An Analysis of Speech during Life Saving Interventions to Inform the Design of a Computerized System for Delay Detection

Katherine Zellner, MPH[1], Louis Jiorgio Villegas[1], Charles Neff[1], Waverly Getrich-Thompson[2], Randall Burd, MD, PhD [2], Ivan Marsic, PhD[3], Aleksandra Sarcevic, PhD[1]
[1]Drexel University, Philadelphia, PA, USA [2]Childrens' National Hospital, Washington D.C., USA, [3]Rutgers University, New Brunswick, NJ

Abstract

We describe an analysis of speech during time-critical, team-based medical work and its potential to indicate process delays. We analyzed speech intention and sentence types during 39 trauma resuscitations with delays in one of three major lifesaving interventions: intravenous/intraosseous (IV/IO) line insertion, cardiopulmonary and resuscitation (CPR), and intubation. We found a significant difference in patterns of speech during delays vs. speech during non-delayed work. The speech intention during CPR delays, however, differed from the other LSIs, suggesting that context of speech must be considered. These findings will inform the design of a clinical decision support system (CDSS) that will use multiple sensor modalities to alert medical teams to delays in real time. We conclude with design implications and challenges associated with speech-based activity recognition in complex medical processes.

Introduction

Trauma resuscitation is a critical step in the early care of severely injured patients that focuses on stabilizing the patient and determining the plan of care. Patients with traumatic injuries are at an increased risk of mortality due to medical errors and process delays when compared to other hospitalized patients[1]. Medical errors and delays during trauma resuscitation contribute to nearly one half of preventable deaths[2]. Patients that require life-saving interventions (LSIs)—procedures that reduce the likelihood of death, such as intubation, blood transfusion, and cardiopulmonary and resuscitation (CPR)—are at even higher risk of mortality[3]. Timely and correct performance of these LSIs can help improve patient outcomes. Prior studies have attempted to address process delays and resuscitation errors using real-time computer-aided decision support[4,5] but their success has been limited because they either require laborious and manual data entry or rely on motion sensing that lacks the information about team activities.

Speech plays a significant role in team coordination during trauma resuscitation as clinicians communicate to request tasks, share information, or confirm completion of tasks. The content of this speech is rich with contextual information about team activities and their progression that can inform the design of a CDSS that relies on speech. Prior work has evaluated the use of speech to automatically triage patients, finding that specific keywords yield accurate classification[6]. However, team-based speech has been underexplored in the context of activity delay detection. The goal of our research is to support clinicians during trauma resuscitation by developing a multi-modal clinical decision support system (CDSS) that will automatically recognize activities and alert teams to delays and errors in real time. In this paper, we focus on the speech modality. We analyzed speech transcripts from 39 resuscitation cases and compared the differences in speech intention and sentence types during delayed and non-delayed work, whether any differences persisted across different interventions, and how might these differences lend to an automatic speech-based delay detection system. Characterizing speech within the context of trauma resuscitation will not only inform a CDSS design but will also contribute to language modeling and automatic speech recognition in other types of time-critical, team-based medical work.

To characterize speech during time-critical activity delays, we focused on three LSIs that were delayed most frequently and contributed most time to delays in our dataset: (1) the process of establishing intravenous or intraosseous (IV/IO) vascular access, which allows physicians to administer medications or fluids[7,8], (2) CPR activity—a lifesaving procedure performed when the heart stops beating; immediate CPR can double or triple the chances of survival after cardiac arrest[7,8], and (3) patient intubation—the placement of an airway tube into the trachea to maintain a clear airway[7,8]. Previously, we analyzed a subset of cases with IV/IO delays[9], finding a significant difference in the proportion of intention labels between speech during delays and speech during non-delayed activities. This paper expands on our prior work by adding analyses of speech intention during delays associated with intubation and CPR and investigating whether any patterns persist across all activities.

Background and Related Work

Trauma resuscitation is a form of time-critical and high-risk teamwork. Trauma teams follow the Advanced Trauma Life Support (ATLS) protocol[10] to achieve consistent care by prioritizing resuscitation activities. The protocol consists of two phases: the primary and secondary surveys. In the primary survey, the team evaluates the patient's Airway, Breathing, Circulation, Disability (neurological assessment) and Exposure (patient clothing removed for further injury assessment). After the patient has been stabilized, the team performs the secondary survey, a detailed evaluation to identify other injuries. The trauma team includes seven to fifteen clinicians from varying disciplines, including a surgical attending, fellow or senior resident (team leader), a junior resident or nurse practitioner (physician surveyor), a scribe, a medication nurse, two or three bedside nurses, an anesthesiologist, and a respiratory therapist.

