A Speech-Based Model for Tracking the Progression of Activities in Extreme Action Teamwork

Swathi Jagannath, Drexel University, USA
Neha Kamireddi, Drexel University, USA
Katherine Ann Zellner, Drexel University, USA
Randall S. Burd, Children's National Hospital, USA
Ivan Marsic, Rutgers University, USA
Aleksandra Sarcevic, Drexel University, USA

Designing computerized approaches to support complex teamwork requires an understanding of how activity-related information is relayed among team members. In this paper, we focus on verbal communication and describe a speech-based model that we developed for tracking activity progression during time-critical teamwork. We situated our study in the emergency medical domain of trauma resuscitation and transcribed speech from 104 audio recordings of actual resuscitations. Using the transcripts, we first studied the nature of speech during 34 clinically relevant activities. From this analysis, we identified 11 communicative events across three different stages of activity performance—before, during, and after. For each activity, we created sequential ordering of the communicative events using the concept of narrative schemas. The final speech-based model emerged by extracting and aggregating generalized aspects of the 34 schemas. We evaluated the model performance by using 17 new transcripts and found that the model reliably recognized an activity stage in 98% of activity-related conversation instances. We conclude by discussing these results, their implications for designing computerized approaches that support complex teamwork, and their generalizability to other safety-critical domains.

CCS Concepts: • **Human-centered computing~Human computer interaction (HCI)~HCI design and evaluation methods~User models •Human-centered computing~Interaction design~Interaction design process and methods~Activity centered design**

Additional Key Words and Phrases: Activity recognition; narrative schemas; speech modeling.

ACM Reference format:

Swathi Jagannath, Neha Kamireddi, Katherine Ann Zellner, Randall S. Burd, Ivan Marsic, and Aleksandra Sarcevic. 2022. A Speech-Based Model for Tracking the Progression of Activities in Extreme Action Teamwork. Proc. ACM Hum.-Comput. Interact, 6, CSCW1, Article 73 (April 2022), 25 pages, https://doi.org/10.1145/3512920.

1 INTRODUCTION

Collaborative, time-critical teamwork requires rapid decisions and timely task completion [2],[12],[17],[21],[22]. Workers must also monitor each other and coordinate tasks, which

This work is supported by the U.S. National Science Foundation, under grant IIS-1763509.
Author's addresses: S. Jagannath, N. Kamireddi, K. A. Zellner, and A. Sarcevic, College of Computing and Informatics, Drexel University, 3675 Market St., Philadelphia, PA 19104, USA; R. S. Burd, Trauma and Burn Surgery, Children's National Hospital, 111 Michigan Avenue NW, Washington, DC, 20010, USA; I. Marsic, Electircal and Computer Engineering, Rutgers University, 94 Brett Road, Piscataway, NJ 08854, USA.

https://doi.org/3512920

increases the mental workload [9],[28],[33],[50],[62]. Although the perception of time is critical for timely performance, it is often skewed, leading to activity delays [30],[36]. For example, tasks during team-based emergency medical scenarios, such as trauma resuscitation, are often delayed [63] despite training [23], protocols [1], and low-tech solutions like checklists [43]. Adding a team member to monitor temporal aspects of work is unfeasible because (1) this task is cognitively demanding [24], and (2) this role would require costly domain expertise while remaining prone to the same timeliness errors as others over time. Several systems have been introduced to support temporal awareness in time- and safety-critical settings [14],[15],[39],[41] but were found intrusive because they relied on manual data entry. Increasing awareness of delays during extreme action teamwork requires an approach that can automatically and unobtrusively monitor the progression of activity and provide real-time data about the team's status.

An automated approach that supports complex situations typically relies on machine learning (ML) to account for the uncertainty and flexibility of the process. In healthcare, the availability of high-power computing and massive datasets has led to an increase in applying ML to disease risk prediction [42],[49]. Other domains, such as weather forecasting have long used ML to detect patterns and make predictions from large datasets [54]. However, this neural-network-based ML approach, where a model learns from raw data, is not feasible in complex teamwork for two related reasons. First, large video- or audio datasets about team activities are not publicly available for neural network training. Second, obtaining large datasets of teamwork is challenging. Events involving knowledge-based, time-critical teamwork occur at unpredictable times and frequencies. Ground-truth coding of these events requires costly domain expertise, takes time, and may be subject to privacy constraints. To compensate for small-size datasets, a common strategy has been to augment neural-network-based ML with heuristics extracted from domain knowledge [52].

In this paper, we determine speech-based heuristics for modeling the temporal progression of activity performance based on team conversations in an extreme action team setting. We use *trauma resuscitation teamwork* as an example of this setting, where highly skilled members cooperate on urgent and unpredictable tasks [28]. During trauma resuscitation, team members work together to rapidly identify and treat life-threatening injuries. Despite established protocols that guide this extreme action teamwork, modeling activity progression is not straightforward. Trauma team members sometimes plan activities but pause or abandon them before completion. Activities may also be attempted several times, repeated, or skipped. While model heuristics can be derived from events observed in different modalities (e.g., visual or touch-based, artifact use), our focus is on speech-based heuristics. We *hypothesize* that verbal communication during team-based activities contains sufficient information for recognizing the stage of activity performance (before, during, and after). This hypothesis was based on prior research showing that speech in time- and safety-critical team-based work is rich with information about task parameters and coordination [22],[24],[53],[65]. Identifying the activity stage is critical for detecting activity delays because the stage indicates whether an activity was only considered (based on the "before" speech) or also performed and completed (based on the "during" and "after" speech).

We tested our research hypothesis in three steps. First, we used 104 transcripts from actual resuscitations and analyzed team conversations during 2,414 activity performances. Using these analyses, we identified 11 types of "communicative events" and their order of occurrence across three stages of activity performance. Second, we developed speech-flows for 34 clinically relevant activities to represent the sequence of activity-related conversations. We adopted the framework of *narrative schemas* from statistical language modeling because it has been successfully used before for representing speech during activity performance [46]. The common features from the 34 narrative schemas were then aggregated to develop a general speech-based model for tracking the activity progression over time. Finally, we evaluated the model using 17 new transcripts with conversations during 451 activity performances. The model successfully identified activity stages in 98% of the conversations. Of 451 conversations, 24 (5%) contained speech about all three stages, 136 (30%) about any combination of two stages, and 284 (63%) about one stage (mostly "after").

The model failed to recognize any stage in seven (2%) conversations. These results confirmed our hypothesis, while also suggesting that heuristics based on other modalities may be needed for recognizing the stages where speech is absent or offers weak cues. Our results also suggest that a system trained on our model could detect an activity delay because it can reliably detect activity completion based on team conversations during the *after* stage. Detecting the *after* stage is critical because it indicates that an activity is completed. If activity completion had not been detected within the expected time, the system can conclude it is delayed. Detecting the *before* and *during* stages is still important to improve the system accuracy and reduce false alarms.

With this work, we make two contributions. First, we contribute two types of speech-based heuristics for modeling the temporal progression of activity in extreme action teamwork: (a) 11 types of communicative events that indicate activity planning, execution, and evaluation, and (b) sequential ordering of these events that form the narrative schema for each of the 34 activities that we considered. Second, we contribute a speech-based model for identifying the stage of activity progression, along with the empirical evidence of how often the model can identify activity stages. A model that can identify the stage of activity progression based on verbal communication can be valuable for any type of teamwork analysis, for designing different types of team decision-support systems, as well as for designing a general system that supports collaborative work.

2 RELATED WORK

We review prior work from three areas of research: (1) communication as a coordination mechanism in extreme action team settings, (2) micro-level analysis of communication, and (3) speech-based modeling and its application in existing activity recognition systems.

2.1 Communication as a Coordination Mechanism in Extreme Action Team Settings

Studies of extreme action teamwork in traffic control rooms [4],[22], airline cockpits [18],[25], emergency response [6],[31],[45], and other "centers of coordination" [58] have shown how team members maintain constant communication to achieve coordination. For example, assignment and management of activities in the London Underground control room are primarily achieved through explicit communication among traffic controllers [22]. Team members also use implicit coordination mechanisms to trigger action, like overhearing conversations or talking out loud [5]. Similarly, speech, gesture, and space are being combined in an airline cockpit to support a shared mental model among team roles [25]. Analyses of situated action in medical scenarios have also found the combined use of speech, gesture, and movement to establish a shared mental model and coordinate activities [20],[51],[53],[65]. For instance, trauma team members constantly monitor each other's talk, gestures, and body movement to make sense of actions and anticipate requests [65]. Although gesture and movement communicate actions, the crowded nature of dynamic work makes it challenging to closely monitor and react to parallel activities [65]. In contrast, speech does not require visual attention, offering an alternative approach for increasing temporal awareness through verbal sharing of activity progress and results.

This prior work offers rich accounts of how embodied action interacts with speech and how team members achieve awareness through spoken interactions in dynamic team processes. Our study extends this line of research by performing a micro-analysis of speech alone, focusing on team member utterances and discussions of activity planning, execution, and evaluation. Through this analysis, we contribute empirical evidence of not only *how* but also *how frequently* teams discuss an activity and its progress. These findings provide new insight into how speech-based heuristics can be used for training ML algorithms to predict activity stages in domains where small datasets make it challenging for ML models to learn from raw data.

2.2 Micro-level Analysis of Verbal Communication

CSCW researchers have long studied how speech shapes and reflects actions during cooperative work [16],[22],[25],[57]. Early work on the Coordinator system, for example, identified four linguistic actions of teamwork (requests, promises, assertions, and declarations), illustrating how teams managed action through conversations [16]. Similarly, Suchman and colleagues [57],[59] derived three forms of communication for achieving coherent teamwork: substantive exchanges about the main topic, annotative exchanges (e.g., questions, clarifications), and procedural exchanges (e.g., transitions to other topics). These early classifications were critical for designing CSCW systems that supported asynchronous communication. However, they are too coarse for systems that use synchronous communication for detecting activities and stages.

Other areas of research, such as interactional sociolinguistics and its analytic methods (e.g., discourse analysis) also provide frameworks for understanding how speech informs social interaction [19]. In CSCW, discourse analysis has been used to analyze small-group conversations during the collaborative creation of drawings [56] and to determine the effects of interface affordances and multimedia content on computer-mediated small-group discussions [60],[61]. Similarly, conversation analysis has been used to unpack the situated practice of using intelligent personal assistants (IPAs) in multi-party conversations [48]. While offering rich insight into the practices of meaning-making in communication, these frameworks do not apply to our work for two reasons: (1) they focus on the form and meaning, rather than on communicative functions of speech, and (2) using these frameworks would result in heuristics that are too complex for augmenting ML models that predict activity stages. We needed a framework that could help us determine what speakers wanted to accomplish with an utterance, so we could model the associations between speech and activity progression.

