plane."	Suggest	Request
FO	Ok	Agree	Agreement
CPT	und ich versuch das abzuarbeiten, was meinst? "and I try to work through this, what do you think?"	Question (opinion)	Question
FO	ja, ist in Ordnung "yes, that's fine"	Confirm	Agreement
CPT	Landing gear not down-locked	ECAM[a]	Procedure
CPT	landing gear recycle	ECAM[a]	Procedure
CPT	Ich probier's mal "I'm going to try it"	Announce	Information
FO	ok, go ahead	Confirm	Agreement
FO	es fehlt vor allem ne green indication hier "Most significantly the green light is missing here"	Point out	Information
CPT	Ja "Yes"	Agree	Agreement
CPT	Der Sektor ist ok, gell auf 4000ft? "This sector is ok, at 4000 ft, right?"	Question (information)	Question
FO	Ja "Yes"	Agree	Agreement

Note. Speech act types and categories according to Krifka et al. (2004). CPT = captain; FO = first officer.
[a]"ECAM" refers to the electronic centralized aircraft monitor, an automated system in an Airbus that monitors aircraft functions and indicates system status and, when necessary, failures and procedures.

meaningful units and attach codes to them. Communication was presented in stereo via headphones in a way that the CPT's voice was presented on the left channel and the FO's on the right channel (reflecting pilots' seating arrangement in the cockpit) to make it easier for a rater to attribute communication to the CPT or FO in a crew. One of the authors (P. G.) and one assistant who had classified the crews' coordination behavior for a different project (Reichling, 2017) and who thus was very familiar with the scenario coded crewmembers' communications using the STACK system. To establish coding reliability, the raters independently coded communications of four crews. Cohen's kappa (κ) was calculated on raters' codes at the category level. A κ value of .79 was obtained, which is considered excellent agreement (Fleiss, Levin, & Paik, 2003). In subsequent analyses, we operationalized the speech act density of a crew by calculating the frequency of speech acts divided by the scenario duration.

Content-based communication coding: Anticipation ratio. The anticipation ratio is a measure of the extent to which crewmembers are sensitive to each other's information needs.

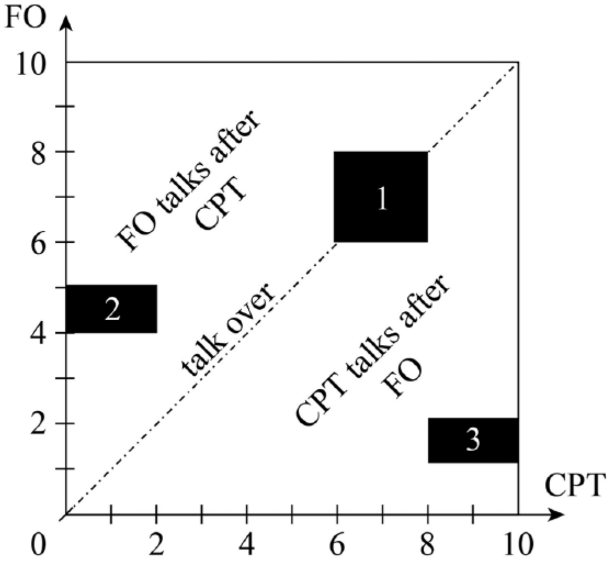

Figure 1. Generic cross-recurrence plot of crewmembers' communication. CPT = captain; FO = first officer.

$$\text{Anticipation Ratio}_{STACK} = \frac{\text{Number of Agreements} + \text{Number of InformationSharing}}{\text{Number of Questions} + \text{Number of Requests}}$$

We adapted the formula suggested by Serfaty et al. (1998) to include the following STACK coding categories:

Content-free communication measures: Relative speech duration. We analyzed the duration of speech as a function of crew position (CPT or FO). To do so, we measured the total time of crew communication and calculated the percentage of verbal communication for each crewmember.

Content-free communication measures: Cross-recurrence analysis. We constructed cross-recurrence (CR) plots to represent the direction and distribution of speech. To analyze the relation between two dynamical systems (CPT and FO, in our case), CR plots offer a nonlinear approach that has also been used to quantify pilots' shared gaze behavior (Gontar & Mulligan, 2016). CR plots are graphs that depict recurring structural or thematic relationships among events or states over time. As our approach aims for a content-free analysis, we were not interested in thematic patterns in crewmembers' communications but in how they coordinate their verbal communications. We thus examined the following relationships between crewmembers' communications: Did FOs consistently speak after CPTs (i.e., in considering the sequential nature of the communications, we assumed that they responded to the CPTs)? Did CPTs take the turn after communications by their FOs? How often did crewmembers speak at the same time (talk over)?