Prior work in computer supported cooperative work (CSCW) has analyzed team-based speech to study collaboration and coordination in several medical settings, including intensive care units[11] and emergency departments[12,13,14]. In emergency medical and trauma resuscitation settings, team members use a combination of speech, gesture, and movement to establish a shared mental model and coordinate activities[12,14]. Trauma team members observe each other's gestures and movement to develop awareness of actions within the room and prepare for potential requests[14]. Gestures and movement are important representations of people's actions, but the chaotic nature of medical work makes it difficult to closely monitor multiple ongoing activities[14]. Speech, however, requires no visual attention and offers an alternative approach to increasing team situational awareness and understanding team activities.

Several real-time computer-aided decision support systems have been implemented to reduce errors during trauma resuscitation[4,5]. The success of these early systems has been limited because they either require manual data entry or use automatic but incomplete data about the process. Other clinical research has attempted to capture automated activity data using sensors to identify process deviations[15,16,17]. In this research, computer vision and radiofrequency identification (RFID) tags are often used to track people and objects[17,18]. Some activities, however, are performed without objects and cannot be tracked using computer vision or RFID. Reliance on speech may address this challenge because trauma teams think aloud and verbalize much of their activity. Speech is used to request, share, and communicate information and tasks, contributing to team situational awareness[17]. Speech can provide useful clues for activity recognition because it contains rich information about team activity that cannot be captured using other modalities. Using speech as an input for real-time decision support can be challenging because we lack knowledge about the speech structure and intention during time-critical medical work like trauma resuscitation.

Related informatics work on speech analysis often focuses on speech mining and natural language processing (NLP). Previous research has defined "intent" as understanding speech by converting speech lines into representations of meaning[19]. Common methods such as semantic analysis[20] and language understanding[21] mainly consider syntax and ignore situational context and acoustic information such as tone. Our use of speech intention differs from the generalized action-based intention recognition in NLP work. We aim to identify the purpose of a team member's speech that is situated in the case. Words can carry different intentions depending on the context as well as speech delivery, which are critical to understanding the underlying meaning.

Context is particularly important in speech during time-critical medical work. Speech during trauma resuscitation contains information about medical procedures and teamwork that can be used to detect the type, duration, and status of medical activities[22]. However, speech is often brief, lacks structure, and its meaning can be dependent on delivery and context. For example, "*tape on the right hand*" can be a report, a reply, a request, or a question depending on the context of the speech line, the inflection, and the punctuation. In prior work, we have shown the ability of a multimodal deep learning model to classify intention based on acoustic information and context[23], but we have not studied if speech intention could be used to detect whether an activity is delayed.

Methods

Research Setting: This study took place at the trauma center of an urban, pediatric teaching hospital in the mid-Atlantic region of the United States. The trauma bay at our research site is equipped with an always-on video and audio recording system for recording live resuscitations under a protocol approved by the hospital's Legal and Risk Management Department. The study was also approved by the hospital's Institutional Review Board (IRB). Video and audio recordings were stored at the research site and were accessed using a remote desktop access granted over video conferencing software by a member of the clinical research team.

Data Collection: To create our dataset, we transcribed the speech from actual trauma resuscitation recordings. We recorded 122 trauma resuscitations with one or more LSIs occurring between January 2016 and December 2019. Of these, 53 cases received caregiver consent for using the video and audio recordings for research purposes. Members

Table 1. Intention labels: Definitions of assigned intention labels adapted from prior work.[13]

Intention Label	Definition
Assess need for an activity	Team members discuss the results of assessment activities or the outcomes of prior actions to decide whether to perform an intervention.
Assess results for an activity	After completing an activity, the team assesses the activity results.
Confirming information receipt or understanding of request	A team member indicates that they received the information or heard a request. This can include a reply to the request confirming or denying it.
Call attention to information, item, or activity	A team member calls attention to a piece of information or an activity.
Interact with patient	A team member interacts with the patient or patient parent. This can include posing questions, requesting actions, providing information.
Preparation to perform an activity	A question or statement intended to prepare the team for an activity.
Provide information	A team member provides information such as an answer to a request for information, clarification, or feedback to the team.
Report progress of an activity	During multi-step activities that take longer to perform, team members continuously communicate their activity progress.
Report results of an activity performance	A team member reports the numerical results of an assessment activity or the result of an intervention
Request action	A team member asks another or the room to perform or modify an activity. This can also include requests to change volume or terminate an activity.
Request information	A team member requests information or clarification about the activity.
State a plan or intention to perform an activity	An individual states that they intend to perform an activity