Several CSCW studies performed similar analyses when studying decision-making among different groups [11],[13],[35],[53]. For example, Feng and Mentis [13] adopted 12 dialogue acts from the conversational games framework [29] to analyze dyadic knowledge sharing during surgical procedures. In another study, Mentis et al. [35] used 11 codes from the rhetorical structural theory [32] to understand rationale development in a complex group decision-making task. Sarcevic et al. [53] analyzed communication patterns among trauma team members through the concept of transactive memory [38], identifying eight types of verbal communication that seek and share information. We build on these current categorizations by identifying and defining the utterance types based on the stage of activity performance in which they occurred.

2.3 Speech-Based Modeling

Related work in statistical language modeling explores the use of *narrative schemas* to represent script knowledge and predict a missing or upcoming event in simple stories about daily activities [8],[37],[40],[46],[47]. This research has mostly used large text-based datasets that were generated by crowdsourcing short descriptions of daily activities [37], or by breaking Wikipedia pages [46],[47] and news text [55] into paragraphs. Each description of an activity created in this way was called a "story" because it had a clear beginning and ending, with no deviations from the topic. More recently, ML and natural language processing (NLP) have been used to explore automated story generation from large text-based datasets through deep recurrent neural networks [27],[34],[44]. As another approach, Xu et al. [58] proposed a skeleton-based model that is learned by a reinforcement learning method to first generate the story outline and then extend it with complete sentences. The ML models in these studies have been trained on the semantics of the story, including the start and end, theme or emotion, and flow of events.

Although we also model stories (i.e., specific sequences of sentences) through narrative schemas, our work differs in three ways. First, we use stories derived from actual conversations during dynamic teamwork, rather than stories generated by third parties. Second, we use speech data to infer the team's status and progression of work, rather than to describe or generate stories

Line #	Timestamp	Speaker	Speech line
20	*00:12:32*	*Team leader*	*Can you check his pupils?*
21	*00:12:45*	*Physician*	*Ok... I am going to shine light into your eyes. Look at me, don't move.*
22	*00:12:55*	*Physician*	*Pupils are 2 millimeters, reactive bilaterally.*
23	00:13:08	Physician	Moving along… I don't see any gross deformity in upper extremities. We are completely exposed. There is no obvious bleeding.
24	00:13:17	Documenter	Do we have a temperature?
25	00:13:19	Bedside nurse	Not yet.
26	*00:13:26*	*Documenter*	*Sorry, what were his pupils? 2?*
27	*00:13:27*	*Physician*	*2 millimeters.*

Figure 1: An excerpt from a transcript, showing line #, timestamp, speaker, and speech line. Italicized lines #20, 21, 22, 26, and 27 represent a single activity story about pupil examination.

about work. Finally, the stories in our dataset are often incomplete and composed of choppy, grammatically incorrect sentences, making the ML model training more challenging.

3 BACKGROUND: TRAUMA TEAMS AND TERM DEFINITIONS

Trauma resuscitations occur in a designated area in the emergency department (ED) called the trauma bay. Patients brought from the injury scene are triaged as "stat" (lower acuity), "attending" (high acuity), or "transfer" (treated at another hospital) trauma team activations. A typical trauma team consists of eight to nine members, including a surgical team leader, an emergency department attending, a surveying physician, a nurse documenter, a medication nurse, two to three bedside nurses, an anesthesiologist, and a respiratory therapist. To prioritize resuscitation activities and achieve rapid diagnoses, trauma teams follow the Advanced Trauma Life Support (ATLS) protocol [1] that consists of two parts—the primary and secondary surveys. The goal of the primary survey is to stabilize the patient by focusing on the patient's major physiological systems. During the secondary survey, the team performs a detailed head-to-toe evaluation of the patient to identify other injuries. The protocol activities are classified into two categories: *assessment* activities (information gathering and evaluating patient status) and *control* activities (taking an action to stabilize the patient based on the assessment).

Depending on patient status, the sequence of activities may vary, leading to repetitions or omissions. This acceptable variability often leads to overlapping or interleaving activity-related conversations. For example, an activity may be planned but abandoned, delayed, repeated, or suspended and then resumed later. These scenarios are usually reflected through speech because teams use verbal communication to plan activities, assign tasks, and report results. Although team members do not always announce the activity that is being performed, they discuss it as they plan (*before*), execute (*during*), and evaluate (*after*). Despite frequent discussions, speech is succinct, simple, and contains domain- and task-specific keywords and phrases [3],[26]. Keywords and phrases alone, however, do not always indicate the stage of an activity performance because the same words may occur across multiple stages. For example, the keyword "pupils" related to the pupil examination activity (Figure 1) can be heard in both the *before* (line #20, Figure 1) and *after* stages (lines #22, #26, Figure 1).

In the context of this study, we define *activity* as any action performed by team members with their hands (e.g., palpation) or eyes (e.g., observation). Depending on the activity, the speech that is accompanying that activity could be either about the activity or the activity itself. For example, when providers report the results of an activity, speech is *about* the activity. When providers ask the patient a few questions to assess their neurological status, speech *is* the activity. We use the term *story* for a sequence of sentences about an activity for which the beginning and ending can be identified. Each resuscitation contains one or more stories about a particular activity, depending on the number of times that activity was performed. Based on the content and order

Table 1: Thirty-four trauma resuscitation activities used for speech modeling.

Assessment Activities		Control Activities
Airway Assessment	Eye Examination	Intubation
Breathing Assessment	Nose Examination	Breathing Control
Blood Pressure (BP) Check	Mouth Examination	Intravenous (IV) Placement
Pulse Check	Neck Examination	Fluid Administration
Capillary Refill Check	Chest Examination	Cardiopulmonary Resuscitation (CPR)
Pupil Examination	Abdomen Examination	Cervical Spine (C-Spine) Precautions
Neurological Examination	Pelvis Examination	Exposure Control
Exposure Assessment	Genitalia Examination	Medication Administration
Oxygen Saturation Check	Back Examination	Wound Care
Heart/Pulse Rate Check	Upper Extremities Exam	
Head Examination	Lower Extremities Exam	
Face Examination	Imaging	
Ear Examination		

of sentences in the stories of a particular activity, we used the *narrative schema* [8] framework to create an abstract representation of these stories. A narrative schema shows a conceptual summary of key "communicative events" that occur across all stories associated with an activity, where each "event" roughly corresponds to a spoken sentence. In other words, a narrative schema provides a generalized workflow-type depiction of speech patterns that occur during activity.

4 METHODS

We developed the narrative schemas and speech-based model based on an empirical study that took place in a level 1 trauma center of an urban, pediatric teaching hospital in the mid-Atlantic region of the United States. We obtained approvals from the hospital's Institutional Review Board (IRB) before the study. All data generated during the study were kept confidential and secure per IRB policies and Health Insurance Portability and Accountability Act (HIPAA). To capture data, we used an always-on video and audio recording system with three cameras and microphones that was installed in the trauma bay for quality assurance purposes. The cameras were positioned at three strategic places to capture activities around the patient bed, among the leadership team, and at the room entrance. An array of high-quality shotgun microphones was also installed between the patient bed and leadership team for continuous capturing of audio data. The ongoing video and audio recording of live resuscitations was approved by the hospital's Legal and Risk Management Department. Because most patients at our research site are children, written consent was obtained from the parent or guardian before using video or audio recordings for research purposes. Assent was obtained for patients >12 years old. For trauma team members, the requirement for obtaining consent was waived under the U.S. Department of Health and Human Service regulation 45 CFR 46.116(c) because video recordings are existing data used for quality assurance. All non-medical members of the research team have been trained in the fundamentals of trauma resuscitation and have acquired the domain knowledge through participant observation and video review of live resuscitations. The first author also has experience in conducting ethnographic fieldwork in emergency departments.

4.1 Dataset and Data Collection

During the 17-month study period (January 2016 – May 2017), 653 trauma activations occurred at our research site. Of these, 190 audio recordings were available for analysis after obtaining consent. The resuscitations lasted from five to 58 minutes, with an average duration of 15 minutes. Because manual transcribing is time-consuming, we could only produce 104 transcripts. The transcripts contained an average of 162 lines of speech (SD = ±83; range: min. 41 – max. 421).

The six transcribers on our research team used a previously developed protocol to ensure consistency and preserve the structure of the discourse. The protocol included instructions for (a) timestamping speech lines to accurately reflect both linear and parallel speech, (b) marking unintelligible speech, (c) censoring sensitive information, and (d) identifying key team roles based on speech. The transcribers relied on their domain knowledge and the differences in voice tones for associating speaker roles with the uttered speech. Each transcriber followed the same steps. First, they filtered out the blank sections of the always-on audio files. Next, they listened through the entire event to remove any information that could identify the patient or a team member. They then repeatedly played the audio to transcribe all uttered speech while preserving its structure, disambiguate overlapping speech, timestamp linear and parallel speech lines, and identify speakers (Figure 1). The transcripts omitted physical interactions among team members and other annotations, including gestures and movements, because those aspects of teamwork were outside this project's scope.

To match the activities from the transcripts with those performed by trauma teams, we used an activity dictionary previously developed by medical experts on our team. This dictionary defines more than 200 resuscitation activities in the ATLS protocol and provides attributes that indicate successful or unsuccessful activity performance. We selected 34 activities (Table 1) based on (1) clinical relevance, (2) whether the speech was used to communicate activity status and progression, (3) clinical goal (assessment vs. control activity), and (4) occurrence in different phases of the protocol (primary vs. secondary survey). Among these 34 activities, 25 were *assessment* activities and the remaining nine were *control* activities.

4.2 Data Analysis

Two researchers analyzed the data in five steps. First, we processed the 104 transcripts, removing about 6% of speech lines that were marked unintelligible by the transcribers. We then performed a line-by-line analysis to associate speech lines with 34 activities (Table 1). Because trauma teams communicate in domain-specific language [26], we used keywords unique to each activity to make these associations. If we observed keywords spanning multiple activities in the same line (e.g., *"Airway is patent, breath sounds bilateral"* for the airway and breathing assessment activities), the line was assigned to more than one activity. Adjacent speech lines that had a clear start and end of conversation about a single activity were grouped into an activity-related *story*. For example, lines # 20, 21, 22, 26, and 27 in Figure 1 are all related to pupil examination and were clustered into one *story*, with each line corresponding to an *event*.