Mathematically, a CR plot of two dynamical systems can be defined as follows (Marwan & Kurths, 2002):

$$C(t_1,t_2) = \begin{cases} 1, x(t_1) = y(t_2) \\ 0, \mathit{otherwise} \end{cases}, \quad t_1, t_2 = 1 \ldots N,$$

where N represents the number of temporal samples, t_1,t_2. In our definition, the function x represents the CPT's speech; the function y represents the FO's speech. Both functions are binary and can take a value of 1 (in case the pilot is speaking) or 0 (if he or she is not speaking). As we are interested only in recurrences of speech, we did not regard both pilots' silences as a recurrence. The resulting matrix C has the size of $N * N$ and can be displayed in the form of a plot with the x- and y-axes indicating progression of time, where values of 1 are displayed as a black dot at (t_1,t_2), constituting the CR plot (see Figure 1).

As shown in the generic example in Figure 1, black dots on the diagonal indicate speech of the two pilots at the same time (1), as all dots on the diagonal represent the same point in time in both systems (CPT and FO). The talkover starts at t_{CPT} = t_{FO} = 6 s and lasts until t_{CPT} = t_{FO} = 8 s. Dots above the diagonal (2) represent points where the FO is talking (from t_{FO} = 4 s until t_{FO} = 5 s) after the CPT, who is talking from t_{CPT} = 0 s until t_{CPT} = 2 s; recurrences below the diagonal (3) represent points in time at which the CPT speaks (t_{CPT} = 8 s until t_{CPT} = 10 s) after the FO (t_{FO} = 1 s until t_{FO} = 2 s). Given a CR plot, one can measure the density of recurrent states (black dots) per plot. This measure is called the *recurrence rate*.

CR plots for crewmembers' communications during the scenario were constructed with a windowing approach (Zbilut, Thomasson, & Webber, 2002). We partitioned crewmembers' communications during the scenario (average duration, ~28 min) into segments of 10 s and calculated the recurrence plot for each segment. Next the average CR plot was calculated and used to assess the average recurrence rate of different events (e.g., CPT talks after FO; FO talks after CPT) in a crew across the scenario. The decision to use the 10-s window was based on previous research on pilots' coordinated gaze behavior (Gontar & Mulligan, 2016) showing that this is an appropriate length for measuring pilots' interaction. For further details on how CR plots and analyses can be used to assess pilots' behavior, see Gontar and Mulligan (2016).

RESULTS AND DISCUSSION

All statistical tests assume a significance level of $\alpha < .05$.

Content-Based Analyses of Crew Communication

As flight times of the scenarios varied by crew, we normalized the frequency of each speech act category by scenario time to yield a measure of speech act density. Data screening (Mauchly's test) revealed that the sphericity assumption was not met for the variable speech act category, $\chi^2(14) = 42.67$, $p < .001$. Consequently, Greenhouse-Geisser corrected degrees of freedom were used in the analysis of this variable. We performed a mixed-design analysis of variance on speech act density. Performance level (poorly performing vs. outstanding) was the between-group variable and speech act category, the within-group variable. A significant main effect of speech act category was found, $F(2.71, 27.12) = 42.98$, $p < .001$, $\eta_p^2 = .81$. As can be seen in Figure 2, crewmembers used speech acts of some categories (agreement, procedure, information) more frequently than others (request, questions, dissent, interaction, emotional). However, poorly performing and outstanding crews did not differ regarding the overall frequency of speech acts. Whereas the effect of performance showed the expected tendency such that poorly performing crews communicated less than outstanding performing crews, it was found to be only marginally significant, $F(1, 10) = 4.26$, $p = .07$, $\eta_p^2 = .30$. Similarly, the interaction of performance level and speech act category was found to be marginally significant, $F(2.71, 27.12) = 2.73$, $p = .07$, $\eta_p^2 = .21$. Bonferroni-adjusted pairwise comparisons revealed that outstanding and poorly performing crews differed only on the rate of their procedure-related communications, with outstanding crews having a higher speech act density ($M = .13$, $SD = .03$) than poorly performing crews ($M = .08$, $SD = .03$, $p = .008$).

Next, we calculated the anticipation ratio for outstanding and poorly performing crews, as described by Serfaty et al. (1998). An independent t test showed that outstanding crews had a significantly higher anticipation ratio ($M = 3.71$, $SD = .99$) than poorly performing crews ($M = 2.52$, $SD = .44$), $t(10) = 2.45$, $p = .018$ (one-tailed). That is, members of outstanding crews, in contrast to their low-performing colleagues, were more likely to volunteer information than to wait for their teammates to request it. This finding suggests that outstanding crewmembers were more attuned to their teammates' information needs than members of poorly performing crews: a finding consistent with past research (e.g., Butchibabu, Sparano-Huiban, Sonenberg, & Shah, 2016).