of our research team transcribed all 53 audio recordings, requiring up to 10 hours of listening for each event. To ensure consistency in transcription, each transcriber followed the following steps. First, they filtered blank sections of the audio and removed identifiable information. They then listened to the audio multiple times to transcribe all uttered speech, separate overlapping speech, timestamp each speech line, and identify speakers by role. Six cases were excluded from the dataset due to large sections of unintelligible audio or corrupted video recordings. Eight cases did not include delays relating to CPR, IV/IO insertion, or intubation. The final dataset included 39 cases (Table 2).

After transcription was completed, a clinical research fellow with multiple years of experience in trauma resuscitation used video recordings to identify "critical windows" where the trauma team performed LSIs. For instance, a critical window would be from the moment when the team decided to intubate to the moment when the team successfully established and secured the airway. Next, this clinical researcher marked any delays that occurred during the critical windows. The researcher used the following criteria to consistently determine whether a delay occurred: (1) if team members had proceeded to the next phase of the protocol without completing required activities in an earlier phase, (2) if a team member required prompting to start an activity, (3) if a team member waited for the completion of the preceding activity before starting their activity, and (4) if a team member executed an activity with pauses and slowness, requiring prompting to continue or accelerate their performance. Case transcripts were imported into a spreadsheet and each speech line was annotated with the timestamp and whether it occurred during a delay based on previously established delay start and end times.

Dataset Overview: Injury types ranged from falls, motor vehicle incidents, and gunshot wounds. The patient age ranged from less than one month old to 14 years old, with an average age of 5 years and 3 months old. The resuscitation

Table 2. Dataset overview per LSI: The number of cases, the number of speech lines during delays and non-delayed activities, and the results of IRR test for intention labels and sentence type.

LSI	Number of Cases	Speech Lines During Delays	Speech Lines During Non-Delayed Activities	Intention Label IRR	Sentence Type IRR
IV/IO	14	899	1,819	0.86	0.91
CPR	10	1,328	2,407	0.89	0.92
Intubation	28	7,450	5,863	0.85	0.89

duration spanned from l0 to 58 minutes, with an average duration of 27 minutes. The average number of speech lines per case was 253.9 (SD=240.9). IV/IO access was delayed in 14 cases and the delays lasted an average of 114 seconds. CPR was delayed in ten cases and each delay lasted an average of 21 seconds. Intubation was delayed in 28 cases and these delays lasted an average of 124 seconds (Table 2).

Data Analysis: A team of five medical students individually labeled all speech lines of the transcribed CPR, IV/IO, and intubation cases. The students assigned one of eleven speech intention labels adapted from prior classification of trauma team communications[2] (Table 1). The label "Other" was reserved for speech of uncertain intention or context that required further video review to discern its purpose. Sentence type labels further characterized the speech as interrogative, imperative, negative (inclusive of words such as "no," "not," and "don't"), or normal (all other speech).

After the initial analysis, speech lines labeled with "Other" or "Hold for review" were reviewed using video recordings to ensure the speech was appropriately labeled based on the context. Some utterances initially marked as unintelligible were successfully deciphered and labeled during this review. The medical students marked lines that remained unintelligible or were too incomplete to determine the intention as "Other." These lines were also not labeled for sentence type analysis.

Following individual labeling, 10% of each case was also labeled by the first author to ensure consistent coding across the research team to achieve valid results. The labeling results from both the individual and 10% coding were compared using an Inter-Rater Reliability (IRR) metric (Cohen's Kappa) to assess concordance of the labels assigned by each researcher. The results showed strong inter-coder agreement for each LSI (Table 2).

To determine if the distribution of intention labels and sentence type differed between the speech during delayed and non-delayed activities, we performed a Chi-Goodness-of-Fit test for each LSI. Our null hypothesis was that the distribution of labels between speech during delayed and non-delayed work was the same. We calculated the expected values by finding the proportion of speech lines assigned to each label and multiplied the percentage to the total number of speech lines during delays. We performed a two-sample z-test to determine if the difference in the proportion of intention labels between speech during delays and non-delayed activities was significant. We also performed a Chi-Goodness-of-Fit test and a two-sample z-tests for sentence type. For all statistical tests, we used a two-tailed alpha of 0.05 to test the level of significance.

