Characterizing Speech in Life Saving Interventions to Inform Computerized Clinical Decision Support for Complex Medical Teamwork

Characterizing Speech in Life Saving Interventions

Katherine Zellner, Matt Coates, Alex Lee, Swathi Jagannath, Aleksandra Sarcevic

Drexel University, {kz276, mpc327, al3548, sj532, as3653}@drexel.edu

Emily Alberto, Allison Harvey, Randall S. Burd

Children's National Medical Center, {ealberto, rburd}@childrensnational.org, allisonharvey@gwmail.gwu.edu

Ivan Marsic

Rutgers University, marsic@rutgers.edu

We describe an initial analysis of speech during team-based medical scenarios and its potential to indicate process delays in an emergency medical setting. We analyzed the speech of trauma resuscitation teams during cases with delayed intravenous/intraosseous (IV/IO) line placement, a significant contributor to delays during life-saving interventions in trauma resuscitation. The insights gained from this experiment will contribute towards the design of a clinical decision support system (CDSS) that will use multiple sensor modalities to alert medical teams to errors in real time. We contribute to the literature by determining how the intention of each speech line and the sentence can support real-time, automatic detection of delays during time-critical team activities.

1 INTRODUCTION

Patients treated during trauma resuscitation, the initial phase of treatment after injury, have a higher risk of morbidity and mortality due to medical errors than other hospitalized patients [2]. Patients who require *life-saving interventions* (LSIs)—treatments that reduce the probability of premature death among a specified target population [1]—have an even higher mortality rate [2]. LSIs during trauma resuscitation include procedures such as intubation (inserting a tube into the airway), blood transfusion, chest tube placement (inserting a tube to drain air or fluid from the chest), cardiopulmonary resuscitation (CPR), and resuscitative thoracotomy (invasive access of the chest cavity to control catastrophic blood loss). Timely performance of LSIs plays a critical role in improving patient outcomes [8]. Understanding the delays that occur during these LSIs and how teams respond to these delays is critical for designing clinical decision support systems (CDSS).

Prior work in HCI and CSCW research has identified challenges in recognizing speech during medical work [4], but the underlying speech patterns and structures remain under studied. Our long-term research goal is to support clinical teamwork during complex workflows by developing a CDSS that will recognize activities and alert teams to delays in real time. Speech is critical during resuscitation activities, as clinicians communicate to request tasks, share information, and confirm completion of activities. Understanding the nature of speech, particularly during LSIs, is necessary to inform the design of an effective, user centered CDSS. Prior work has evaluated the feasibility of detecting clinical activities through speech [3], but it has not yet explored delay detection.

In this paper, we report our initial analysis of speech associated with delays during LSIs performed in 53 pediatric trauma resuscitations. A challenging activity necessary for the success of every LSI during pediatric trauma resuscitation is the establishment of vascular access, either intravenously (IV) or intraosseously (IO) [8]. In our dataset, IV/IO insertion was the most common delayed activity (36% of the cases) and had the longest delays (mean = 114 seconds). Reducing IV/IO insertion delays could make LSIs more efficient and improve patient outcomes. We focused our initial analysis on speech surrounding delayed IV/IO placement. Through this analysis, we have identified patterns of speech associated with delays to assist in delay detection for the future development of the CDSS.

2 METHODS

2.1 Dataset and Data Preparation

Six members of our research team transcribed audio recordings of actual trauma resuscitations performed at a pediatric urban teaching hospital in the US Northeast region. During our data collection period (January 2016-December 2019), 122 trauma resuscitations containing one or more LSIs occurred. Among these resuscitations, 53 had video and audio recordings that were available for transcription and analysis after obtaining patient/caregiver consent. Transcribing each event required up to 10 hours of active listening and to capture the uttered speech. The transcripts were chronological and included every utterance in the event. Each transcriber filtered out the blank sections of the audio files. Second, they removed any personal health information that could identify the patient. They then listened to the audio multiple times to transcribe all speech, separate overlapping speech, timestamp each speech line, and identify speaker roles.

Next, a clinical researcher with multiple years of trauma resuscitation experience reviewed video recordings of all 53 cases and identified 122 LSI episodes. We call these episodes “critical windows.” An example of a critical window is from the moment when the team decided to insert a chest tube to the moment when the team successfully inserted and secured the chest tube. Next, this same clinical researcher and a medical student identified delays that occurred during the critical windows. To ensure consistency, we used the following criteria to determine if a delay occurred: (1) if team members had proceeded to the next phase of the protocol without completing the previous required, (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 needed a prompt to speed up their performance. Using this approach, the clinical researchers identified 328 activity delays across all 53 cases. Of these 328 delays, 76 were related to IV/IO placement occurring in 14 cases.

Using the timestamps, we marked the speech related to IV/IO delays in the corresponding transcripts and determined whether the speech lines were uttered before delays, during delays, or after delays were remedied. Through this analysis, we identified 2,625 speech lines, 840 occurring during delays and 1,785 during non-delayed activities.