Content-Free Analyses of Crew Communication

To conduct the content-free analyses, crewmembers' audio recordings were processed with

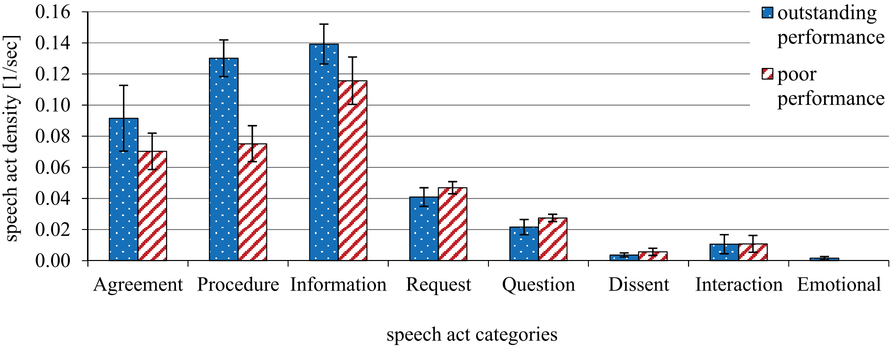

Figure 2. Speech act density for different categories of the expanded STACK system. Error bars refer to the standard error.

the software tool Audacity. The software's filter feature allowed us to identify and delete noise from the recordings (e.g., simulator noise or breathing). The remaining recordings were analyzed with a sampling rate of 25 Hz, and binary codes were assigned to indicate segments with a pilot talking (1) or segments without speech (0). CR plots relating crewmembers' segments of speech and silence across time were drawn with the toolbox provided by Marwan, Carmen Romano, Thiel, and Kurths (2007).

We conducted a 2 (performance level) × 2 (crew position) analysis of variance on communication rate (defined as the percentage of total time a crew spent communicating). Performance level (poorly performing vs. outstanding) was the between-group variable; crew position (CPT vs. FO) was the within-group variable. As can be seen in Figure 3, there were no significant main effects of crew position, $F(1, 20) = 1.38$, $p = .25$, or performance, $F(1, 20) = .001$, $p = .98$, nor was there a significant interaction effect of performance level by crew position, $F(1, 20) = 10$, $p = .75$. This result indicates that in our sample, performance differences among crews were not related to how much the members of a crew talked with each other.

Next, we examined the direction of crewmembers' communication—that is, who was communicating and whether there was a turn. To tackle this question, we constructed CR plots for each crew using a windowing approach with a bin size of 10 s and a step width of 40 ms (sample rate). Specifically, we analyzed whether, after communications by one pilot, his or her colleague took the turn within 10 s.

By applying this windowing approach to the communications of a crew throughout the flight scenario, we obtained about 45,000 plots for each crew. Plots were subsequently averaged and normalized by the duration of the flight segment. Figure 4 shows the averaged CR plots of sequential patterns calculated for the six poorly performing crews (two left columns) and the six outstanding crews (two right columns). The recurrence rate ranged from 0 (indicated by points in dark blue) to 0.1 (shown in bright yellow). The axes of a plot indicate a point in time (e.g., 2 s, 4 s) within a given 10-s window. For instance, in Plot 2, we can see that whenever the FO talked at Second 2, the CPT was likely to take the turn at Second 6. This results in a time lag of 4 s between CPT and FO (see Figure 4, bottom). The coloring of the area above the diagonal in a CR plot indicates the recurrence rate of communication sequences in which the CPT initiated a conversation and the FO spoke in the subsequent turn. The recurrence of the converse relationship is depicted in the area below the diagonal. The CR plots as well as the recurrence lag distribution revealed that the communication sequence (i.e., who leads the conversation and who follows) was less fixed (recurrent) in outstanding crews ($M = .057$, $SD = .010$) than in poorly performing crews ($M = .069$, $SD = .006$), $U = 5$, $p = .037$ (two-tailed). This finding is reflected by the different coloring of the CR plots obtained for poorly performing and outstanding crews (see top panel of Figure 4). As shown, the plots of poorly performing crews tend to be brighter than those of outstanding

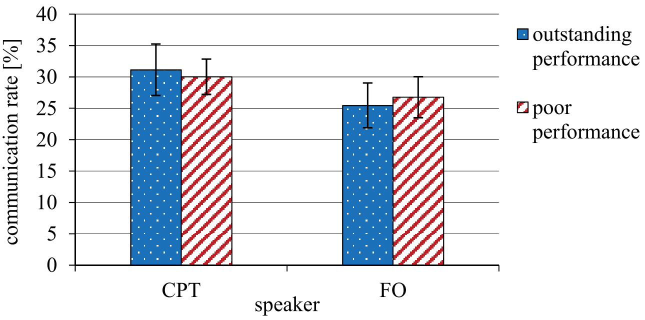

Figure 3. Analysis of communication rate as a function of crew position and performance. Error bars refer to the standard error. CPT = captain; FO = first officer.

crews. Again, as illustrated in the bottom panel of Figure 4, the mean recurrence rate shown across time tended to be higher for poorly performing crews versus outstanding crews. That is, members of poorly performing crews were more likely than members of outstanding crews to respond to a teammate's communication within the 10-s window.

