An Analysis of Speech as a Modality for Activity Recognition during Complex Medical Teamwork

Swathi Jagannath[1], Aleksandra Sarcevic[1], and Ivan Marsic[2]

[1]Drexel University
Philadelphia, PA, United States
{sj532, aleksarc}@drexel.edu

[2]Rutgers University
Piscataway, NJ, United States
marsic@rutgers.edu

ABSTRACT

We analyzed the nature of verbal communication among team members in a dynamic medical setting of trauma resuscitation to inform the design of a speech-based automatic activity recognition system. Using speech transcripts from 20 resuscitations, we identified common keywords and speech patterns for different resuscitation activities. Based on these patterns, we developed narrative schemas (speech "workflow" models) for five most frequently performed activities and applied linguistic models to represent relationships between sentences. We evaluated the narrative schemas with 17 new cases, finding that all five schemas adequately represented speech during activities and could serve as a basis for speech-based activity recognition. We also identified similarities between narrative schemas of different activities. We conclude with design implications and challenges associated with speech-based activity recognition in complex medical processes.

CCS CONCEPTS

• **Human-centered computing~Activity centered design** • Computing methodologies~Activity recognition and understanding

KEYWORDS

Speech analysis; speech modeling; narrative schema; activity recognition; decision support; emergency medicine.

1 INTRODUCTION

Trauma is the leading cause of death and disability in children and young adults. Early care after injury has an important impact on outcome [6,28], making the initial management of injured patients (*trauma resuscitation*) a critical phase in their care. A standardized evaluation protocol (Advanced Trauma Life Support [ATLS] [1]) has been shown to improve patient outcomes, but errors and process deviations persist [3,8]. While most deviations represent permissible variations, up to 40% have been classified as errors associated with adverse outcomes, including long-term disability and death [7,12].

To reduce the number of errors during trauma resuscitation, prior research has implemented real-time computer-aided decision support [5,8]. These initial systems have had limited success because they either require manual data entry or use automatic but incomplete data about the process (e.g., patient data from vital sign monitors but no data about team activities). Other clinical settings have experimented with automated activity data capture using environmental sensors in an attempt to infer workflow from these data and identify process deviations [9,19,30]. Common approaches have included radiofrequency identification (RFID) tags to track people and objects [16,30], or integration of low-level activity data from multiple sensors to infer high-level activities [17].

Much of the resuscitation process, however, relies on speech [2,25]. Verbal communication plays a key role in team situational awareness during resuscitations [32], and involves assigning tasks, requesting, sharing or confirming information, and reporting activity completion. This content-rich speech is a useful source for activity recognition because it has unique information that cannot be captured by other sensor modalities, like computer vision or RFID. Even so, using speech as an input for activity recognition-based decision support is challenging for two reasons. First, our understanding of the nature and structure of speech during complex medical processes such as trauma resuscitation is still incomplete. Prior studies have shown the challenges in using speech during medical work [2,24], but speech structures and models in these processes remain understudied. Second, verbal exchanges during trauma resuscitations are often interleaved due to parallel activity performance, further complicating the use of speech as a clue for automatic activity recognition. Previous work on language modeling used coherent textual content from simple everyday activities [18,20,21,22]. The context of medical emergencies with distinct, yet grammatically incorrect verbal communication provides new opportunities for language modeling.

PervasiveHealth '18, May 21–24, 2018, New York, NY, USA

ACM ISBN 978-1-4503-6450-8/18/05$15.00
https://doi.org/10.1145/3240925.3240941

Our long-term research goal is to leverage speech as a powerful clue for automatic activity recognition, and alert trauma teams to errors and process deviations by developing an activity recognition-based decision-support system. In this paper, we focus on understanding and modeling speech by addressing three research questions: (1) How is speech structured during trauma resuscitation? (2) How do speech workflows differ based on activity types? (3) How representative are the speech "workflow" models of conversation during activities? To answer these questions, we selected five most frequently performed activities during trauma resuscitation—Blood Pressure (BP) Check, Pupil Examination, Exposure Assessment, Neurological Score (Glasgow Coma Score) Calculation, and Intravenous (IV) Placement. We used transcripts from 20 actual resuscitations to first identify speech patterns and commonly occurring keywords for each of the five activities. We then constructed narrative schemas—representations of *speech workflow models* during an activity performance, which have some similarity to *speech acts* [31]. Narrative schemas are critical because they provide a conceptual basis for an activity recognition system. Additionally, we extended the event notation described by Pichotta and Mooney [21] to better represent the structure of speech in our domain. To evaluate our schemas, we used 17 new cases and analyzed the extent to which these schemas needed modifications to account for the previously unseen cases, as well as how were those modifications distributed across activities.