Second, we applied Sarcevic et al. [53] coding scheme to identify activity-related communication types. The eight verbal communication types required extension because they were not granular enough to identify the nuances in speech and how it related to different stages of activity performance. We used open coding to identify these nuances and add new activity-related types of communication encountered in the transcripts. To facilitate the construction of narrative schemas, we renamed communication types into "communicative events." This analysis resulted in 11 communicative events (building blocks for the narrative schemas).

Third, two researchers labeled every speech line with a communicative event, and by extension with an activity performance stage (as each of the 11 communicative events belongs to only one stage). To ensure this mapping was accurate and consistent between the researchers, we used the ground truth data generated by medical experts through a video review of 17 (out of 104, 16%) resuscitation videos. The experts marked the start and end times for each performed activity and whether that performance was successful. For these 17 cases, we aligned the timestamps of activity performance from the videos with those of transcribed speech lines, subsequently separating speech lines into stages. Using these data, we created a protocol for mapping speech lines to the appropriate stage of activity performance. For example, any speech lines associated with airway assessment that occurred before the physician talked to the patient belong to the

Table 2: Summary statistics for features of transcribed and not-transcribed cases during the study period.

Case Characteristics	Transcribed (n=104)	Not Transcribed (n=86)	*p* -value
Age (years, mean±SD)	6.7 (5.4)	6.2 (4.9)	0.76
Male (%)	76.9	64.0	0.06
Activation Type (%)			0.29
Stat	57.7	68.6	
Transfer	29.8	20.9	
Attending	12.5	10.5	
Injury Type (%)			0.59
Blunt	93.3	90.7	
Penetrating	4.8	4.7	
Other	1.9	4.7	
No pre-notification (%)	12.5	16.3	0.53
Daytime (%)	70.2	59.3	0.13
Weekend (%)	19.2	24.4	0.48

before stage; any speech lines associated with the physician's assessment of the patient's airway belong to the *during* stage; any speech lines occurring after the assessment belong to the *after* stage. We then used this protocol for assigning the communicative events based on the content of the speech line and the context within which the line occurred. For speech lines with similar content, we determined the label based on the occurrence of the speech line in relation to activity performance. The context was determined based on the lines that preceded or succeeded the target speech line in the transcript. Any uncertainties about label assignments were first discussed and then resolved between the two researchers. To further ensure consistency in labeling, a third researcher independently assigned speech lines by using ten transcripts that were aligned with the ground truth video data (9.6% of all cases). We compared the labels of the two initial researchers with those of the third and calculated a Cohen's Kappa to determine intercoder reliability. The results showed an almost perfect inter-rater reliability score of 0.82.

Fourth, we identified the characteristics of each activity so we could determine which communicative events to include in the narrative schemas. We then constructed narrative schemas for each of the 34 activities using the 11 communicative events.

Finally, we analyzed the differences and similarities across the communicative events and the overall structure of the constructed schemas. Using this analysis, we derived generalizations across the 34 narrative schemas and constructed a speech-based model that represents a synopsis of the key communicative events and their order of occurrence.

5 RESULTS

We present our findings in four parts. First, we summarize the features of selected resuscitation cases (Table 2). We then describe the 11 communicative events and how they relate to activity performance. We explain the narrative schemas that emerged from our data by providing an example schema for both the assessment and control activities. Finally, we describe our speech-based model for tracking activity progression over time and how we developed it.

5.1 Dataset Overview: Features of Selected Resuscitation Cases

To assess whether the randomly selected sample of transcribed cases was biased towards any patient or resuscitation feature, we compared them with not-transcribed cases using seven features (Table 2). A univariate analysis (Fisher's exact test) showed no significant differences between groups (Table 2). Among the 104 patients, most were male (76.9%) with a mean age of 6.7±5.4 years. Most patients were triaged as a "stat" (lower acuity) activation (57.7%) and arrived

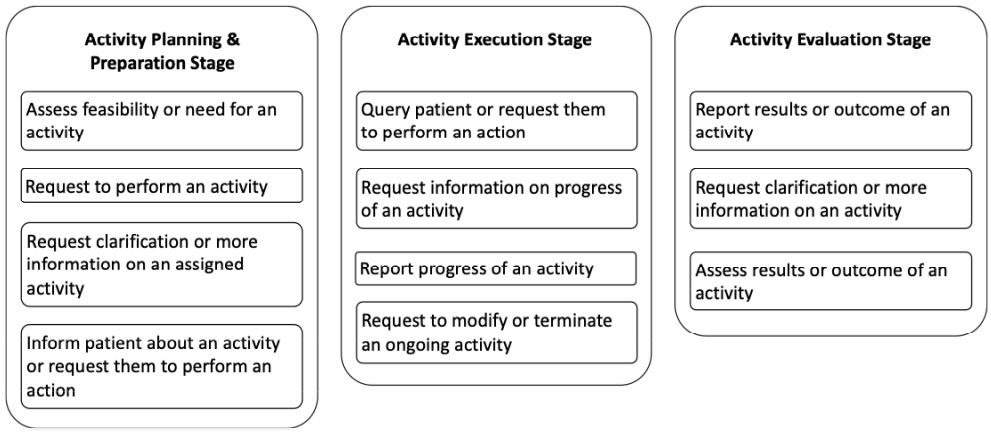

Figure 2: Types of communicative events across the three stages of activity performance.

after the team had been notified (87.5%). Resuscitations mostly occurred on weekdays (80.8%) and during the daytime (70.2%). Most patients were injured by a blunt injury mechanism (93.3%).

5.2 Communicative Events and their Occurrence in Relation to Activity Performance

Our analysis of transcripts showed that trauma team members communicated about resuscitation activities using 11 different communicative events (CEs) (Figure 2). We grouped these events based on their occurrence across the stages of activity performance: before, during, and after.

5.2.1 Before Activity Performance. We identified four communicative events in this stage:

(1) CE-1 *Assess feasibility or need for an activity:* The teams sometimes discussed the results of pre-hospital interventions to decide if a control activity is needed. For example, patients often received fluids on their way to the hospital. Upon their arrival, the trauma team discussed with the transport team (EMT) if more fluids were needed, e.g., "*Did you guys give him fluids*?" When the EMT confirmed that the patient received fluids, the leader decided "*He doesn't need more fluids, he's already got enough.*"

The team also evaluated the need for a control activity based on preceding assessments. For example, before proceeding with intubation (an airway control activity), the anesthesiologist confirmed with the leader, "*Should we intubate him now*?" The leader's response in these situations either initiated or terminated a control activity (e.g., "*Yes, go ahead and intubate*").

(2) CE-2 *Request to perform an activity*: The leader told other team members what actions to perform, either by requesting that they begin an activity or by inquiring about the activity status. For example, the leader requested the physician to begin airway assessment by stating "*Check the airway.*" For other activities, the leader asked if they performed it (e.g., "*Did you check his pupils*?"), which then prompted the physician to perform the activity.

(3) CE-3 *Request clarification or more information on an assigned activity*: When a team member wanted to clarify or confirm their task assignment, they asked for additional information or repeated what they initially heard. For example, the nurse confirmed with the leader "*Did you say normal saline*?" to ensure they administered the correct fluids.

(4) CE-4 *Inform patient about an activity or request them to perform an action:* Team members often interacted with the patient before performing an activity. For instance, the physician informed the patient "*We are going to roll you on your left side, ok*?" These announcements also signaled to the team that an activity was about to start. If an activity required the patient to perform an action (e.g., opening mouth during the mouth exam), the physician requested the action by saying, "*Could you please open your mouth*?"

5.2.2 During Activity Performance. In this stage, we identified four communicative events:

(5) CE-5 *Query patient or request them to perform an action:* For some activities, posing questions to patients to assess their status signaled the activity performance. For example, to perform the neurological exam (a three-part Glasgow Coma Score [GCS] for verbal, visual, and motor functions), the physician asked the patient, *"What is your name*? *How old are you*?" For other activities, patients were asked to move legs or arms (e.g., *"Can you move your arms*?" during the extremities exam). In contrast to physician-patient interaction before an activity, physicians in this communicative event request actions from patients *while* an activity is being performed.

(*6)* CE-6 *Request information on progress of an activity*: Requests about activity progress are phrased as questions, either to understand how long an activity will take or to decide whether and when to proceed with the next activity. For example, during the IV placement activity (insertion of an intravenous catheter for administering fluids or medications), the documenter frequently asked the bedside nurse, "*How is the access*?" to check if the IV line had been inserted.

(7) CE-7 *Report progress of an activity:* During multi-step activities that take longer to perform, team members continuously communicated about their progress. For example, while intubating a patient, the anesthesiologist stated, "*Still working on the airway.*" In another story about this activity, the physician notified the leader about the first failed attempt and announced the start of another attempt, "*Second attempt on the airway starting now.*"

(8) CE-8 *Request to modify or terminate an ongoing activity:* Team members sometimes asked each other to adjust activity performance, e.g., "*Slow down your compressions*" during cardiopulmonary resuscitation (CPR). If an ongoing activity required termination, the leader requested the performing team member to cease their operation, e.g., "*You can take the oxygen off*" to stop the oxygen flow during the breathing control activity.

5.2.3 After Activity Performance. We identified three communicative events in this stage:

(9) CE-9 *Report results of an activity:* When reporting the numerical results of an assessment activity, team members verbalized a calculated value (e.g., "*GCS is 15*"), a measured value (e.g., "*I got the blood pressure of 122 over 82*"), or a read-out value from the monitor (e.g., "*Saturation is 98 percent*"). The reports were also based on observation (e.g., "*Pupils equal and bilateral*" after inspecting the patient's eyes) or palpation (e.g., *"Abdomen is non-distended*" after pressing the patient's abdomen). When reporting the result of a control activity, team members communicated both successful and failed attempts. For example, after the first intubation attempt failed, the anesthesiologist reported, "*Can't get this tube in. I am going to try a smaller one*," communicating the restart of the activity.

(10) CE-10 *Request clarification or more information on an activity:* Requests for information in the after stage mostly came from the leader or documenter. Their requests sought (a) clarifications about reported results, (b) details missing from the initial report, and (c) a repeat of the report. For example, after the physician examined the patient's head, the documenter clarified, "*Did you say there is a laceration on her head*?" The documenter was also requesting additional information about a completed activity, "*What size tube did you put in*?" The leader sometimes clarified the results with the physician, e.g., "*Any step-offs or deformities*?"

(11) CE-11 *Assess results of an activity:* After completing an activity, the team would assess the results. If the results were not satisfactory, the team would repeat the activity. For example, after hearing the physician's report from the neurological exam, the leader commented "*Can't give him a 15. He does not seem completely oriented. Does he answer your questions*?" With this comment, the leader asked the physician to recalculate the score.