One possible explanation for this finding is that members of outstanding crews spoke >10 s; thus, our selected window size may have been too short to capture the teammate's response. This explanation is consistent with the finding reported earlier that outstanding crews had significantly more procedure-related communications than poorly performing crews. Communications relating notes in checklists and procedures may take >10 s, thus exceeding our preset analytic window. This fact may have contributed to the low recurrence rate obtained for outstanding crews.

Note also that the diagonals in the CR plots in Figure 4 are drawn in blue, indicating that crewmembers rarely spoke at the same time. The recurrence rate of the outstanding crews at a lag of 0 ± 40 ms (talk at the same time) was lower (M = .032, SD = .008) than that of poorly performing crews (M = .044, SD = .010), U = 5, p = .037 (two-tailed). This finding suggests that members of outstanding crews showed fewer instances at which they talked over a teammate or interrupted him or her, in contrast to members of poorly performing crews. This finding is also consistent with Nevile (2007) who posits that overlapping talk by crewmembers may indicate workflow management problems or rushed performance. Moreover, since the width of the diagonal represents the average time lag between communications by crewmembers, inspection of Figure 4 suggests that turn taking or answering time differed among crews (cf. Crew 2 vs. Crew 11).

In most crews, CPTs and FOs were equally likely to initiate a communication. This finding is reflected in the CR plots of Figure 4 by the fact that areas above and below the diagonal have the same coloring. However, some crews (see Crews 6, 9, and 10) seem to be different in this respect. In the CR plot representing sequential patterns in Crews 9 and 10, the area below the diagonal is brightly colored, indicating that the CPT responded more frequently to the FO than vice versa. The CR plot of Crew 6 shows the reverse pattern (area above the diagonal is brightly colored).

LIMITATIONS

We need to point out that our content-free analysis of two-member teams provides for limited variance among teams. Teams may differ only with respect to the length of members' utterances and the time lag between utterances. Teams with more than two members would have more sources of variance; for instance, they could vary in terms of the sequence of speakers. For instance, Fischer et al. (2007) studied flow patterns in four-person teams and noted that, in high-performing teams, all members contributed to the discussion, whereas poorly performing teams had one or two members who rarely participated. Furthermore, as we analyzed only