Our findings showed that the identified speech patterns and keywords for each activity could serve as differentiators for activity-related verbal communication in a dynamic medical setting. We also observed that the narrative schemas were stable for all five activities, with only minor modifications triggered during evaluation. Similarities were found between the narrative schemas of different activities, which could simplify speech-based activity recognition. Major challenges for modeling speech to facilitate speech-based activity recognition included incomplete or unintelligible communication, concurrent and interleaved activities, planned but abandoned activities, and repeated or possibly unsuccessful activity performances. We contribute to pervasive health and computing literature by describing design and research considerations for developing narrative schemas to facilitate speech-based automatic activity recognition in complex medical processes.

2 BACKGROUND: DOMAIN DESCRIPTION AND TERM DEFINITIONS

The goal of trauma resuscitation is to rapidly stabilize the patient, identify major injuries, and develop a treatment plan. Trauma teams follow the ATLS protocol to achieve a reliable diagnosis by ordering and prioritizing resuscitation activities. The protocol consists of two phases: the primary and secondary surveys. In the primary survey (coined ABCDE), the team evaluates the patient's Airway, Breathing, Circulation, Disability (neurological assessment) and Exposure (patient is disrobed for identifying injuries that may not be evident initially). The secondary survey is performed after the patient had been stabilized and includes a detailed head-to-toe evaluation to identify other injuries. While the protocol allows for a hierarchically-ordered process, the execution of the protocol depends on changes in patient status, requiring some activities to be repeated. Trauma teams consist of seven to 15 providers from multiple 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. Additional providers may be called based on the extent and nature of patient injury.

Team communication occurs at a high level, when discussing intentions and plans, and at a low level, when coordinating tasks and specifying task parameters [32]. Trauma team members, however, rarely name resuscitation activity while it is being performed. Rather than *describing* the activity, communication during resuscitation *supports* the activity. Activities are discussed and reported before the execution, during the planning and preparation, or after the completion of the activity to report the results. This communication pattern requires drawing associations between what the team is *saying* and what the team is *doing*, as well as defining these associations based on the stages of activity performance (e.g., before-during-after the activity or preparing-performing-assessing).

In the context of our work, we define *activity* as purposeful work that people do with their hands or eyes (observation). We consider speech as a facilitator of activity rather than being activity itself. For some activities, however, speech *is* the activity. For example, providers can ask the patient for their name (e.g., "Tell me your name!") to assess the patient's neurological status or if their airway is obstructed. We use the term *story* to represent all verbal communications related to a *single performance* of an activity. If an activity is performed multiple times, each performance may result in a different story. For example, the BP Check activity was performed twice in one of the cases we analyzed, resulting in two stories at two different times (Figure 1). A *speech-event* roughly corresponds to each line of speech in a story. For example, "One ten over sixty-eight" (line 25) and "Oh see…lot better blood pressure" (line 224) represent *speech-events* in the first and second story, respectively (Figure 1).

A set of stories capturing the conversation during different instances of a given type of activity is represented as a single *narrative schema*—one narrative schema per one activity [4]. A narrative schema is a conceptual summary of key events (in terms of speech) that occur in a story; it is a *pattern of events* (or lines of speech) that

Line #	Time	Speech
23	(0:03:02)	One ten over sixty-eight.
24	(0:03:04)	say again.
25	(0:03:05)	One ten over sixty-eight.
28	(0:03:15)	And that was manual.
		[…]
214	(0:20:38)	Right, cause I'm not, I didn't write that eighty over forty because I want to see if we get a better reading. I don't believe it.
217	(0:21:04)	It's okay, It's okay it's just hugging. It hug your arm and checks your pressure.
224	(0:22:43)	Oh see … lot better blood pressure.
225	(0:22:45)	Way better blood pressure, he's not moving, he's not dancing, he's calm.

Figure 1: All communications related to Blood Pressure Check activity in an example resuscitation. (case ID: 160649)

typically occur during an activity performance. A narrative schema provides a generalized workflow-type depiction of the stories. For the purpose of speech-based activity recognition, this depiction does not require any additional annotation (e.g., exclamation marks or attribution to different team members) because speech recognizer cannot interpret different expressions and tones as punctuation marks. Depending on different contexts and findings during an activity, the course of the activity may take different turns. We represent these different storylines during different performances of the same activity as transitions between events in the narrative schema.