Table 1: Intention Labels and Associated Activity Stage

Intention Label	Activity Stage
Assess need for an activity	Before
Request to perform an activity	Before
Request clarification or more information on an assigned activity	Before
Interact with patients about an activity	Before
Query patients or request them to perform an action	During
Request information on progress of an activity	During
Report progress of an activity	During
Request to modify or terminate an ongoing activity	During
Report results of an activity performance	After
Request clarification or more information on an activity performance	After
Assess results of an activity performance	After

2.2 Data Analysis

Three researchers (a human-centered computing researcher and two medical students, all trained in the basics of trauma resuscitation) reviewed all speech lines and labeled them with one of 11 intention labels adapted from prior classifications of trauma team communications [6] (Table 1). Each label aligns with an "activity stage" to indicate whether the speech line occurred before, during, or after the activity performance. We analyzed speech lines associated with both delayed and non-delayed activities to find whether the speech differs between the two groups to aid in delay detection. We noted the sentence type using four categories (normal, negative, interrogative, and imperative) to investigate whether questions, commands, or negative words like "no" or "not" occurred more frequently during delayed activities.

To ensure consistency in data coding researchers reviewed and discussed speech lines from an older dataset to practice coding. Next, the human-centered computing researcher labeled all speech lines in our dataset, while the medical students each labeled 50% of speech lines. After labeling a subset of four cases, we compared the labels of the researchers. The results showed a moderate agreement with an interrater reliability score of 0.56 (Cohen's Kappa). The researchers reviewed disagreements and discussed the correct labels. Most disagreements were for intention labels with similar functions. For example, "request clarification" and "request information" were often swapped as "report progress" and "report results."

After discussion and recalibration, the researchers completed the remaining 10 cases and the recalculated interrater reliability showed near perfect agreement with a score of 0.86 (Cohen's Kappa). To determine if the speech during delayed activities had the same distribution of intention labels and sentence types as speech during non-delayed activities differed, we conducted a Chi-square goodness of fit test. We then performed a two-sample t-test to compare the differences of the proportions per individual intention label. For all statistical tests, we used an alpha of 0.05.

Table 2: Summary Statistics for Intention Labeling

Intention Label	Speech Lines during delays (n = 840)	Speech Lines during routine work (n=1,785)	P-Value
Assess need for activity	17 (2.02%)	67 (3.75%)	**0.01875***
Assess result of an activity performance	3 (0.36%)	14 (0.78%)	0.21008
Interact with patient about an activity	13 (1.55%)	57 (3.19%)	**0.00748***
Query patients or request them to perform an action	4 (0.48%)	12 (0.67%)	0.27975
Report progress of an activity	263 (31.31%)	475 (26.61%)	**0.00623***
Report results of an activity performance	122 (14.52%)	275 (15.41%)	0.55273
Request clarification or information on an activity performance	53 (6.31%)	126 (7.06%)	0.47705
Request clarification or more information on an assigned activity	9 (1.07%)	41 (2.30%)	**0.01579***
Request information on progress of an activity	31 (3.69%)	85 (4.76%)	0.20891
Request to modify or terminate an ongoing activity	78 (9.29%)	95 (5.32%)	**0.00013***
Request to perform an activity	173 (20.60%)	327 (18.32%)	0.16525

3 FINDINGS

3.1 Speech Intentions during Activity Delays

We observed a significant difference in the distribution of intention labels between speech during delayed and non-delayed activities, X2 (10, n =2625), p < .001. More speech lines with the intentions "report progress of an activity" and "request to modify or terminate an activity" could be heard during delays than during non-delayed activities (Table 2). Although "report progress of an activity" was the most common intention label in both categories (delayed = 26.61%, non-delayed = 31.31%), the clinical team would more frequently report progress during delays. For example, in cases where the team had made multiple IV/IO insertion attempts during delays, a clinician would report "*I'm trying to get a line in."* The clinical team also requested team members to modify actions or terminate activities nearly twice as often during delays than during non-delayed activities than (9.29%, 5.32%), e.g., "*slow down your compressions*" or "*use another line, use another line*." During delayed activities, the clinical team requested changes or terminations more frequently than during non-delayed activities. The frequency of this intention label could indicate when delays occur. During delays, the clinical team assessed need for activities, interacted with patients less, and requested clarification or more information about an assigned activity significantly less often than in non-delayed activities (Table 2).

3.2 Sentence Types during Activity Delays

The distribution of sentence type did not differ between speech occurring during delays and speech during non-delayed activities, X2 (3, n =2625), p = 0.086. About five percent of speech lines were negative sentences, indicating that something was incorrect or that personnel could not accomplish a task. Most negative sentences occurred while activity was being performed (75%) and when clinicians were reporting progress of an activity (62%), such as "*this IO is not working.*" Eighteen percent of the speech lines were imperative sentences. Imperative speech lines occurred while activities were being performed and were intended to modify or terminate an activity (e.g., "*yeah, try a PIV then*"). Twenty percent of the sentences were interrogative sentences. Almost all of these speech lines were requesting information (e.g., *"is the IV on the left not working*?"). Lastly, 57% of speech lines were normal sentences, which were most often either reporting progress of an activity (e.g., "*significant blood in the mouth*") or reporting results of an activity (e.g., "*we got the second line*").