Findings

We present our findings in four parts, with the first three focusing on intention and sentence type during each LSI and the last part presenting the results across all LSIs: (1) IV/IO access delay related speech, (2) CPR delay related speech, (3) intubation delay related speech, and (4) speech across all LSI delays.

Speech Intention and Sentence Type During Delays Associated with IV/IO Access

Fourteen cases had delays during the IV/IO insertion activity. A total of 2,933 were associated with this activity across all cases, with 899 occurring during delay, 1,819 occurring during non-delayed work, and 215 unintelligible or were too incomplete to determine the intention.

We observed a significant difference in the distribution of intention labels between speech during delayed and non-delayed work, X^2 (11, n = 2718), $p < .001$. Speech with the intentions "Assess need for an activity" and "Interact with patient" was significantly more frequent during non-delayed work than during the delays (Table 3).

Team members assessed the need for activities in 3.68% of all speech lines during non-delayed work, which is more than twice the percent of speech lines where team members assessed results during delays (1.89%) (Table 3). For example, clinicians would discuss whether they needed additional vascular access, stating "*do we need another*

Table 3. Summary statistics for speech intentions during the IV/IO access task per intention label: The proportion and count of speech lines during delays and non-delayed work, the standard z-score, and the p-value.

Intention Label	Non-Delay	Delay	Z-Score	P-Value
Assess need for an activity	**3.68% (67)**	**1.89% (17)**	**2.53**	**0.0114***
Assess results of an activity performance	0.77% (14)	0.33% (3)	1.35	0.1762
Call attention to information, item, or activity	6.98% (274)	7.68% (160)	-0.606	0.5440
Confirming information receipt	6.27% (127)	8.45% (69)	-0.673	0.5001
Interact with patient	**2.25% (41)**	**1.00% (9)**	**-2.102**	**0.0355***
Preparation to perform an activity	3.13% (57)	2.89% (26)	0.344	0.7305
Provide information to the team	4.67% (85)	3.45% (31)	1.485	0.1376
Report progress of an activity	12.64% (230)	12.46% (121)	-0.595	0.5513
Report results of an activity performance	9.40% (171)	8.90% (80)	0.4249	0.6709
Request information	12.70% (231)	10.79% (97)	1.4367	0.1508
Request action	**23.20% (422)**	**27.92% (251)**	**-2.68**	**0.0074***
State a plan or intention to perform an activity	5.83% (10)	6.45% (58)	-0.6426	0.5205

*Indicates significant results

access?" or "*we need to have IO access."* During delays, the team would have often already assessed the need for activities and would work towards completing the activity to mitigate those delays.

We also observed more interactions with the patient during non-delayed work than during delays. During non-delayed work, team members interacted with the patient in 2.25% of speech lines, compared to 1% of speech during delays (Table 3). Clinicians would query patients (e.g., "*can you say your name?*") or request patients to perform an action (e.g., "*open your eyes please*") during routine work to determine the status of the patient. These assessment inquiries and requests were less frequent during delays, when the team was more focused on completing the interventions.

Speech during delays had significantly more speech lines with the intention labels "Confirming information receipt or understanding of request" and "Request action." Team members confirmed information receipt in 8.45% of speech lines during delays as opposed to 6.27% of speech lines during non-delayed work. Phrases such as "*okay,*" "*yes I will*," and "*alright*" were more frequently heard during delays. Team members requested actions in 27.92% of all speech lines during delays, while 23.2% of speech lines with this label occurred during non-delayed work. For example, in cases where the team had made multiple IV/IO insertion attempts during delays, the leader told team members to modify actions or terminate activities, making statement such as "*slow down your compressions*" or "*use another line, use another line.*"

The distribution of sentence type (interrogative, imperative, negative, and normal sentences) did not significantly differ between speech occurring during delays and non-delayed work related to the IV/IO access task, X^2 (4, n = 2718), p=0.181.

Speech Intention and Sentence Type During Delays Associated with CPR

Ten of 39 cases in our dataset contained delays during the CPR activity. A total of 3,894 were associated with this activity across all cases, with 1,328 occurring during delays, 2,407 occurring during non-delayed work, and 159 classified as unintelligible.

We observed a significant difference in the distribution of intention labels between speech lines associated with delays during CPR and speech during non-delayed work, X^2 (11, n = 3,735), p < .001. We also found a significant difference in the proportion of speech during CPR for intention labels "Interact with patient," "Request information" and "Report progress of an activity" (Table 4).