5.2.4 Summary of Communicative Events. After grouping the communicative events based on their occurrence in relation to activity stages, we found that not all events appeared in every story or always in the same order. For example, requests for information about the activity progress were not observed during short, single-step activities, such as the pupil exam. Similarly, requesting patients to perform an action was not possible when the patient was unconscious. In

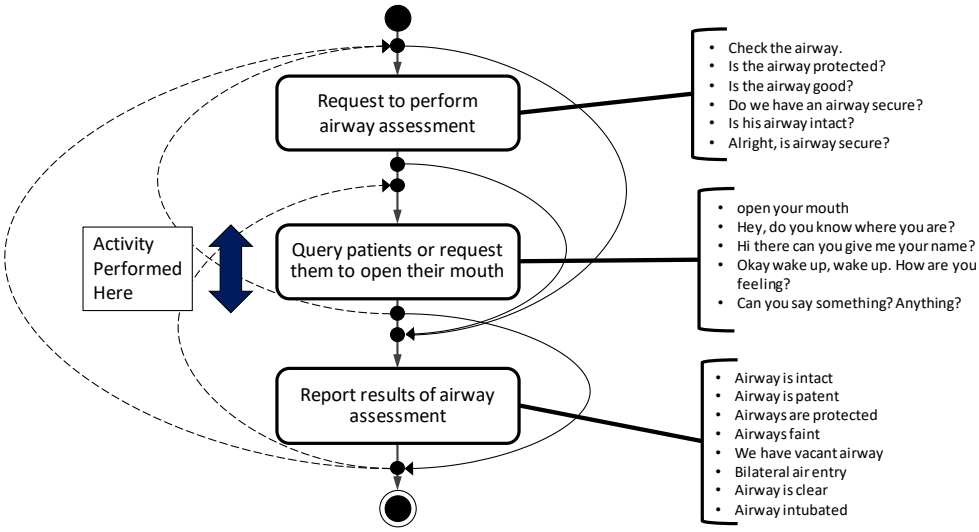

Figure 3. Narrative schema for the Airway Assessment activity. Rectangular boxes represent key communicative events. Example lines from actual speech associated with each event are shown on the right side. Arrows between the boxes indicate possible transitions between key events in a single activity-related story (solid lines for going forward and dashed lines for going back). A double-headed vertical arrow indicates the likely time of activity performance.

contrast, some communicative events appeared multiple times during a single activity story, such as multi-step reports from the neurological exam.

We also found two major syntactical differences between the 11 communicative events across three activity performance stages. First, most of the events in the before and during stages involved the use of verbs, such as "Did you check his pupils?" (before) or "I am still working on it" (during). In contrast, verbs were largely absent from the events during the after stage and were instead replaced by domain-specific keywords that indicated activity completion (e.g., "Airway is patent" in an airway assessment report). Second, two of the eight communicative events in the before and during stages exhibited similar sentence structures: *request to perform an activity* and *query patient or request them to perform an action*. Both were used to issue commands but were directed to different people in the room. These patterns suggest that distinguishing between these two events will require considering direct objects in the sentence and keywords associated with those objects. For example, in the *request to perform an activity* (e.g., "Did you check his pupils?"), *you* refers to the care provider and *his* refers to the patient. The keyword "pupils" further indicates that this request was directed to a care provider. In the *query patient or request them to perform an action* event (e.g., "Can you move your arms?"), both *you* and *your* refer to the patient. This communicative event also uses the words such as "sweetie" or "honey" to comfort the patient.

The 11 communicative events extend the previous classification [53] by focusing on activity-related speech and by including the sequence of communicated information. We reframed *directives* as *requests to perform an activity* to represent a broader set of requests through five sub-categories. *Reports* were separated into two sub-categories to better distinguish between the *during* and *after* stages. We adopted *inquiries* to indicate questions for the patient during activity performance. Inquiries from other team members became requests for information. We added a new communicative event called *assessments* and identified two sub-categories to distinguish between assessments occurring *before* and *after* the activity. *Responses* and *clarifications* were combined into *reports* because these messages contained a repeat of the results. Finally, we excluded *acknowledgments*, *message relays*, and *summons* because these types were not relevant for our activity-centered analysis.

Table 3: Number of speech lines per *primary survey assessment* activity associated with 11 communicative events in 104 transcripts. The total number of stories per activity (n) is in row #2.

		Airway Assessment	Breathing Assessment	BP Check	Pulse Check	Cap. Refill	Pupil Exam	Neuro Exam	Exposure Assessment	Oxygen Sat.	Heart/Pulse Rate Check
#	Total stories (n) in 104 transcripts	96	94	100	104	14	101	104	89	14	26
Before	Assess feasibility or need for an activity										
	Request to perform an activity	17	19	36	56		46	60	31	6	9
	Request clarification or more information on an assigned activity		12					22			
	Inform patient about an activity or request them to perform an action			7			36		6		
During	Query patient or request them to perform an action	67	54	7			25	378	34		
	Request information on progress of an activity								29		
	Report progress of an activity			24							
	Request to modify or terminate an ongoing activity										
After	Report results of an activity	129	170	192	238	14	143	369	104	14	36
	Request clarification or more information on an activity		15	30	22	1	27	51	7		3
	Assess results of an activity			6				56			

5.3 Narrative Schemas for Assessment and Control Activities

Using the 11 communicative events as the building blocks, we constructed narrative schemas for 25 assessment and nine control activities performed during trauma resuscitation (Table 1). We explain how we developed these 34 narrative schemas by describing one representative schema for each activity category: (1) airway assessment for the assessment category and (2) fluid administration for the control category.

5.3.1 Narrative Schemas for Assessment Activities. The goal of an assessment activity is to gather information about the patient and evaluate their status. Airway assessment is one of several assessment activities in the ATLS protocol. The physician performs this activity by inspecting the patient's mouth, followed by an exam of patient consciousness. We identified three key communicative events associated with airway assessment (Figure 3): (1) *request to perform airway assessment*—the leader initiates the activity by requesting the physician to begin airway assessment; (2) *query patient or request them to open their mouth*—the physician evaluates the patient's airway by asking them to open their mouth or to answer simple questions; and, (3) *report results of airway assessment*—the physician verbalizes the activity results. The actual assessment of the airway usually occurs during an event (2). Talking to the patient is skipped if the patient is unconscious and the decision to establish a definite airway is made.

The remaining 24 narrative schemas for assessment activities follow a similar pattern of communicative events and their order (Table 3 and Table 4). Specifically, for 19 other assessment activities (e.g., evaluating breath sounds, eyes, head, and abdomen), the physician directly addresses the patient. Activities such as heart/pulse rate check and oxygen saturation check use proxies (e.g., vital sign monitor) for obtaining information about the patient and do not require direct interactions with patients. We also observed that all but one assessment activity (capillary refill) started with a *request to perform an activity* event (Table 3 and Table 4). In addition, all assessment activities included *report results of an activity* at the end of the schema.

Table 4: Number of speech lines per *secondary survey assessment* activity associated with 11 communicative events in 104 transcripts. The total number of stories per activity (n) is in row #2.

	Head	Face	Ear	Eye	Nose	Mouth	Neck	Chest	Abdomen	Pelvis	Genitalia	Back	Upper Extr.	Lower Extr.	Imaging
Total stories (n) in 104 transcripts	91	83	91	29	78	96	64	94	95	86	47	100	103	99	104
Assess feasibility or need for an activity															98
Request to perform an activity	20	12	60	3	15	13	15	12	12	5	11	147	10	13	326
Request clarification or more information on an assigned activity												48			91
Inform patient about an activity or request them to perform an action	80	11	25	7		37	29	6	34	2	10	138	20	48	34
Query patient or request them to perform an action	90	36	13	5		60	201	38	67	33	13	408	246	434	16
Request information on progress of an activity	20	22	11	6	7	21	11	6	7	13	1	27	31	34	27
Report progress of an activity												176			54
Request to modify or terminate an ongoing activity							4								
Report results of an activity	163	145	171	37	103	120	82	140	137	111	56	212	221	263	59
Request clarification or more information on an activity	22	14	8	2	3	9	6	11	8	7	1	41	19	17	9
Assess results of an activity															98

Finally, we found that the *report progress of an activity* event mostly occurred in multi-step activities that took longer to complete, such as BP check, back exam, and imaging. In BP check, for example, team members first reported progress when the BP cuff was placed on the patient's arm and then again when the device started calculating the value.

5.3.2 Narrative Schemas for Control Activities. The purpose of a control activity is to manage an injury by intervening. The fluid (or bolus) administration activity belongs to the circulatory control category in the ATLS protocol. Its goal is to replace the fluids lost due to hemorrhage or other types of cardiovascular compromise. The narrative schema for fluid administration consists of four key communicative events (Figure 4): (1) *assess feasibility or need for fluid administration*—the leader decides if fluid administration is needed; if fluids are not required, the activity ends; (2) *request to start fluid administration*—if fluids are needed, the leader requests a bedside nurse to start fluid administration; (3) *request clarification or more information on fluid administration*—the bedside nurse confirms the type and amount of fluid with the leader; and, (4) *report results of fluid administration*—the bedside nurse reports the results to the team. The actual fluid administration usually starts during events (3) and (4), as indicated by the nurse's updates, e.g., "*I put him on fluids*" or "*Saline is connected.*" The activity then continues until the end of the resuscitation.

Based on the narrative schemas for the remaining eight control activities, activity performance was initiated shortly after the leader's *request to perform a control activity* (Table 5). All but one control activity (CPR) contained at least one decision-making exchange as the team was *assessing feasibility or need for an activity*, usually at the beginning. Five out of nine control activities contained *interactions with the patient* to offer comfort before (e.g., inserting a needle in the IV placement activity) or during the activity (e.g., applying wound care). For some control activities such as intubation, breathing control, or CPR, team members did not interact with the patient because the mouth and face were covered with oxygen bags, or the patient was unconscious.