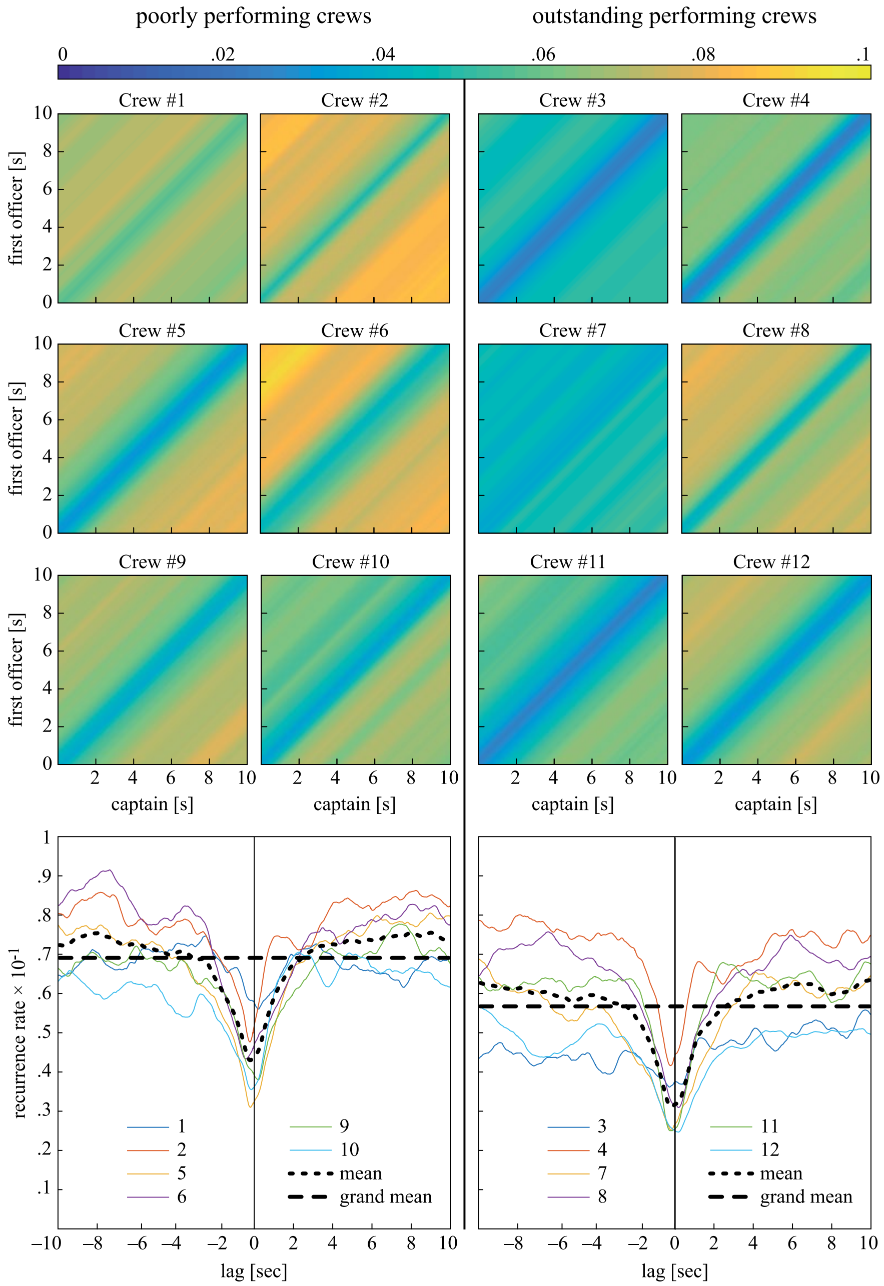

Figure 4. The top panel shows the average cross-recurrence plots representing the rate at which sequential patterns recur in poorly performing crews (two left columns) and outstanding crews (two right columns). The bottom panel shows the recurrence rate as a function of the time lag between the communications by members of poorly performing crews (left panel) and outstanding crews (right panel).

the best- and worst-performing crews, we do not know whether our results also apply to crews in the middle of the performance spectrum. We focused on high- and low-performing crews since instructor pilots were more consistent in their ratings of superior and poor crew performance than when they rated mediocre performance. Although the performance ratings did not concern crew communication but instead addressed discrete crew behavior (e.g., reasonable contingency management), instructors had to consider what pilots said in addition to what they did to make their assessments. Aspects of pilots' communications may have thus contributed to instructors' performance ratings. However, since our communication analyses targeted speech acts rather than the presence/absence of specific information, the danger of circularity in our data seems negligible.

CONCLUSION

The present research was conducted to compare the effectiveness of two approaches to assess crew communication. One method was content based and relied on the analysis of crewmembers' speech acts (i.e., the task-related functions of their communications) to identify differences between high- and low-performing crews. The second approach was content-free and sought to identify recurrent structural patterns in crewmembers' communications associated with different performance levels.

The speech act–based analysis of crew communication adopted the categories of the STACK coding system developed by Krifka et al. (2004). In contrast to their research, our analysis did not yield significant differences between outstanding and poorly performing crews in the extent to which they used speech acts of different categories; the exception was the procedure-related utterances. That is, the STACK categories did not reliably discriminate between crews of different performance levels. This finding may be the result of our small sample size. It may also reflect situational constraints, as our data collection was confined to off-nominal events. It is possible that during these events, performance differences between crews are less apparent at the level of speech acts but instead may be reflected in the extent to which they address different aspects of the off-nominal event and its management (see, for instance, Bourgeon et al., 2013; Orasanu & Fischer, 1992). Although our analysis of speech acts was inconclusive, we did observe differences in crewmembers' anticipation ratio. Consistent with past research (Nonose et al., 2015), we found that members of outstanding crews were more likely than members of poorly performing crews to volunteer relevant information at a time when their teammates needed it but had not yet asked for it. Even though a team's anticipation ratio may be a reliable indicator of the team's performance, it presently is an impractical tool for use during team training. Currently, it cannot be automatized, as the judgments that it requires exceed the capabilities of available technology. Likewise, human raters would not be able to provide the necessary judgments and computations in real time to be useful as performance feedback. However, it is conceivable that future speech recognition software will be sufficiently sophisticated to support this type of analysis and thus would enable the automatic assessment of crew communication during training.

The amount of communication, operationalized in the content-based approach as speech act density and in the content-free approach as communication rate, did not differentiate between outstanding and poorly performing crews. This finding contrasts with past research that observed more verbal communication by high-performing crews as compared with low-performing crews (Krifka et al., 2004); however, in other research (e.g., Gillan, 2003; Helmreich & Sexton, 2004; Orasanu & Fischer, 1992), quantitative differences in crew communication concerned specific categories of crewmembers' problem-solving talk rather than their overall talk.

Content-free analyses of crew communication identified significant differences between outstanding and poorly performing crews. Specifically, we observed that the recurrence rate of sequential patterns (i.e., who is talking when, after whom, and for how long) was lower for outstanding crews than for poorly performing crews. This finding may indicate that outstanding crews were more adaptive than poorly performing crews, in terms of the timing and length of their responses as well as with respect to the initiation of turns. This is consistent with Gorman et al. (2012) who also analyzed team interaction using CR approaches and concluded that flexible interaction patterns are associated with superior performance.