The requests for information were significantly higher in frequency during CPR delays because team members often requested information through statements such as "*how long has [nurse] been doing CPR?*" or "*are you ok with CPR?*" These statements served to assess the CPR status and ensure that team members were correctly performing CPR and for the appropriate amount of time. The proportion of the "Interact with patient" intention label appears to be significantly associated with delays, but upon closer examination, we observed only four speech lines with this

Table 4. Summary statistics for speech intentions during CPR per intention label: The proportion of speech lines during delays and non-delayed work, line count, standard z-score, and p-value.

Intention Label	Non-Delay	Delay	Z-Score	P-Value
Assess need for an activity	12.46% (300)	11.44% (152)	0.9129	0.36125
Assess results of an activity performance	2.28% (55)	1.73% (23)	1.1315	0.25784
Call attention to information, item or activity	7.60% (183)	8.43% (112)	-0.91025	0.36745
Confirming information receipt	11.38% (274)	12.08% (160)	-0.60682	0.54397
Interact with patient	**0.04% (1)**	**0.30% (4)**	**-2.0077**	**0.03775***
Preparation to perform an activity	2.29% (55)	2.78% (37)	-0.94585	0.34423
Provide information to the team	14.79% (356)	15.96% (212)	-0.95615	0.339
Report progress of an activity	**5.82% (140)**	**4.22% (56)**	**2.09845**	**0.03587***
Report results of an activity performance	7.31% (176)	7.38% (98)	0.73795	0.93962
Request information	**9.26% (223)**	**11.59% (154)**	**-2.26439**	**0.02355***
Request action	17.08% (411)	16.34% (217)	0.57481	0.56541
State a plan or intention to perform an activity	4.90% (118)	4.44% (59)	0.63281	0.52686

*Indicates significant results

intention label during CPR delays. These speech lines occurred during a single delay event, where the team discussed the patient's symptoms with a caregiver before proceeding with CPR.

Team members were significantly less likely to "Report progress of an activity" during delays. Most of the time, progress reports were associated with non-delayed work, when team members shared their status through statements such as "*CPR is good… CPR is continuous*" or "*on one minute for CPR*."

No significant difference was observed in the overall distribution of sentence type (interrogative, imperative, negative, and normal sentences) between speech during delayed events and speech during non-delayed work ($p = 0.374$).

Speech Intention and Sentence Type Associated with Delays During Intubation

Of 39 cases in our dataset, 28 contained delays during patient intubation and intubation related activities. Within these cases, we identified a total of 8,240 total speech lines. Of these, 5,863 speech lines occurred during non-delayed work, 1,607 occurred during delays, and 770 speech lines were classified as unintelligible or too incomplete to determine the intention label.

We observed a significant difference in the distribution of intention labels between speech during delayed and non-delayed work, X^2 (11, n = 7470), $p < .001$. Speech lines with the intentions "Assess results for an activity," "Interact with patient" and "Report results" were significantly more frequent during non-delayed work than during intubation-related delays (Table 5).

Team members assessed the results of their activities in 1.62% of all speech lines during non-delayed work, while this assessment was twice as lower during delays (0.087%) (Table 5). After completing the activities, team members would discuss results before moving onto the next step. For example, a team leader would state "*alright, [the patient] got good chest rise with bagging*" or "*3 and 5, pupils are dilated and non-reactive, we are going to start hyperosmolar therapy with hypertonic saline.*" Speech lines with this intention are often longer and more conversational. In contrast, team members assessed the results of their activities less often during delays as they were focused on administering the appropriate interventions rather than discussing results of prior activities.

During non-delayed work, team members also more frequently interacted with the patient (3.1% of speech lines labeled with "Interact with patient" during non-delayed work vs. 2.1% of speech lines during delays) (Table 5). Team members typically explained the procedure or comforted the patient before or after major interventions such as intubation using statements such as "*we're going to move you buddy, okay*," "*hey wanna see this teddy bear*," and "*I know, we're going to have all your [toys] taken care of, okay, you just have to relax, can you take my hand?*" In contrast, speech lines when clinicians are interacting with the patient were less frequent during delays, as they were usually performing activities and focused on collaborating with each other.

Team members reported results in 10.8% of speech lines during non-delayed work, while 7.71% of speech lines contained the reports during delays. In these speech lines, clinicians reported assessment results such as "*pupils equal*

Table 5. Summary statistics for speech intentions during intubation delays per intention label: The proportion of speech lines during delays and non-delayed work, line count, standard z-score, and p-value.