Because control activities usually took longer to perform than assessment activities, we found that the *report progress of an activity* event was more common for control than assessment

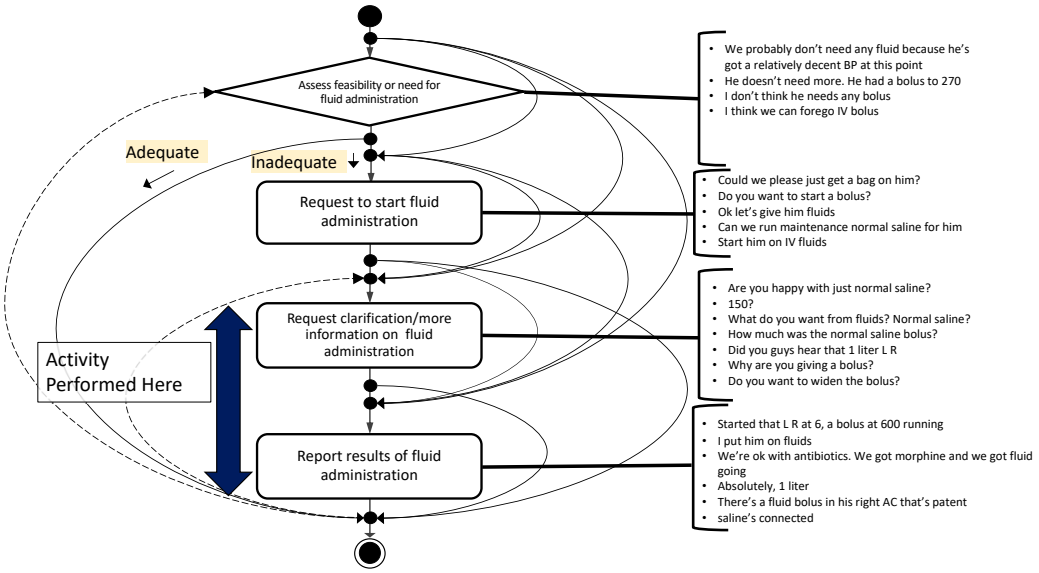

Figure 4. Narrative schema for the Fluid Administration activity. Rectangular boxes represent key communicative events. Example lines from actual speech associated with each key event are shown on the right side. Arrows between the boxes indicate possible transitions between key events in a single activity story (solid lines for going forward and dashed lines for going back). A double-headed vertical arrow indicates the likely time of activity performance.

activities. In five of the nine control activities, team members provided updates throughout the during stage of activity performance (e.g., progress on intubation or IV placement attempts).

5.4 Speech-based Model for Tracking Activity Progression in Teamwork

Our model represents a typical flow of speech among team members of an emergency medical team (Figure 5). The model consists of the generalized aspects of the 34 narrative schemas. Because the model provides the basis for ML algorithm training to recognize current activity and stage based on speech, we used the cost-benefit analysis when deciding which communicative event to include in the model. We considered factors such as the expected cost of acquiring sufficient training data to recognize each event versus the benefit of being able to differentiate between the communicative events.

The model development proceeded in several steps. We first identified the number of speech lines associated with each communicative event per activity (Table 3, Table 4, Table 5). Through this analysis of 10,166 speech lines, we determined the frequency of communicative events across all 34 activities. For example, the *report results of an activity* event was heard in 3,697 (36%) lines, while the *request to modify or terminate an ongoing activity* event was heard in 56 (0.5%) lines. These frequencies are important for model building because they show a typical amount of data available for training the ML algorithm on each communicative event. Acquiring sufficient training data for rare events would require many cases, which is costly. At the same time, the benefit of being able to recognize these events is low because their recognition will be rarely needed. Training data for safety-critical domains are already hard to obtain due to the small number of cases (e.g., ~600 resuscitations per year) and the limited availability of domain experts for data labeling.

With this challenge in mind, we excluded two communicative events with the lowest number of speech lines: *assess results of an activity* (n=66) and *request to modify or terminate an ongoing activity* (n=56). Next, we included the four events that contained the highest number of speech

Table 5: Number of speech lines per *control* activity associated with 11 communicative events in 104 transcripts. The total number of stories per activity (n) is in row #2.

	Intubation	Breathing	IV Placement	Fluid Admin.	CPR	C-Spine Prec.	Exposure Ctrl.	Medications	Wound Care
Total stories in 104 transcripts	8	64	90	24	2	73	72	69	10
Assess feasibility or need for an activity	22	20	118	27		77		113	
Request to perform an activity	11	27	50	18	3	101	110	74	4
Request clarification or more information on an assigned activity				6	10	32		19	6
Inform patient about an activity or request them to perform an action			22			46	70	46	
Query patient or request them to perform an action						66			9
Request information on progress of an activity	10	21	17				5	65	
Report progress of an activity	32		22		4	2	16	154	
Request to modify or terminate an ongoing activity		29			11	12			
Report results of an activity	53	30	61	16		41	10	54	3
Request clarification or more information on an activity	12		24			2	2	12	
Assess results of an activity		4							

lines and that were strong indicators of activity stages: *report results of an activity*, *query patient or request them to perform an action*, *request to perform an activity*, and *inform patient about activity or request them to perform an action*. We combined two of these four events that contained interactions with patients into one—*interact or request to perform actions*—because the speech lines were almost identical and the training effort to differentiate between the two is high. We also removed the word "patient" from this merged event to make the model applicable to other extreme action team settings, where the request could be directed to any actor in the setting. For the remaining five events that occurred less frequently (Figure 6), we focused on (1) their strength as predictors of activity stages and (2) the uniqueness of keywords for each activity and stage.

The *report progress of an activity* event had a total of 484 speech lines, occurring in nine out of 34 activities, mostly control and multi-step assessment activities (Table 3, Table 4, Table 5). This event was a strong indicator of the during stage because the speech lines contained verbs and keywords like "working on IV access" and "getting the tube in right now." For activities that took longer to perform, reports about progress also indicated the during stage. For this reason, we included this communicative event in the model.

The *assess feasibility or need for an activity* event had 475 speech lines, occurring in seven out of 34, mostly control activities (Table 3, Table 4, Table 5). These speech lines were in the form of questions, like "does he have an IV" and "should we intubate now?" Because responses to these questions triggered either activity performance or termination, the questions can serve as strong cues for recognizing the before stage. We, therefore, included this event in the model.

The *request information on progress of an activity* event had 391 speech lines, occurring in 21 out of 34 activities (Table 3, Table 4, Table 5). Although this communicative event occurred in many activities, the content of speech lines was similar to that of reports about activity progress. For example, a request "are you working on the access?" vs. a report "I am still working on the access;" a request "do you feel any hematoma?" vs. a report "no hematoma." Because we already included the *report progress of an activity* event in the model, a similar event with no distinctive speech patterns would be inefficient for activity stage recognition.

The *request clarification or more information on an activity* and *request clarification or more information on an assigned activity* events had 385 and 246 speech lines occurring in 28 and nine out of 34 activities, respectively (Table 3, Table 4, Table 5). These events repeated previous speech

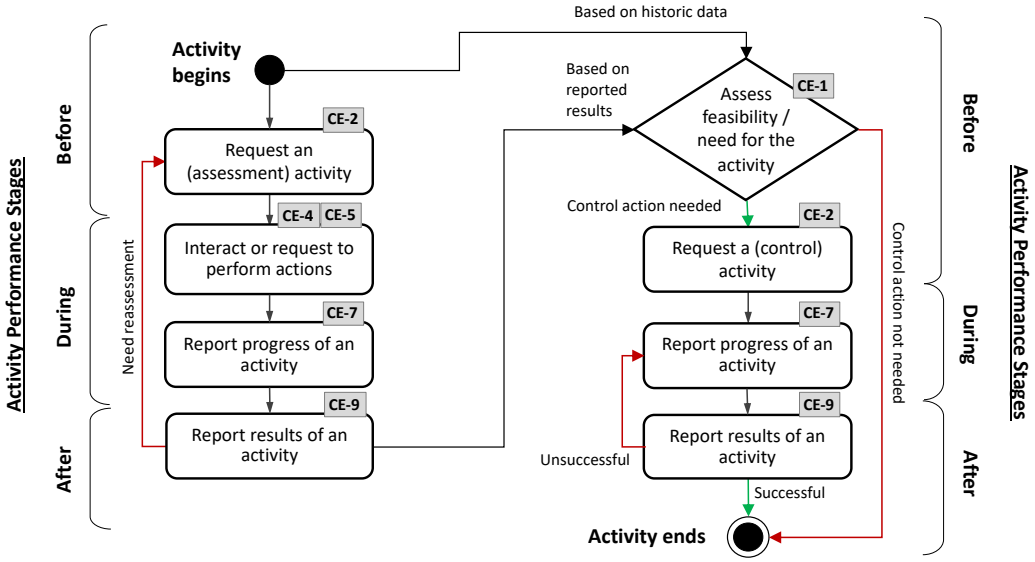

Figure 5: Model for tracking activity progression over time in extreme action teamwork.

lines from reports about results and requests for activity performance. For example, team members requested clarifications like "did you say no abrasions?" and "what was the BP again?" These communicative events may support activity stage recognition because they call for repeating the results (after stage) or for repeating the assigned activity (before stage). However, the frequency of clarifications is lower than that of reports and requests, which makes the system training more challenging. We, therefore, omitted these two events from the model.

As a result of this cost-benefit analysis, the final model includes five of the initial 11 communicative events (Figure 5, Figure 6). The left side of the model represents speech flow associated with assessment activities and the right side corresponds to control activities. Some communicative events occur in both types of activities (e.g., *report results of an activity*). To facilitate the model's generalizability to other extreme action team settings, we removed medical terms from the event labels and simplified them using broader terms.

Although our model represents a typical flow of speech associated with activity progression in an example extreme action team setting, the actual work practice may diverge from the model depending on different scenarios. For example, the team may perform several assessment activities before a control activity is executed, or a control activity may be followed by an assessment activity due to changes in the environment. Even so, the model can continue to track the progression of individual activities and be used for stage recognition and delay detection.

6 EVALUATION OF THE SPEECH-BASED MODEL

To assess the robustness of the speech-based model when applied to new scenarios, we evaluated the model using 17 previously unseen resuscitation transcripts from the April 2017 – May 2018 period with a total of 1,466 speech lines. We compared this testing dataset with the original 104 transcripts on the same seven features using Fisher's exact test, finding that all p values were insignificant (>0.05). These results confirmed equal distribution of all features in both datasets.