Our analysis revealed that members of poorly performing crews were more likely than members of outstanding crews to talk over each other. Simultaneous talk may reflect a breakdown in a crew's work flow or may be the result of disharmony between crewmembers (Nevile, 2007). Moreover, research on the impact of interruptions has shown that operators commit more errors (e.g., Bailey & Konstan, 2006) and perceive increased workload (Gontar, Schneider, Schmidt-Moll, Bollin, & Bengler, 2017; Weigl, Antoniadis, Chiapponi, Bruns, & Sevdalis, 2015) when they are interrupted. Measuring simultaneous talk between crewmembers thus seems to be a critical aspect of crew communication.

Although the present content-free CR analysis showed some promising results in distinguishing between crews of different performance levels, additional research should refine this approach to include analyses that consider contextual variables, such as the pilot's role (i.e., pilot flying vs. pilot monitoring) and tasks (e.g., decision making, procedures, manual approach). A more granulated approach of analysis could yield task-specific—and thus more meaningful and trainable—communication patterns associated with superior performance.

The advantages of using a content-free approach to the assessment of crew communication during CRM training are obvious: It is fast, objective, and reliable. CR algorithms are already capable of real-time analysis and thus could be implemented into a tool for use during simulator training (see also Gorman et al., 2012). Such a tool would provide pilot instructors with quantifiable information about a crew's communication behavior and could inform their feedback to the crew during the debrief. Moreover, the communication rating given by instructors would be more transparent to the trainees. Such a tool would also reduce instructors' workload during a simulation session, enabling them to focus on crew behavior that is more reliably assessed, such as trainees' adherence to and execution of procedures.

Such an objective communication measurement approach might also apply to team training in other high-reliability domains, such as health care, off-shore oil drilling, or nuclear power plants. As in aviation, analyses of incidents and accidents in these domains have revealed breakdowns in teamwork among operators and led to the adoption of CRM-inspired modules into training programs (Flin, O'Connor, & Mearns, 2002). It is thus likely that trainers face comparable challenges in rating operators' communication skills as were observed for instructor pilots. Although the current research addresses approaches to assessing pilot-crew communication, findings might generalize to these domains as well.

Our ongoing research focuses on joint recurrence analyses that combine communication behavior and joint visual attention data collected during the flight simulation with dual eye tracking. This approach attempts to quantify multimodal meaning making in the cockpit, an idea advanced by Hutchins, Weibel, Emmenegger, Fouse, and Holder (2013). In combining information on pilots' joint visual attention with their communication behavior, we can answer questions such as the following: How much time does a team need to develop shared situation models? That is, how much time elapses between the detection of a problem and a shared understanding of the situation? This type of analysis further allows us to study how a team achieves shared attention and to understand the extent to which highly coordinated gaze behavior requires (or does not require) verbal communication. An integrated analysis of joint visual attention and crewmembers' verbal communication would deepen our understanding of how both processes jointly facilitate team coordination and performance.

ACKNOWLEDGMENTS

We acknowledge the support of Alexandra Ruhl for processing audio data for the content-free analyses and Julia Reichling for her support in the coding process. This article uses unpublished data from a previously published experiment (Gontar & Hoermann, 2015a, 2015b; Gontar et al., 2014; Gontar et al., 2015; Gontar & Mulligan, 2016). None of the results reported here overlap with any prior published publication. Part of the content-based analysis (speech act density) was adopted from P. G.'s diploma thesis. 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

Angus, D., Smith, A. E., & Wiles, J. (2012). Human communication as coupled time series: Quantifying multi-participant recurrence.

IEEE Transactions on Audio, Speech, and Language Processing, *20*(6), 1795–1807. doi:10.1109/TASL.2012.2189566

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, *22*(4), 685–708. doi:10.1016/j.chb.2005.12.009

Bourgeon, L., Valot, C., & Navarro, C. (2013). Communication and flexibility in aircrews facing unexpected and risky situations. *International Journal of Aviation Psychology*, *23*(4), 289–305. doi:10.1080/10508414.2013.833744

Brandt, T. (2010, March). *Basic performance of flight crew*. Paper presented at the Next Generation of Aviation Professionals Symposium, Montreal, Canada.

Butchibabu, A., Sparano-Huiban, C., Sonenberg, L., & Shah, J. (2016). Implicit coordination strategies for effective team communication. *Human Factors*, *58*(4), 595–610. doi:10.1177/0018720816639712

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of Cognitive Engineering and Decision Making*, *3*(1), 27–46. doi:10.1518/155534309X433302

Cooper, G. E., White, M. D., & Lauber, J. K. (1980). Resource management on the flightdeck. In *Proceedings of a NASA/Industry Workshop (NASA CP-2120)*. Moffett Field, CA: NASA–Ames Research.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, *95*(1), 32–53. doi:10.1037/a0017328

Fischer, U., McDonnell, L., & Orasanu, J. (2007). Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. *Aviation, Space, and Environment Medicine*, *78*(1), B86–B95.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley.

Flin, R., O'Connor, P., & Mearns, K. (2002). Crew resource management: Improving team work in high reliability industries. *Team Performance Management: An International Journal*, *8*(3/4), 68–78. doi:10.1108/13527590210433366

Gillan, C. A. (2003). Analysis of multicrew decision making from a cognitive perspective. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 427–432). Dayton, OH: Wright State University.