Intention Label	Non-Delay	Delay	Z-Score	P-Value
Assess need for an activity	6.77% (397)	7.84% (152)	-1.488	0.1362
Assess results of an activity performance	**1.62% (95)**	**0.87% (23)**	**2.2188**	**0.02634***
Call attention to information, item, or activity	5.33% (313)	5.85% (112)	1.5143	0.13104
Confirming information receipt	10.8% (635)	9.52% (160)	-0.7993	0.42372
Interact with patient	**3.08% (181)**	**2.12% (34)**	**2.0635**	**0.0394***
Preparation to perform an activity	2.11% (124)	1.49% (24)	1.581	0.11382
Provide information to the team	1.24% (728)	1.31% (210)	0.5732	0.97606
Report progress of an activity	5.34% (313)	4.98% (80)	3.643	0.5686
Report results of an activity performance	**10.81% (634)**	**7.72% (124)**	**-4.2744**	**0.00028***
Request action	**15.23% (893)**	**19.67% (316)**	**1.3206**	**<.00001***
Request information	12.74% (747)	11.51% (185)	-0.4098	0.18684
State a plan or intention to perform an activity	3.71% (217)	3.92% (63)	-1.488	0.6818

*Indicates significant results

and reactive bilaterally," patient values such as vital signs (e.g., "*130 over 90*" to report blood pressure), or activity results (e.g., "*okay he has got an oral airway*" to indicate the airway had been established). These short and often numeric sentences are representative of reporting results. Fewer speech lines had the intention "Report results" during delays because delays occurred during the activities and not after the activity was completed.

Speech during delays had significantly more lines with the intention label "Request action." Team members requested actions in 19.7% of all speech lines during intubation delays, as opposed to 15.2% of speech lines during non-delayed work. Team members communicated to ensure actions are completed or modified appropriately. For instance, a clinician would say "*can we get suction please*" to request assistance or "*let's hold on the intubation, it would be better for anesthesia to get here first*" to deliberate the best course of action during a delay.

The distribution of sentence type significantly differed between speech occurring during delays and speech during non-delayed work, X^2 (4, n = 6696), p=0.007. Team members uttered significantly more imperative sentences during delays (7.96%) than during non-delayed work (6.33%), (Z= -2.19, P = 0.03). The higher frequency of the "Request action" intention label aligns with this increased prevalence of imperative sentences. No other sentence type significantly differed between speech during delays and non-delayed work.

Speech Intention and Sentence Type across All LSIs

When combining the cases for all three LSIs, we observed a significant difference in the distribution of intention labels between speech during delays and non-delayed work, X^2 (11, n = 8096), p < .001. Speech lines with the intention labels "Assess need for an activity," "Interact with patient," and "Report results" were significantly more frequent during non-delayed work than during delays (Table 6). The only significant result for intention labels during delays was observed for the intention "Request action" (Table 6). The results of this combined analysis for all three LSIs most closely reflects the findings observed for speech during intubation and IV/IO access.

No significant difference was observed in the overall distribution of four sentence types (interrogative, imperative, negative, and normal sentences) between speech during delays and speech during non-delayed work (p = 0.234).

Discussion

The findings from our analysis of speech intention and sentence types during the three commonly performed LSIs in trauma resuscitation provided several insights into speech as a sensor modality for delay detection during a complex medical process. As we found in this study, speech can be used to extract rich information about ongoing activities and their progression. Reports about the activity results indicated that the activity was completed, while reports about the progress of an activity indicated that the activity was still ongoing. Similarly, when teams were assessing the need for an activity, it implied that the activity did not start yet.

Table 6. Summary statistics for speech intentions for all LSIs per intention label: The proportion of speech lines during delays and non-delayed work, line count, standard z-score, and p-value.

Intention Label	Non-Delay	Delay	Z-Score	P-Value
Assess need for an activity	**448 (7.32%)**	**107 (5.41%)**	**2.91**	**0.0035***
Assess results of an activity	105 (1.72%)	29 (1.47%)	0.752	0.4523
Call attention to information,	422 (6.89%)	148 (7.49%)	-0.898	0.3690
Confirming information receipt	519 (8.48%)	179 (9.059%)	-0.796	0.4258
Interact with patient	**144 (2.35%)**	**28 (1.42%)**	**2.51**	**0.0121***
Provide information	845 (13.81%)	275 (13.92%)	-0.123	0.9022
Report progress of an activity	434 (7.10%)	159 (8.05%)	-1.42	0.1566
Report results of an activity	**586 (9.58%)**	**159 (8.05%)**	**2.04**	**0.041***
Request action	**932 (15.23%)**	**349 (17.66%)**	**-2.58**	**0.0099***
Request information	663 (10.83%)	218 (11.03%)	-0.247	0.805
State a plan or intention to perform an activity	284 (4.64%)	104 (5.26%)	-1.13	0.2599