We prepared the testing dataset for evaluation by following the same steps we did for the modeling dataset: line-by-line analysis to associate speech lines with activities based on domain-specific keywords, grouping the speech lines into 451 activity stories based on the start and end of each conversation, and labeling each of the new 1,466 speech lines with one of the 11

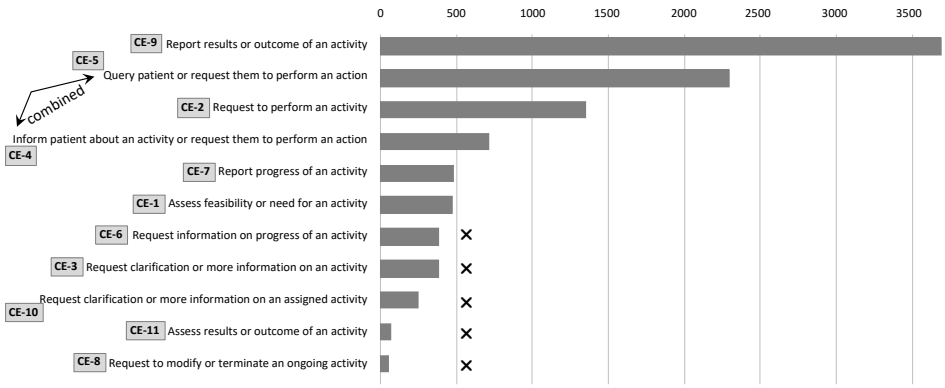

Figure 6: Frequency of speech lines for each of the 11 communicative events across 34 activities in 104 transcripts. Two events with patient interactions were merged into one for the model. Events excluded from the model are marked with an "x."

communicative events. To validate the labeling process, a second researcher independently also labeled each new speech line. After comparing the labels, the inter-rater reliability showed near-perfect agreement with a score of 0.91 (Cohen's Kappa). Grouping the speech lines into stories supported the evaluation process because each story contained one or more speech lines in each activity stage (Figure 1). Through this evaluation, we analyzed if (1) the association of speech line to communicative event remained consistent between the two datasets, thereby justifying our decision to include 6 out of 11 events in the model, and (2) the speech-based model accurately recognized the activity stages in individual stories of the testing dataset.

6.1 Evaluation of Speech Line-Communicative Event Associations Between Datasets

Our findings showed that the top four communicative events with the highest number of speech lines from the testing dataset mirrored those from the modeling dataset (Figure 6). The *report results or outcome of an activity* event remained at the top, with a total of 623 (42%) speech lines, followed by *query patients or request them to perform an activity* (335 lines, 23%), *request to perform an activity* (231 lines, 16%), and *inform patients about an activity or request them to perform an action* (93 lines, 6%) events. The bottom five events with the lowest number of speech lines that were excluded from the model (Figure 6) were also consistent with the bottom five from the testing dataset, although ranked in a different order. The number of speech lines for these bottom five events ranged from 37 to five (out of 1,466). The remaining two communicative events included in the speech-based model—*assess feasibility or need for an activity* and *report progress of an activity*—had a similar ranking in the modeling dataset but in a reversed order. This consistency in the frequency of communicative events across the two datasets validated our decisions on which events to include in the model.

6.2 Evaluation of the Robustness of the Speech-based Model

We next manually "ran" each of the 451 activity-related stories through the speech-based model to assess if the model could accurately recognize the activity performance stages. Our model successfully recognized an activity stage if at least one communicative event from that stage in the model was also present in a given story (i.e., the events in the story and the model matched). As an example, when we "ran" the pupil examination story from Figure 1 through the model, it successfully "recognized" all three stages because this story contained at least one communicative event from each stage that is also included in the model. Of the five speech lines in the story, the first occurred in the *before* stage, the second was in the *during* stage, and the last three occurred in the *after* stage (Figure 1):

Table 6: Number of stories with only speech lines relating to the three communicative events that were excluded from the model.

	BP Check	Abdomen Exam	Pelvis Exam	Breathing Control	C-spine Precautions	Nose Exam
Request clarification or more information on an activity	1	1	1			
Request to modify or terminate an ongoing activity				2	1	
Request information on progress of an activity						1

1. #20 "*Can you check his pupils*?" > REQUEST TO PERFORM AN ACTIVITY > **before** stage
2. #21 "*Ok... I am going to shine light into your eyes. Look at me, don't move.*" > QUERY PATIENTS OR REQUEST THEM TO PERFORM AN ACTION > **during** stage
3. #22 "*Pupils are 2 millimeters, reactive bilaterally.*" > REPORT RESULTS OF AN ACTIVITY > **after** stage
4. #26 "*Sorry, what were his pupils? 2*?" > REQUEST CLARIFICATION OR MORE INFORMATION ON AN ACTIVITY > **after** stage
5. #27 "*2 millimeters*" > REPORT RESULTS OF AN ACTIVITY > **after** stage

The model recognized the event *request to perform an activity* as belonging to the before stage and the event *query patients or request them to perform an action* as belonging to the during stage. For the last three speech lines, the model recognized lines #22 and #27 as belonging to the after stage because their label is also in the after stage of the model. Although the model does not contain the communicative event for line #26, it was still successful in recognizing the after stage based on the communicative event in lines #22 and #27.

6.2.1 Unsuccessful Model Runs. The model was unsuccessful in recognizing activity stages when the stories *only* contained speech lines associated with the communicative events excluded from the model. Of 451 stories, the model failed to recognize stages in seven stories (2%) because they contained speech lines associated with events that were excluded from the model (Table 6).

6.2.2 Successful Model Runs. From the remaining 444 stories, the model successfully recognized the activity stage when the story contained speech about that stage: it recognized all three activity stages in 24 (5%) stories; any combination of two activity stages in 136 (30%) stories; and one stage in 284 (63%) stories (Figure 7). These results confirmed our hypothesis, as most conversations contained sufficient information for recognizing at least one stage of activity performance. Although the model performs most reliably when an activity story contains speech lines from all three stages, activity progression was also recognized with only one or two stages. For example, 202 (45%) stories had speech lines only from the *after* stage, 53 (12%) stories had speech lines from the *before* and *after* stages, 74 (16%) stories had speech lines from the *during* and *after* stages, and 24 (5%) stories had speech from all three stages (Figure 7). In these stories, the model could not predict the start of an activity when speech data from the *before* stage was missing, but it could always predict the after *stage* and, by extension, the activity completion.

7 DISCUSSION

In this work, we determined speech-based heuristics for modeling the temporal progression of activity in extreme action teamwork. Because large datasets for training ML models about complex team activities are hard to obtain, a computerized approach for supporting this type of teamwork can use heuristics to augment machine learning. Our speech-based heuristics contain information about the (a) communication types (i.e., communicative events), (b) order of these communication types (i.e., stages within which they occur), and (c) keywords associated with specific activities and communication types. These structured representations of speech (i.e.,

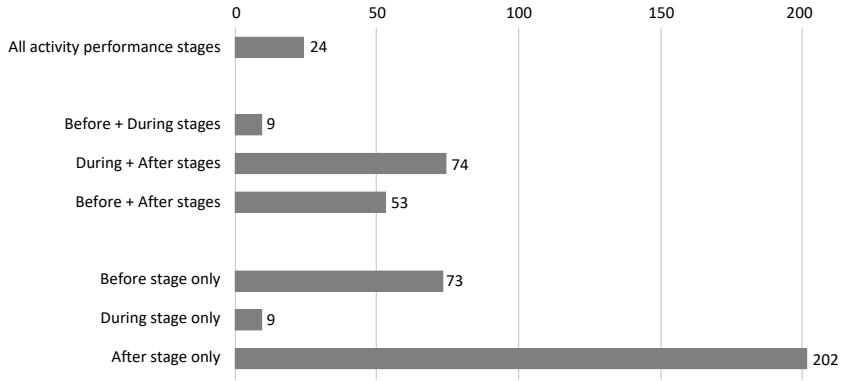

Figure 7: The number of activity-related stories (out of 451) in which the model reliably recognized some or all activity performance stages across the 17 testing transcripts.

narrative schemas) can now be used to support the modeling of the team's state by recognizing the team's current activity and the stage of that activity. Specifically, the model can guide the use of keywords for training ML algorithms to recognize the communicative events (i.e., parts of the model), the activity stage, and the activity type. We next discuss (1) the results of the model performance and their implications for designing computerized approaches that support complex teamwork; (2) how can the model be used with a neural network for machine learning, and (3) how can the model be used in other extreme action team settings.

7.1 Performance of the Model and Implications for Supporting Complex Teamwork

Most speech lines in both the modeling and testing datasets were associated with the *after* stage, as team members made sure to verbalize the results of their activities. When we grouped the speech lines in the testing dataset into activity-related stories (Figure 7), speech associated with the *after* stage was found in 353 stories (out of 444, 80%), speech about the *before* stage was found in 159 stories (36%), and about the *during* stage in 107 stories (24%). Because most stories contained speech in the *after* stage, the speech-based model detected this stage more reliably than the other two stages. The model could not detect the start of an activity in 64% of the stories, but it was still successful in detecting activity completion in 80% of the stories.

As our analysis has shown, the *after* stage is the key in detecting activity delays because speech lines from this stage indicate that activity was completed. The *before* stage implies the activity was planned, but not necessarily started. Similar uncertainties about the activity status can emerge with speech in the *during* stage. The *during* stage is also relatively short in most activities and may not guarantee successful completion. In contrast, if the model did not detect the *after* stage within the expected time (e.g., calculated relative to the start of the entire case), the system can reliably issue alerts about delays. Detecting the *before* and *during* stages is important to avoid false alarms for delayed activities that are still being performed.

Our results have several implications for designing a computerized approach for detecting delays in time-critical teamwork. First, the system design should focus on the *after* stage given its critical role in detecting activity completion. Twenty percent of the stories in our testing dataset were missing speech lines from the *after* stage. Three explanations may account for these omissions: (a) information was not communicated or it was communicated through non-verbal modalities (e.g., gesture, artifact use); (b) some lines were censored; and (c) some speech was unintelligible. While censoring should not be an issue in the real-world application—the system will be running in real time and no censoring will be needed—the other reasons suggest that other modalities may be needed to complement the lack of speech in this stage. Second, other modalities may be more suitable for detecting the *before* stage because the planning steps may be more visible

or detectable through artifact use. Additional barriers, where other modalities may better support activity and stage detection, include misalignment between activity performance and verbalization of activity status, as well as overlapping activities with parallel verbalizations from multiple team members [26].

Finally, even when speech is present, it is often incomplete and succinct. Prior work on learning models of events from large corpora included event inference [8],[46], learning a structured collection of events [7], and learning real-world situations from small corpora of events [40]. The text input in these script models consists of stories with complete, grammatically accurate sentences, which could be parsed into tuples of verb, subject, object, and prepositional relations [46]. However, a computerized approach for supporting complex teamwork will need to rely on imperfect input because it will have missing parts of speech (i.e., missing subjects or objects). The performance of the system in these cases could be improved by considering the context of adjacent sentences, as discussed next.