3 RELATED WORK

3.1 Speech-Based Activity Recognition

To our knowledge, little research exists on speech-based activity recognition. Stork *et al*. [27] proposed a method for recognizing daily human activities in the kitchen and bathroom contexts by using sounds produced during these activities. Giannakopoulos and Siantikos [11] developed an activity recognition system for elderly monitoring that uses non-verbal information from the audio channel. Other research on acoustic-based activity recognition followed similar approaches [15,29]. This prior work, however, focuses on audio signals and sounds from sensors, and has not used textual content from speech to recognize activities.

More recently, several studies used deep learning techniques to predict intentions from speech [13] and detect medical phases during trauma resuscitation [14]. These studies have focused on deriving the meaning of the sentences using feature extraction from speech logs. Unlike this prior work, our goal is to understand speech patterns and develop models of speech to inform the design of a speech-based activity recognition system.

3.2 Modeling of Speech Patterns

Related research in speech pattern modeling includes language models for script inference. Orr et al. [20], Pichotta and Mooney [21,22], and Modi et al. [18] worked with stories about daily activities that were created by either crowdsourcing short and coherent descriptions with clear beginning and ending and no digressions [18], or breaking Wikipedia pages into paragraphs [21,22]. Although researchers have modeled situations based on real-world stories about activities, these stories were directly described in textual format, as opposed to speech associated with activities [10,20,23]. In contrast, our speech models use stories that are based on actual conversations heard before, during and after activity performance.

In this paper, we model patterns of speech by constructing narrative schemas using the event notation as described by Pichotta and Mooney [21]. This representation provides a standardized language modeling approach for understanding the structure of speech. Specifically, it divides every *speech-event* associated with an activity into elements known as parts of speech, allowing classification of the keywords as verb, subject and object. This classification in turn enables the speech recognition system to map the commonly occurring keywords to elements in the event notation. Pichotta and Mooney [21] depict an event using a *tuple* of five elements (v, e_s, e_o, e_p, p), where v = verb, e_s = subject, e_o = direct object, e_p = prepositional relations, and p = preposition relating v and e_p. For our domain, we extend the above event tuple by two elements—adjective and adverb. These two elements were critical to include because many keywords that characterize speech during trauma resuscitation fall under these two categories. Our seven-tuple event notation included:

$$\boldsymbol{(v, e_s, e_o, e_p, p, adj, adv)}$$

where v = verb, e_s = subject, e_o = direct object, e_p = prepositional relations, p = preposition relating v and e_p, *adj* = adjective, *adv* = adverb. By modeling speech patterns using text-based content derived from verbal conversations during complex teamwork, we show the kinds of research and design considerations that researchers must make when using speech for activity recognition.

4 METHODS

This study took place in a Level 1 trauma center of an urban, pediatric teaching hospital in the mid-Atlantic region of the United States. In addition to medical instruments and tools typically seen in the resuscitation rooms, the trauma bay at our research site has 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).

4.1 Data Collection

We collected 37 audio recordings captured in the trauma bay for patients that were admitted to the hospital over a period of 11 months. Injury mechanisms in the captured dataset ranged from fall injuries, motor vehicle crashes, pedestrians struck by motor vehicles, and gunshot wounds. The length of resuscitations was from 10 to 58 minutes, with an average duration of 27 minutes. We manually transcribed the audio recordings to represent every activity

and speech utterance in the resuscitation in a chronological order. For each utterance, the transcripts included timestamps, role uttering the speech, and roles to which speech was directed. On average, the transcripts consisted of 200 lines of speech (SD = ± 108). We removed roles from the transcripts because they are difficult to identify from speech, leaving only timestamps, transcribed communications and activities.