*Indicates significant results

Our findings showed that overall speech intentions significantly differed during delayed and non-delayed work for all three LSIs. Speech with the intention "Request action" was significantly more frequent during delays, while speech with the intentions "Report results," "Interact with patient," and "Assess need for an activity" was less frequent during delays. This overall analysis, however, hid the significant speech intentions observed in CPR cases.

When we took a closer look at verbal communication associated with delays during each of the LSIs, we observed certain patterns in speech intentions for different LSIs. During CPR, for example, speech with the intention "Request information" was significantly more prevalent during delays. This finding suggests that delays in CPR could be expected if the team makes multiple requests for information while the activity is being performed. In contrast, delays in establishing IV/IO access and performing intubation could be expected if the team makes multiple "Requests for action." This difference in speech characteristics during delays between the LSIs can be attributed to the nature of the intervention. During CPR delays, team members request information more frequently to assess the status of CPR. During delays in intubation or IV/IO insertion, team members issue more requests for action to mitigate delays and ensure that these multi-step activities keep proceeding.

The type of sentence and keywords associated with speech intentions could also facilitate delay and activity recognition. Key phrases such as "*can you,*" "*can someone please*," and "*do you want me to*" were frequently heard during delays in IV/IO access tasks, while keywords such as "*how long has*," "*can I have an update on*," and *"what is the*" were frequently heard during delays in CPR. The results for sentence types across all three LSIs showed that only intubation contained speech that had a significantly higher proportion of imperative sentences. Imperative sentences were likely more prevalent due to the higher proportion of speech lines with the "Request action" intention. More action requests were phrased as commands rather than questions during intubation. For instance, during intubation delays, team members more frequently exchanged phrases like "*let's change it to*," "*please give,*" or "*do [subject] now*."

Our findings build the evidence base that a speech-based activity recognition system must consider the context within team-based, time-critical work. This finding introduces new challenges for designing a real-time delay detection system. The future system must be able to capture the context – the activity being performed – as well as the phrases and words that indicate delays for those activities. By developing a multimodal system that uses video capture, motion sensing, and speech recognition, the system should be able to capture the necessary context to detect delays.

This study also exposed several challenges when studying speech in a complex medical setting. First, we encountered many incomplete or unintelligible communications while transcribing the resuscitation cases, which could have affected our results. However, speech during both delays and non-delayed work had a similar proportion of unintelligible speech in our dataset, removing a potential source of bias in our data. Second, trauma teams perform many overlapping activities, which added a challenge in determining the flow of activities and whether speech was relevant to delays. Finally, capturing high quality audio was difficult due to the noisy environment and frequent movement of team members. The shotgun microphones installed at our research site and directed towards key team roles helped address this challenge.

Conclusion

In this paper, we analyzed speech intention and sentence types during three commonly delayed LSIs during trauma resuscitation: IV/IO insertion, CPR, and intubation. Our goal was to understand how speech patterns change during delays. We transcribed 39 resuscitation cases, producing a total of 8,831 speech lines. Each speech line was labeled with one of eleven intention labels and one of four sentence types. We then ran statistical tests to identify differences in speech intentions during delayed and non-delayed work.

Through this analysis, we found a significant difference in patterns for speech associated with delays vs. non-delayed work. During delays associated with IV/IO insertion and intubation, team communications included more requests for actions, while requests for information were significantly more common during delays associated with CPR. We found that the patterns of speech intention differed depending on the LSI. These findings support the importance of context when analyzing speech as well as the possibility that a system could use speech intention as an input to detect when delays are occurring.

Acknowledgements

We thank the Children's National Medical Center staff for their participation in this research. We also thank Alex Lee, Matt Coates, Swathi Jagannath, and Emily Alberto for their contributions to this research. This research has been funded by the National Science Foundation under the collaborative research grant numbers IIS-1763509, IIS-1763355, and IIS-1763827.