7.2 Using the Speech-based Model for Machine Learning

The training of the ML algorithms based on our model can proceed in two steps. In the first step, an algorithm can be trained to recognize the communicative event for each speech line. Under the assessment activity, the model contains two events with *requests* for provider activity or patient action, and two events with *reports* about activity results or progress (Figure 5). Under the control activity, the model contains two communicative events with *reports* about the activity results or progress, one *request* for activity, and one *decision* to assess the need for activity. To distinguish between these communicative events in the model, the algorithm should primarily rely on the subject, verb, and object parts of the speech line. As our analysis showed, the syntactical patterns in speech lines can support differentiating between the communicative events that belong to the same category (e.g., requests and reports), but occur in multiple stages. The presence and absence of a verb could be used to differentiate between the *report progress of an activity* (verb present in the during stage) and *report results of an activity* (verb absent in the after stage). Direct objects and associated keywords could be used to distinguish between a *request* directed to a care provider in the before stage and a *request* directed to the patient in the during stage. Once the algorithm predicts the communicative event for the speech line, it can assign the activity stage by looking up this event's location in the model. To organize inputs that the ML algorithm can use for recognizing the communicative event, researchers can adopt approaches such as those of Pichotta and Mooney's [46]: identify parts of speech for a given sentence and organize them into a fixed order, such as subject, followed by a verb, object, and other prepositional relations. A similar neural network structure can also be used to predict communicative events based on the current speech line.

The results of this first step will be in the form of probabilistic predictions for isolated speech lines. However, as we observed, not all speech during extreme action teamwork is grammatically correct, most likely leading to lower performance than that of Pichotta and Mooney's [46]. To strengthen these predictions, we may consider using the model and the context of predictions made for preceding speech lines. If the current prediction from the first step does not follow a previous prediction relative to the model, it is less likely to be correct. If it follows the previous prediction, it is more likely to be correct. This second step will result in the adjusted probabilities of activity stage predictions. In addition, a third ML algorithm can be trained to recognize the activity type, such as pupil examination or airway assessment in our case. Although the speech-based model is activity-nonspecific, it will help distinguish between assessment and control versions of activities that have both (e.g., airway management comprises both an assessment and a control activity), as they differ in one communicative event (decision).

7.2 Applying the Speech-based Model to other Extreme Action Team Settings

Although we developed our model within the context of an emergency medical domain, the model applies to other extreme action team settings because of the similarities in speech and activity patterns. To illustrate the generalizability of the model, we draw parallels between our results and those from two classic studies of safety-critical teamwork—the London underground control room [22] and the airline cockpit [25]. The line controller in the control room [22] and the captain in the cockpit [25] could be considered analogous to the trauma team leader in our study.

Our findings showed that the trauma team leader explicitly assigned tasks to other team members, which was also observed in the airline cockpit, but not always in the control room. For example, the controller's discussion with the driver to delay the train implicitly became a task for the Divisional Information Assistant (DIA) to announce this delay to passengers. Both trauma team leaders and line controllers issued requests to terminate or modify an activity, e.g., the controller requested the driver to reverse the train or wait until further instructions. In the airline cockpit, modifications to an ongoing activity were more implicit due to physical arrangements. When the second officer (SO) announced a potential fuel leak, the captain and first officer (FO) turned around, suggesting they could help even though they did not receive an explicit request.

All three domains also relied on "self-talk" [22] to make the status of the actions more visible. The controller, for example, announced the completion of his changes to the timetable, which prompted the DIA to announce those changes to the passengers. Similarly, bedside physicians called out the results of their activity performance to keep others aware of the activity progress and completion. Descriptions of the activity progress in the airline cockpit were also in the form of continuous conversations among team members. Through phrases such as "*you see, right now*" or "*but we're still*," their verbalizations indicated the temporal relationships among their actions, and by extension, the stages of their activities.

These parallels show that team activities in other safety-critical team settings also exhibit the same three stages of activity performance: preparation, execution, and assessment. Similar to our results, explicit verbalizations of these stages may sometimes be missing or communicated through other modalities. While applying our model to different domains may require extracting keywords that are specific to that domain, we expect that the communicative events and their ordering in the narrative schemas will remain the same or be only slightly modified because of the same activity pattern (preparation, execution, and assessment).

Another generalizable aspect of our work is in the approach that we used to extract speech-based heuristics and model the progression of team activity based on those heuristics. As other domains may need to extract heuristics specific to their work, they could apply our approach to analyzing team actions or conversations: identify events that indicate the activity stage, order them based on their occurrence, and aggregate them into a generalized model. A model that can determine the stage of activity progression using verbal communication can be valuable for any type of teamwork analysis, for designing different types of team decision-support systems, and for designing a general system that supports collaborative work.

8 CONCLUSION, STUDY LIMITATIONS, AND FUTURE WORK

Verbal communication among the members of extreme action teams contains rich information about task assignments, work coordination, progress, and results, providing an insight into the team's overall state. These verbalizations of activity progress offer an alternative approach for increasing temporal awareness in time- and safety-critical settings, where the crowded nature of teamwork makes it challenging to visually monitor multiple activities. In this work, we set out to determine speech-based heuristics for modeling the temporal progression of activity. We hypothesized that verbal communication during team-based activities contains sufficient information for recognizing the stage of activity performance (before, during, and after). To test this hypothesis, we analyzed 10,166 speech lines from 104 actual trauma resuscitations and

identified 11 types of communicative events and their order of occurrence in the process. Using these results, we developed narrative schemas for 34 assessment and control activities, and then aggregated those schemas into an activity-nonspecific speech-based model for tracking the progression of individual activities over time. Evaluation of the model performed on a testing dataset showed that the model reliably recognized at least one activity stage in 98% of the activity stories. This result confirmed our hypothesis, while also suggesting the need for other modalities to improve the detection of stages when speech is missing or offering weak cues. Our analyses of team conversations and their modeling also helped shine the light on the prevalent "black box" of many current ML systems, as we now show how the model works and what information it uses to predict activity stages.

Our study has two limitations. First, due to the limited availability of medical experts on our team, we were unable to generate video-based ground truth data for all transcripts in our dataset. Even so, close to 10% of our transcripts had the ground truth data available, which supported our inter-rater reliability assessments and ensured consistency in associating speech lines with the activity performance stages. Second, we evaluated the speech-based model with the testing data from the same domain. However, based on our discussion of parallels in communicative functions of speech across three safety-critical domains, we expect that both the model and communicative events can generalize to other speech-heavy extreme action team settings because they are not specific to activities.

We already begun developing the ML algorithm for activity stage recognition based on the speech-based heuristics presented in this paper. Our next steps will focus on determining the thresholds for activity delays so we can predict delays based on the detected activity stage and its typical execution time relative to other activities in the workflow. Given the challenges in implementing algorithms for speech-based activity recognition, we will complement speech with other modalities and sensors in the environment. By combining multiple modalities, our overall research goal is to develop a robust computerized approach for tracking the team's state and alerting them to potentially delayed critical activities.

ACKNOWLEDGMENTS

We thank the medical and research staff at Children's National Hospital for their support. This research has been funded by the National Science Foundation under grant number IIS-1763509.

REFERENCES

[1] American College of Surgeons, Advanced Trauma Life Support®(ATLS®), 7th Edition, Chicago, IL, 2005.

[2] Jakob E. Bardram 2000. Temporal coordination–on time and coordination of collaborative activities at a surgical department. *Computer Supported Cooperative Work* (CSCW) 9, 2, 157-187.

[3] Engelbert A.G. Bergs, Frans L.P.A Rutten, Tamer Tadros, Pieta Krijnen, and Inger B. Schipper. 2005. Communication during trauma resuscitation: do we know what is happening? *Injury* 36, 8, 905-911. DOI: https://doi.org/10.1016/j.injury.2004.12.047

[4] Johan Berndtsson, and Maria Normark. 1999. The coordinative functions of flight strips: air traffic control work revisited. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, ACM, 101-110.

[5] Claus Bossen. 2002. The parameters of common information spaces: The heterogeneity of cooperative work at a hospital ward. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '02)*, ACM, 176-185.

[6] John Bowers, and David Martin. 1999. Informing collaborative information visualisation through an ethnography of ambulance control. In *Proceedings of the 1999 European Conference on Computer-Supported Cooperative Work (ECSCW '99)*, 311-330.

[7] Nathanael Chambers. 2013. Event schema induction with a probabilistic entity-driven model. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1797-1807.

[8] Nathanael Chambers, and Daniel Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2, 602-610.

[9] Yu-Hern Chang, and Chung-Hsing Yeh. 2010. Human performance interfaces in air traffic control. *Applied ergonomics* 41, 1, 123-129.

[10] Snigdha Chaturvedi, Haoruo Peng, and Dan Roth. 2017. Story comprehension for predicting what happens next. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1603-1614.

[11] Gregorio Convertino, Helena M. Mentis, Aleksandra Slavkovic, Mary Beth Rosson, and John M. Carroll. 2011. Supporting common ground and awareness in emergency management planning: A design research project. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 4, 1-34.

[12] Cleidson RB De Souza, and David F. Redmiles. 2011. The awareness network, to whom should I display my actions? and, whose actions should I monitor? *IEEE Transactions on Software Engineering* 37, 3, 325-340.

[13] Yuanyuan Feng, and Helena M. Mentis. 2016. Supporting common ground development in the operation room through information display systems. In *AMIA Annual Symposium Proceedings*, 1774–1783.

[14] Javier Martínez Fernández, Juan Carlos Augusto, Ralf Seepold, and Natividad Martínez Madrid. 2010. Why traders need ambient intelligence. In *Ambient Intelligence and Future Trends-International Symposium on Ambient Intelligence (ISAmI 2010)*, Springer, Berlin, Heidelberg, 229-236.

[15] Mark Fitzgerald, Peter Cameron, Colin Mackenzie, Nathan Farrow, Pamela Scicluna, Robert Gocentas, Adam Bystrzycki et al. 2011. Trauma resuscitation errors and computer-assisted decision support. *Archives of Surgery* 146, 2, 218-225.

[16] Fernando Flores, Michael Graves, Brad Hartfield, and Terry Winograd. 1988. Computer systems and the design of organizational interaction. *ACM Transactions on Information Systems (TOIS)* 6, 2, 153-172.

[17] David M. Gaba, Steven K. Howard, and Stephen D. Small. 1995. Situation awareness in anesthesiology. *Human Factors* 37, 1, 20-31.

[18] Charles Goodwin, and Marjorie Harness Goodwin. 1996. Seeing as situated activity: Formulating planes. *Cognition and Communication at Work*, 61-95.

[19] John J. Gumperz. 1982. *Discourse Strategies. Studies in Interactional Sociolinguistics 1.* Cambridge: Cambridge University Press.