4.2 Data Analysis

We performed a three-step analysis corresponding to our three research questions. In *step one*, we used 20 transcripts and identified speech patterns and commonly occurring keywords for every observed activity "before", "during" and "after" the activity completion. In *step two*, we selected five most frequently performed activities, which also occur in different phases of the protocol and differ in complexity, duration, frequency, and number of steps involved. The five activities included (1) *Blood Pressure (BP) Check*—a repetitive assessment of patient blood pressure; (2) *Glasgow Coma Score (GCS) Calculation*—a complex, multi-step assessment of patient neurological status including verbal, visual, and motor responses; (3) *Pupil Examination*—a brief assessment of patient pupil size and reaction to light; (4) *Exposure Assessment*—a brief assessment of patient body temperature; and, (5) *Intravenous (IV) Placement*—a complex, multi-step activity for controlling patient circulation. Among these activities, BP Check, GCS Calculation, Pupil Examination, and Exposure Assessment are *assessment* activities performed to assess the patient status. IV Placement is a *control* activity performed to stabilize patient condition based on the assessments. We constructed narrative schemas for these five activities to provide a conceptual representation of key events during a single activity performance. We analyzed 20 transcripts line by line, adding new steps to the schema to represent key events as we encountered each story of an activity. We then applied our seven-tuple event notation to *speech-events* (i.e., speech sentences in activity performances) and classified the words in each to fit this event notation.

Once we constructed the initial narrative schemas, we "froze" the schemas and marked them as "standard" for these five activities. In *step three*, using the additional 17 transcripts, we assessed how well the standard schemas represented previously unseen resuscitation cases. We analyzed new transcripts line by line to determine if a sentence from a new transcript corresponded to an existing event in the standard schemas or a new event was needed. For each activity, we also examined if new tuples were needed and tracked the number of changes to the standard schemas. This evaluation helped us identify how much the standard schemas got affected by the modifications, if any, triggered by the new cases.

5 RESULTS

We first present the patterns of speech observed across all resuscitation activities, as well as commonly occurring keywords identified for the five selected activities. We then describe the standard narrative schemas for each of the five activities. Finally, we report the evaluation results.

5.1 Speech Patterns and Common Keywords

Most speech during trauma resuscitation is not in complete and grammatically correct sentences, but in short words and phrases. We identified three types of patterns based on our analysis: (1) activities with a definite speech pattern, (2) activities distinctly reporting the numerical results, and (3) activities reporting the results with specific keywords.

Activities with a definite speech pattern contain specific speech attributes that clearly indicate the activity being performed. Examples include GCS Calculation, Pupil Examination, and Breathing Assessment. We found that the activities in this category have the same lines of speech with little or no variation in wording before, during and after the activity performance across different resuscitation cases. In GCS Calculation, if the patient is conscious and obeying commands, the exam always starts with questions directed towards the patient, typically asking for their name or the day of the week, to assess their verbal ability. These questions are followed by requests to move extremities (e.g., "move your toes" or "squeeze my hand") to assess their motor abilities.

Activities distinctly reporting numerical results do not follow a definite pattern of speech, but they could be recognized by the unique way of reporting values of activity outcomes. Examples include BP Check, Exposure Assessment, and Heart Rate/Pulse Rate Check. In the example of BP Check activity, the results are always reported as a set of two numbers separated by the word "over". Sometimes, the values are accompanied by identifier words such as "BP", "blood pressure", "manual", or "cuff", or reported with other vitals, like heart rate and oxygen saturation. In some examples, reporting of the BP values alone was triggered by a request, (e.g., "what is the blood pressure?"), which could also act as an identifier.

Activities with specific keywords for reporting the results typically contain little to no speech before or during the activity. Examples include Abdomen and Pelvis Examination, Ear Examination, Chest Examination, Peripheral and Central Pulse Check. These activities can be recognized by the specific keywords found in the reports. The results of patient Abdomen and Pelvis Examination are reported as "Abdomen is soft and non-distended" and "Pelvis is stable", where "soft", "non-distended", and "stable" represent activity-specific keywords.

The speech patterns allowed us to derive common keywords for the activities and understand the nature of speech before, during, and after the activity is performed

Table 1: Common keywords/phrases for BP Check, GCS Calculation, Pupil Examination, Exposure Assessment, IV Placement before, during and after activity performance. No speech was observed *during* BP Check, Pupil Exam and Exposure Assessment.