Gontar, P., & Hoermann, H.-J. (2014). Flight crew performance and CRM ratings based on three different perceptions. In A. Droog (Ed.), *Aviation psychology: Facilitating change(s): Proceedings of the 31st EAAP Conference* (pp. 310–316). Amsterdam, The Netherlands: European Association for Aviation Psychology.

Gontar, P., & Hoermann, H.-J. (2015a). Interrater reliability at the top end: Measures of pilots' nontechnical performance. *International Journal of Aviation Psychology*, *25*(3–4), 171–190. doi:10.1080/10508414.2015.1162636

Gontar, P., & Hoermann, H.-J. (2015b). 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., Hoermann, H.-J., Deischl, J., & Haslbeck, A. (2014). How pilots assess their non-technical performance: A flight simulator study. In N. A. Stanton, S. J. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in human aspects of transportation: Part I* (pp. 119–128). Danvers, MA: AHFE International.

Gontar, P., & Mulligan, J. B. (2016). Cross recurrence analysis as a measure of pilots' coordination strategy. In A. Droog, M. Schwarz, & R. Schmidt (Eds.), *Proceedings of the 32nd Conference of the European Association for Aviation Psychology* (pp. 524–544). Amsterdam, The Netherlands: European Association for Aviation Psychology.

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*. Geneva, Switzerland: International Ergonomics Association.

Gontar, P., Schneider, S. A. E., Schmidt-Moll, C., Bollin, C., & Bengler, K. (2017). *Hate to interrupt you, but . . . Analyzing turn-arounds from a cockpit perspective*. Manuscript submitted for publication.

Gorman, J. C., & Cooke, N. J. (2011). Changes in team cognition after a retention interval: The benefits of mixing it up. *Journal of Experimental Psychology Applied*, *17*(4), 303–319. doi:10.1037/a0025149

Gorman, J. C., Cooke, N. J., Amazeen, P. G., & Fouse, S. (2012). Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(4), 503–517. doi:10.1177/0018720811426140

Haslbeck, A., & Hoermann, H.-J. (2016). Flying the needles: Flight deck automation erodes fine-motor flying skills among airline pilots. *Human Factors*, *58*(4), 533–545. doi:10.1177/0018720816640394

Helmreich, R. L., & Foushee, H. C. (2010). Why CRM? Empirical and theoretical bases of human factors training. In B. G. Kanki, R. L. Helmreich, & J. M. Anca (Eds.), *Crew resource management* (2nd ed., pp. 3–57). Amsterdam, The Netherlands: Academic Press.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *International Journal of Aviation Psychology*, *9*(1), 19–32. doi:10.1207/s15327108ijap0901_2

Helmreich, R. L., & Sexton, B. J. (2004). Group interaction under threat and high workload. In R. Dietrich & T. M. Childress (Eds.), *Group interaction in high risk environments* (pp. 9–23). Aldershot, UK: Ashgate.

Holt, R. W., Hansberger, J. T., & Boehm-Davis, D. A. (2002). Improving rater calibration in aviation: A case study. *International Journal of Aviation Psychology*, *12*(3), 305–330. doi:10.1207/S15327108IJAP1203_7

Hutchins, E., Weibel, N., Emmenegger, C., Fouse, A., & Holder, B. (2013). An integrative approach to understanding flight crew activity. *Journal of Cognitive Engineering and Decision Making*, *7*(4), 353–376. doi:10.1177/1555343413495547

Kiekel, P. A., Cooke, N. J., Foltz, P. W., Gorman, J. C., & Martin, M. J. (2002). Some promising results of communication-based automatic measures of team cognition. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(3), 298–302. doi:10.1177/154193120204600318

Kiekel, P. A., Gorman, J. C., & Cooke, N. J. (2004). Measuring speech flow of co-located and distributed command and control teams during a communication channel glitch. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 683–687. doi:10.1177/154193120404800387

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: ISAP.

Kozlowski, S. W. J., Gully, S. M., Nason, E. R., & Smith, E. M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. In D. R. Ilgen & E. D. Pulakos (Eds.), *The changing nature of performance: Implications for staffing, motivation and development* (pp. 240–292). San Francisco, CA: Jossey-Bass.

Krifka, M., Martens, S., & Schwarz, F. (2004). Linguistic factors. In R. Dietrich & T. M. Childress (Eds.), *Group interaction in high risk environments* (pp. 75–85). Aldershot, UK: Ashgate.

Lacity, M. C., & Janson, M. A. (1994). Understanding qualitative data: A framework of text analysis Methods. *Journal of Management Information Systems*, *11*(2), 137–155. doi:10.1080/07421222.1994.11518043

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

Mjos, K. (2001). Communication and operational failures in the cockpit. *Human Factors and Aerospace Safety*, *1*(4), 323–340.