[20] Maria Härgestam, Magnus Hultin, Christine Brulin, and Maritha Jacobsson. 2016. Trauma team leaders' non-verbal communication: video registration during trauma team training. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 24, 1, 37.

[21] Christian Heath, Marina Jirotka, Paul Luff, and Jon Hindmarsh. 1994. Unpacking collaboration: the interactional organisation of trading in a city dealing room. *Computer Supported Cooperative Work (CSCW)* 3, 2, 147-165.

[22] Christian Heath, and Paul Luff. 1992. Collaboration and control crisis management and multimedia technology in London Underground Line control rooms. *Computer Supported Cooperative Work (CSCW)* 1, 1-2, 69-94.

[23] Vaughan JG Higgins, Melanie J. Bryant, Elmer V. Villanueva, and Simon C. Kitto. 2013. Managing and avoiding delay in operating theatres: a qualitative, observational study. *Journal of Evaluation in Clinical Practice* 19, 1, 162-166.

[24] Edwin Hutchins. 1995. *Cognition in the Wild*, MIT Press, Cambridge, MA.

[25] Edwin Hutchins, and Leysia Palen. 1997. Constructing meaning from space, gesture, and speech. In *Discourse, Tools and Reasoning*. Springer, Berlin, Heidelberg, 23-40.

[26] Swathi Jagannath, Aleksandra Sarcevic, and Ivan Marsic. 2018. An analysis of speech as a modality for activity recognition during complex medical teamwork. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ACM, 88-97. DOI: https://doi.org/10.1145/3240925.3240941

[27] Parag Jain, Priyanka Agrawal, Abhijit Mishra, Mohak Sukhwani, Anirban Laha, and Karthik Sankaranarayanan. 2017. Story generation from sequence of independent short descriptions. *arXiv preprint arXiv:1707.05501*

[28] Katherine J. Klein, Jonathan C. Ziegert, Andrew P. Knight, and Yan Xiao. 2006. Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly* 51, 4 (2006), 590-621. DOI: https://doi.org/10.2189/asqu.51.4.590

[29] Jacqueline C. Kowtko, Stephen D. Isard, and Gwyneth M. Doherty. 1991. Conversational games within dialogue. In *Proceedings of the ESPIRIT Workshop on Discourse Coherence*.

[30] Diana S. Kusunoki, and Aleksandra Sarcevic. 2015. Designing for temporal awareness: The role of temporality in time-critical medical teamwork. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 1465-1476.

[31] Jonas Landgren. 2006. Making action visible in time-critical work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 201-210.

[32] William C. Mann, and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse* 8, 3, 243-281.

[33] Tanja Manser. 2009. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiologica Scandinavica* 53, 2, 143-151.

[34] Lara Martin, Prithviraj Ammanabrolu, Xinyu Wang, William Hancock, Shruti Singh, Brent Harrison, and Mark Riedl. 2018. Event representations for automated story generation with deep neural nets. In *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1.

[35] Helena M. Mentis, Paula M. Bach, Blaine Hoffman, Mary Beth Rosson, and John M. Carroll. 2009. Development of decision rationale in complex group decision making. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1341-1350.

[36] Susan Mohammed, Katherine Hamilton, Rachel Tesler, Vincent Mancuso, and Michael McNeese. 2015. Time for temporal team mental models: Expanding beyond "what" and "how" to incorporate "when". *European Journal of Work and Organizational Psychology* 24, 5, 693-709.

[37] Ashutosh Modi, Ivan Titov, Vera Demberg, Asad Sayeed, and Manfred Pinkal. 2017. Modeling semantic expectation: Using script knowledge for referent prediction. *Transactions of the Association for Computational Linguistics* 5, 31-44, DOI: https://doi.org/10.1162/tacl_a_00044

[38] Richard L. Moreland, Linda Argote, and Ranjani Krishnan. 1986. Socially shared cognition at work: Transactive memory and group performance. In Nye, J.; Brower, A (eds.). *What's social about social cognition? Research on socially shared cognition in small groups*. Thousand Parks, 57–84.

[39] Randall J. Mumaw, Emilie M. Roth, Kim J. Vicente, and Catherine M. Burns. 2000. There is more to monitoring a nuclear power plant than meets the eye. *Human Factors* 42, 1, 36-55.

[40] John W. Orr, Prasad Tadepalli, Janardhan R. Doppa, Xiaoli Fern, and Thomas G. Dietterich. 2014. Learning scripts as Hidden Markov Models. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI '14)*. 1565-1571.

[41] Van Paassen, M. M., Clark Borst, Rolf Klomp, Max Mulder, Pim van Leeuwen, and Martijn Mooij. 2013. Designing for shared cognition in air traffic management. *Journal of Aerospace Operations* 2, 1-2, 39-51.

[42] Yoonyoung Park, Jianying Hu, Moninder Singh, Issa Sylla, Irene Dankwa-Mullan, Eileen Koski, and Amar K. Das. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. 2021. *JAMA Network Open* 4, 4, e213909-e213909.

[43] Samantha E. Parsons, Elizabeth A. Carter, Lauren J. Waterhouse, Jennifer Fritzeen, Deirdre C. Kelleher, Karen J. O'Connell, et al. 2014. Improving ATLS performance in simulated pediatric trauma resuscitation using a checklist. *Annals of Surgery* 259, 4, 807-813.

[44] Nanyun Peng, Marjan Ghazvininejad, Jonathan May, and Kevin Knight. 2018. Towards controllable story generation. In *Proceedings of the First Workshop on Storytelling*, 43-49.

[45] Mårten Pettersson, Dave Randall, and Bo Helgeson. 2004. Ambiguities, awareness and economy: a study of emergency service work. *Computer Supported Cooperative Work (CSCW)* 13, 2, 125-154.

[46] Karl Pichotta, and Raymond J. Mooney. 2016. Learning statistical scripts with LSTM recurrent neural networks. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI '16)*. 2800-2806.

[47] Karl Pichotta, and Raymond J. Mooney. 2016. Using sentence-level LSTM language models for script inference. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL '16)*. 279-289.

[48] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. "Do animals have accents?" Talking with agents in multi-party conversation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*, ACM, 207-219. DOI: https://doi.org/10.1145/2998181.2998298

[49] Mathieu Ravaut, Vinyas Harish, Hamed Sadeghi, Kin Kwan Leung, Maksims Volkovs, Kathy Kornas, Tristan Watson, Tomi Poutanen, and Laura C. Rosella. 2021. Development and validation of a machine learning model using administrative health data to predict onset of Type 2 diabetes. *JAMA Network Open* 4, 5, e2111315-e2111315.

[50] Tom W. Reader, Rhona Flin, and Brian H. Cuthbertson. 2007. Communication skills and error in the intensive care unit. *Current Opinion in Critical Care* 13, 6, 732-736.

[51] Madhu Reddy, Paul Dourish, and Wanda Pratt. 2001. Coordinating heterogeneous work: information and representation in medical care. In *Proceedings of the 2001 European Conference on Computer-Supported Cooperative Work (ECSCW '01)*, Springer, Dordrecht, 239-258.

[52] Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch et al. 2021. Informed machine learning-A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*. DOI: https://doi.org/10.1109/TKDE.2021.3079836

[53] Aleksandra Sarcevic, Ivan Marsic, Michael E. Lesk, and Randall S. Burd. 2008. Transactive memory in trauma resuscitation. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, 215-224. DOI: https://doi.org/10.1145/1460563.1460597

[54] M. G. Schultz, Clara Betancourt, Bing Gong, Felix Kleinert, Michael Langguth, L. H. Leufen, Amirpasha Mozaffari, and Scarlet Stadtler. 2021. Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A* 379, 2194, 20200097.

[55] Dan Simonson, and Anthony Davis. 2018. Narrative schema stability in news text. In *Proceedings of the 27th International Conference on Computational Linguistics*, 3670-3680.

[56] Jaime Snyder. 2012. Let me draw you a picture: coordination in image-enabled conversation. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion (CSCW '12)*. ACM, 219–222. DOI: https://doi.org/10.1145/2141512.2141582

[57] Lucy A. Suchman. 1993. Do categories have politics? The language/action perspective reconsidered. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work*, Springer, Dordrecht, 1-14.

[58] Lucy A. Suchman. 1997. Centers of coordination: A case and some themes. In: Resnick L.B., Säljö R., Pontecorvo C., Burge B. (eds) *Discourse, Tools and Reasoning*. NATO ASI Series (Series F: Computer and Systems Sciences), vol 160. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-03362-3_3

[59] Lucy A. Suchman, and Randall H. Trigg. 1986. A framework for studying research collaboration. In *Proceedings of the 1986 ACM Conference on Computer Supported Cooperative Work (CSCW '86)*, ACM, New York, NY, 221-228. DOI: https://doi.org/10.1145/637069.637097

[60] Jozsef A. Toth. 1994. The effects of interactive graphics and text on social influence in computer-mediated small groups. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94)*. ACM, 299–310. DOI: https://doi.org/10.1145/192844.193032

[61] Yi-Chia Wang, Mahesh Joshi, and Carolyn Penstein Rosé. 2008. Investigating the effect of discussion forum interface affordances on patterns of conversational interactions. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08)*. ACM, 555–558. DOI: https://doi.org/10.1145/1460563.1460650

[62] Joel S. Warm, Raja Parasuraman, and Gerald Matthews. 2008. Vigilance requires hard mental work and is stressful. *Human factors* 50, 3, 433-441.

[63] Rachel Webman, Jennifer Fritzeen, JaeWon Yang, Grace F. Ye, Paul C. Mullan, Faisal G. Qureshi, Sarah H. Parker, Aleksandra Sarcevic, Ivan Marsic, and Randall S. Burd. 2016. Classification and team response to non-routine events occurring during pediatric trauma resuscitation. *The Journal of Trauma and Acute Care Surgery* 81, 4, 666-673.

[64] Jingjing Xu, Xuancheng Ren, Yi Zhang, Qi Zeng, Xiaoyan Cai, and Xu Sun. 2018. A skeleton-based model for promoting coherence among sentences in narrative story generation. *arXiv preprint arXiv:1808.06945*.

[65] Zhan Zhang and Aleksandra Sarcevic. 2015. Constructing awareness through speech, gesture, gaze and movement during a time-critical medical task. In *Proceedings of the 2015 European Conference on Computer-Supported Cooperative Work (ECSCW '15)*, 163-182. DOI: https://doi.org/10.1007/978-3-319-20499-4_9

Received January 2021; revised July 2021; accepted November 2021.