	Blood Pressure Check	**GCS Calculation**	**Pupil Examination**	**Exposure Assessment**	**IV Placement**
Before	blood pressure, BP, pressure, heart rate, saturation, BP cuff, vitals, manual cuff, pressure cuff	GCS	open your eyes, shine, light, pupils	temperature, temp, exposure	IV access, left, right, IV, line
During	--	open your eyes, what is your name, what is your date of birth, squeeze my hand, wiggle your toes, move your hand, lift your leg, remember	--	--	IV, pinch, hurts
After	blood pressure, heart rate, saturation, BP, pressure, manual, (value)	GCS, Glasgow coma score, (value)	pupil, equal, reactive, bilateral, minimal, sluggish, brisk, millimeters, (value)	temperature, temp, tympanic, axillary, oral, (value)	IV access, left, right, IV, line, gauge, fluids, (value)

(Table 1). For example, only a few activities contain speech during performance. Keyword analysis also highlighted the unique vs. common keywords across activities. For instance, "bilaterally" is used in more than one activity, including Breathing Assessment, Pupil Examination, and Ear Examination. Keywords like "reactive" and "sluggish" are unique for Pupil Examination. Understanding the keywords provided us with a platform for analyzing the structure of *speech-events*.

5.2 Narrative Schemas

We constructed narrative schemas for five different activities based on 20 transcripts of actual resuscitations (Figure 2, Figure 3, and Figure 4). The rounded rectangle boxes in the schemas represent key narrative events such as information requests or reports, and the diamond shaped boxes represent decision-making events such as assessment during an activity performance. The arrows between the boxes indicate possible transitions between key events in a single performance of an activity. Because transitions between events depend on the changes in patient condition, the events in any story may occur in any order as indicated by the arrows. The vertical double-headed arrow indicates the actual activity performance in the workflow (Figure 2). Below we describe each narrative schema in detail.

Blood Pressure (BP) Check activity occurs during the Circulation Assessment phase of the primary survey when a bedside nurse places an automatic BP cuff on the patient to obtain the BP value. Blood pressure is one of the vital signs, along with temperature, heart and respiratory rates, and oxygen saturation, that is repeatedly measured throughout the resuscitation. The standard narrative schema for BP Check activity consists of six key events (Figure 2): (1) requesting to measure the patient's BP, (2) talking to the patient to check BP, (3) reporting on progress of BP measurement, (4) reporting on the measured value of BP, (5) requesting a (repeated) report or additional clarification of BP value, and (6) assessing the measured BP value. Most of the time, the bedside nurse performs BP Check after talking to the patient to check BP event, as indicated by the double-headed red arrow. In some cases, a trauma team member may request to check the patient's BP. The nurse may also report the BP without the request to measure, talking to the patient or reporting the progress of BP measurement. Generally, due to the noisy nature of the setting, the report may be inaudible, hence a request to report the previously reported BP value event might follow the BP report. At the end of the activity performance, the team leader may assess the measured BP value to determine if they need another measurement.

Glasgow Coma Score (GCS) Calculation activity is performed during the Disability Assessment phase of the primary survey. A physician surveyor assesses the patient's visual, verbal and motor abilities based on their responses to the surveyor's commands. For example, if the patient is conscious, the surveyor may ask them to open their eyes, answer specific questions or move their arms and legs to assess visual, verbal and motor abilities, respectively. The patient is given a neurological score ranging from 3-15, where 15 is a fully alert patient and 3 is an unresponsive patient. GCS Calculation activity is normally not repeated unless there is a significant change in the patient condition. GCS Calculation activity is speech-intensive because it has multiple steps and may take longer to complete. The standard narrative schema for GCS Calculation consists of five key events (Figure 3): (1) requesting to measure the patient's GCS value, (2) talking to the patient to determine GCS value, (3) reporting on the calculated GCS value, (4) requesting a (repeated) report or additional clarification of GCS value, and (5) assessing the calculated GCS value. GCS Calculation is also one of the few activities that contain speech during the activity performance. The surveyor performs this activity while they are talking to the patient to determine GCS value event. Similar to BP Check activity, the surveyor reports the GCS value after the activity is performed and other team members may request additional clarification. Occasionally, the team leader may initiate the activity by requesting the patient's GCS value. At the end of the activity performance, the team leader assesses the value to determine if the calculation is appropriate. If the assessment is unreliable, the activity is repeated to confirm the calculation.