References

1. Gruen R, Gregory J, Jurkovich L, McIntyre H and Maier, R. Patterns of errors contributing to trauma mortality. Annals of Surgery 244, no. 3 (September 2006): 371–80.
2. Kusunoki D and Sarcevic A. Designing for temporal awareness: the role of temporality in time-critical medical teamwork. CSCW: Proceedings of the Conference on Computer-Supported Cooperative Work. 2015 (March 2015): 1465–76.
3. Cales RH, Trunkey DD. Preventable trauma deaths. A review of trauma care systems development. JAMA. 1985 Aug 23-30;254(8):1059-63.
4. Clarke J, Webber B, Gertner A, Kaye J, and Rymon R. On-line decision support for emergency trauma management. In Proc. Annual Symposium on Computer Application in Medical Care, American Medical Informatics Association. 1994 (1028).
5. Fitzgerald M, Cameron P, Mackenzie C, Farrow N, Scicluna P, et al. Trauma resuscitation errors and computer-assisted decision support. Archives of Surgery. 2011;146(2):218-225.
6. Kim D, Oh J, Im H, Yoon M, Park J, Lee J. Automatic Classification of the Korean Triage Acuity Scale in Simulated Emergency Rooms Using Speech Recognition and Natural Language Processing: a Proof of Concept Study. J Korean Med Sci. 2021;36(27):e175.
7. Tengs T, Adams M, Pliskin J, Safran D, Siegel J, Weinstein M, et al. Five-hundred life-saving interventions and their cost-effectiveness. Risk analysis 1995;15(3):369-390.
8. Swendiman R, Sharoky C, Russell K, Goldshore M, Blinman T, Nance M. Life-saving interventions in pediatric trauma: a National Trauma Data Bank Experience. Journal of trauma and acute care surgery. 2019;87(6):1321-1327.
9. Zellner K, Coates M, Lee A, Jagannath S, Sarcevic A, Alberto E, et al. Characterizing speech in life saving interventions to inform computerized clinical decision support for complex medical teamwork. Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing. Association for Computing Machinery. 2021;199–202.
10. Advanced trauma life support. Chicago, IL: American College of Surgeons; 2018.
11. Reddy M, Dourish P, and Pratt W. Coordinating heterogeneous work: information and representation in medical care. In ECSCW 2001:239-258.
12. Härgestam M, Hultin M, Brulin C, Jacobsson M. Trauma team leaders' non-verbal communication: video registration during trauma team training. Scandinavian journal of trauma, resuscitation, and emergency medicine. 2016;24(1):37.
13. Sarcevic A, Marsic I, Lesk ME, Burd RS. Transactive memory in trauma resuscitation. Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08. 2008;215-224.

14. Zhang Z, Sarcevic A. Constructing Awareness Through Speech, Gesture, Gaze and Movement During a Time-Critical Medical Task. ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway. 2015;163–82.
15. Forestier G, Lalys F, Riffaud L, Trelhu B, Jannin P. Classification of surgical processes using dynamic time warping. Journal of Biomedical Informatics. 2012 Apr;45(2):255–64.
16. Neumuth T, Meißner C. Online recognition of surgical instruments by information fusion. International Journal of Computer Assisted Radiology and Surgery. 2011 Oct 18;7(2):297–304.
17. Vankipuram M, Kahol K, Cohen T, Patel VL. Toward automated workflow analysis and visualization in clinical environments. Journal of Biomedical Informatics. 2011 Jun;44(3):432–40.
18. Li X, Yao D, Pan X, Johannaman J, Yang J, Webman R, et al. Activity recognition for medical teamwork based on passive RFID. 2016 IEEE International Conference on RFID (RFID). 2016 May
19. Ruiter D, Cummins, C. A model of intentional communication: AIRBUS (Asym-metric Intention Recognition with Bayesian Updating of Signals). Proceedings of SemDial 2012;149-50.
20. Wang, Y.-Y., Deng, L., Acero, A.: Spoken language understanding. IEEE Signal Processing Magazine. 2005(22):16–31
21. Tur G, Mori, R. Introduction. Spoken language understanding. 2011:1–7
22. Jagannath S, Sarcevic A, Marsic I. An Analysis of Speech as a Modality for Activity Recognition during Complex Medical Teamwork. Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018
23. Williams JD, Kamal E, Ashour M, Amr H, Miller J, Zweig G. Fast and easy language understanding for dialog systems with Microsoft Language Understanding Intelligent Service (LUIS). Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2015;
24. Gu Y, Li X, Chen S, Zhang J, Marsic I. Speech intention classification with multimodal deep learning. Adv Artif Intell. 2017;10233:260-271.