Mosier, K. L., & Fischer, U. M. (2010). Judgment and decision making by individuals and teams: Issues, models, and applications. *Reviews of Human Factors and Ergonomics*, *6*(1), 198–256. doi:10.1518/155723410X12849346788822

Nevile, M. (2007). Talking without overlap in the airline cockpit: Precision timing at work. *Text & Talk—An Interdisciplinary Journal of Language, Discourse Communication Studies*, *27*(2), 225–249. doi:10.1515/TEXT.2007.009

Nonose, K., Kanno, T., & Furuta, K. (2015). An evaluation method of team communication based on a task flow analysis. *Cognition, Technology & Work*, *17*(4), 607–618. doi:10.1007/s10111-015-0340-4

O'Connor, P., Hoermann, H.-J., Flin, R., Lodge, M., & Goeters, K.-M. (2002). Developing a method for evaluating crew resource management skills: A European perspective. *International Journal of Aviation Psychology*, *12*(3), 263–285. doi:10.1207/S15327108IJAP1203_5

Orasanu, J. (1990). *Shared mental models and crew decision making* (CSL Technical Report No. 46). Princeton, NJ: Princeton University.

Orasanu, J., & Fischer, U. (1992). Team cognition in the cockpit: Linguistic control of shared problem solving. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 189–194). Hillsdale, NJ: Erlbaum.

Orasanu, J., Fischer, U., & Davison, J. (1997). Cross-cultural barriers to effective communication in aviation. In S. Oskamp & C. Granrose (Eds.), *Cross-cultural work groups: The Claremont Symposium on Applied Social Psychology* (pp. 1–23). Thousand Oaks, CA: SAGE.

Reichling, J. (2017). *Analysis of pilots' coordination strategies and their influence on performance* (Master thesis). Technical University of Munich, Garching, Germany.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, *50*(3), 540–547. doi:10.1518/001872008X288457

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small Group Research*, *36*(5), 555–599. doi:10.1177/1046496405277134

Sassen, C. (2005). *Linguistic dimensions of crisis talk: Formalising structures in a controlled language. Pragmatics and beyond*. Amsterdam, The Netherlands: Benjamins.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Reprinted). Cambridge, UK: Cambridge University Press.

Serfaty, D., Entin, E. E., & Johnston, J. H. (1998). Team coordination training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 221–245). Washington, DC: American Psychological Association.

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: The links between language, performance, error, and workload. *Journal of the Society for Human Performance in Extreme Environments*, *5*(1), 63–68.

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Human Factors*, *52*(2), 234–245. doi:10.1177/0018720809350882

Sperling, B. K. (2006). Information distribution and team situational awareness: An experimental study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 477–481. doi:10.1177/154193120605000356

Weber, D. E., Mavin, T. J., Roth, W. M., Henriqson, E., & Dekker, S. W. A. (2014). Exploring the use of categories in the assessment of airline pilots' performance as a potential source of examiners' disagreement. *Journal of Cognitive Engineering and Decision Making*, *8*(3), 248–264. doi:10.1177/1555343414532813

Weigl, M., Antoniadis, S., Chiapponi, C., Bruns, C., & Sevdalis, N. (2015). The impact of intra-operative interruptions on surgeons' perceived workload: An observational study in elective general and orthopedic surgery. *Surgical Endoscopy*, *29*(1), 145–153. doi:10.1007/s00464-014-3668-6

Yule, S., Rowley, D., Flin, R., Maran, N., Youngson, G., Duncan, J., & Paterson-Brown, S. (2009). Experience matters: Comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ Journal of Surgery*, *79*(3), 154–160. doi:10.1111/j.1445-2197.2008.04833.x

Zbilut, J. P., Thomasson, N., & Webber, C. L. (2002). Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Medical Engineering & Physics*, *24*(1), 53–60. doi:10.1016/S1350-4533(01)00112-6

Patrick Gontar studied mechanical engineering at Technical University of Munich with specializations in aeronautics and human factors. He received his diploma with distinction in 2014. During his time at the institute, he conducted research in different fields of flight safety, such as effects of cyber threats or unforeseen events during flight missions. His research interests include visual attention and verbal communication of pilot teams. He is currently working on his PhD thesis, supervised by Klaus Bengler.

Ute Fischer is a research scientist in the School of Literature, Communication, and Culture at the

Georgia Institute of Technology. After receiving her PhD in cognitive psychology from Princeton University, she was a postdoctoral fellow and then a senior research scientist at NASA Ames Research Center. Her research addresses aspects of individual and team decision making in complex, high-technology environments—in particular, factors affecting practitioners' risk assessment, communication, and decision strategies.

Klaus Bengler graduated in psychology at the University of Regensburg and received his PhD in 1995 in cooperation with BMW at the Institute of Psychology. Since 2009 he has been the head of the Institute of Ergonomics at the Technical University of Munich. The research areas of the institute include flight safety, digital human modeling, human-robot cooperation, highly automated driving, and human reliability.