4 CONCLUSION

Through this analysis, we found a significant difference in speech patterns in speech associated with delays vs. speech during non-delayed work. During delays, the clinical team less frequently interacted with patients, talked about assessing need for activities, assessed results for activity performances, and requested for clarification. Rather, clinicians more frequently issue requests for modifications or terminations of activities and for the team to perform activities during delays. By identifying these intentions, a system would be able to detect when delays are likely occurring. When clinicians frequently discuss modifying or terminating activities, the system could recognize that a delay is occurring and issue alerts to the clinical team. Further speech analysis is needed to determine how the structures of speech lines relate to the intention labels and delays.

The distribution of sentence type was not significantly different between speech during delayed and non-delayed activity. Negative sentences, questions, or commands were not strong indicators of delays. Sentence type did align with activity stages, suggesting that this speech pattern can be used for activity stage recognition. CDSS would also need to recognize the activity stage and use this input to accurately determine whether a delay is occurring. Further research should investigate how speech patterns differentiate between intention labels since sentence type does not. For example, how does speech differ when clinicians are requesting modifications or terminating activities versus when just requesting to perform an activity or reporting progress?

Our findings suggest the following challenges in future work. First, our transcribers were humans who were able to rewind and re-listen to parts of speech that were not initially clear. In a real-time, critical speech lines and key words could be omitted due to technical difficulties, quiet speech, or too many people speaking simultaneously. Second, the intention label coding was challenging due to its subjective nature. However, through discussion and training protocols, we were able to establish high inter-rater reliability. Third, this study was our initial analysis of speech during non-delayed work, and we found that 8.6% speech-lines did not fit into one of the 11 predefined intention labels. Over the course of the analysis, we discovered new labels that could have fit some speech lines during non-delayed activities, such as "confirming information or request" or "personnel inquiry." These new labels would not have changed analysis of speech lines during delays. Future work should incorporate the new intention labels to investigate whether this difference between speech during delayed activities and speech during non-delayed activities continues.

We will continue to characterize speech during additional life-saving interventions, such as intubation, blood transfusion, and CPR. In addition, we will explore how speech could be combined with modalities such as video, computer vision, and other sensors to support the automatic delay detection in a complex medical setting such as trauma resuscitation.

ACKNOWLEDGMENTS

We thank the medical staff at Children's National Medical Center for their participation in this research. This research has been funded by the National Science Foundation under grant number IIS-1763509 and partly by National Library of Medicine of the National Institutes of Health under grant number 2R01LM011834-05.

REFERENCES

[1] Greene, N., Bhananker, S., & Ramaiah, R. (2012). Vascular access, fluid resuscitation, and blood transfusion in pediatric trauma. International journal of critical illness and injury science, 2(3), 135–142. https://doi.org/10.4103/2229-5151.100890

[2] Gruen, Russell L., Gregory J. Jurkovich, Lisa K. McIntyre, Hugh M. Foy, and Ronald V. Maier. “Patterns of Errors Contributing to Trauma Mortality.” Annals of Surgery 244, no. 3 (September 2006): 371–80. https://doi.org/10.1097/01.sla.0000234655.83517.56

[3] Jagannath S, Sarcevic A, Marsic I. An Analysis of Speech as a Modality for Activity Recognition during Complex Medical Teamwork. Int Conf Pervasive Computer Technology Healthcare. 2018 May;2018:88-97. doi: 10.1145/3240925.3240941

[4] Kusunoki, Diana S., and Aleksandra Sarcevic. “Designing for Temporal Awareness: The Role of Temporality in Time-Critical Medical Teamwork.” CSCW : Proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work 2015 (March 2015): 1465–76. https://doi.org/10.1145/2675133.2675279

[5] Mina, Michael J., Rashi Jhunjhunwala, Rondi B. Gelbard, Stacy D. Dougherty, Jacquelyn S. Carr, Christopher J. Dente, Jeffrey M. Nicholas et al. 2017. Factors affecting mortality after penetrating cardiac injuries: 10-year experience at urban level I trauma center. The American Journal of Surgery 213, no. 6, pp. 1109-1115.

[6] Sarcevic, Aleksandra, Ivan Marsic, Michael E. Lesk, and Randall S. Burd. "Transactive memory in trauma resuscitation." In Proceedings of the 2008 ACM conference on Computer supported cooperative work, pp. 215-224. 2008.

[7] Swendiman, Robert A., Catherine E. Sharoky, Katie W. Russell, Matthew A. Goldshore, Thane A. Blinman, and Michael L. Nance. 2019. Life-saving interventions in pediatric trauma: a National Trauma Data Bank Experience. Journal of trauma and acute care surgery 87, no. 6, pp. 1321-1327.

[8] Tengs, Tammy O., Miriam E. Adams, Joseph S. Pliskin, Dana Gelb Safran, Joanna E. Siegel, Milton C. Weinstein, and John D. Graham. 1995. Five-hundred life-saving interventions and their cost-effectiveness. Risk analysis 15, no. 3, pp. 369-390.