Pupil Examination is also performed during Disability Assessment phase of the primary survey, either with or

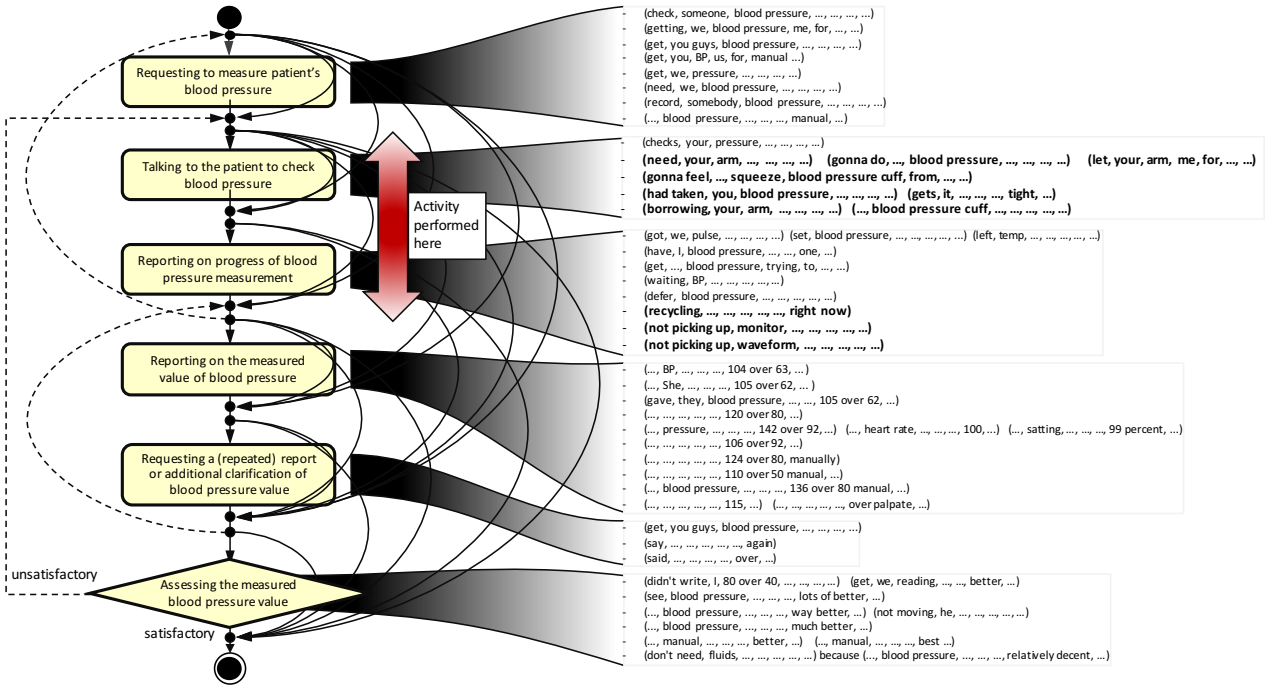

Figure 2: Narrative schema for BP Check activity with "speech-events" (left) and example utterances for each event represented in seven-tuple notation (right). Any event may transition to any subsequent event because of incomplete (lacking or nonverbal) communication or problems with speech recognition. New event tuples added during evaluation of the narrative schemas using 17 new resuscitations are indicated in bold text.

immediately after GCS Calculation. During the exam, a physician surveyor assesses the patient pupils' condition and responsiveness to light. This activity is performed only once and there is less speech associated with it. The standard narrative schema for pupil examination consists of four events (Figure 3): (1) requesting to perform pupil examination, (2) talking to the patient to perform examination, (3) reporting the results, and (4) requesting a (repeated) report or additional clarification of results. The team leader usually initiates the exam with a request and the surveyor performs the activity after talking to the patient by asking them to open their eyes. The surveyor reports the results immediately after the examination. In some cases, other providers in the team may request an additional clarification on the reported value. The schema for this activity does not contain a decision-making event because we did not observe any events associated with assessment of pupil results. However, the activity may be repeated if the patient condition changes.

Exposure Assessment is performed during the primary survey to measure the patient's temperature using a thermometer. A bedside nurse measures temperature multiple times during every resuscitation, along with other vital signs. The five key events in the standard narrative schema include (Figure 3): (1) requesting to measure the patient's temperature, (2) reporting on progress of measuring temperature, (3) reporting on the measured value of temperature, (4) requesting a (repeated) report or additional clarification of temperature value, and (5) assessing the measured temperature value. A team leader may request to measure the patient's temperature. Commonly, the bedside nurse performs the activity and reports the value along with other vital signs. We observed multiple requests to repeat previously reported values because this activity is performed several times. The assessment event does not have a decision point leading to repeat the measurement within this activity because temperature assessment leads to controlling the patient temperature, an activity called Exposure Control (e.g., covering the patient with blankets).

IV Placement is performed in the Circulatory Control phase of the primary survey following the Circulatory Assessment activities such as BP, Pulse and Heart Rate Check. A thin tube is inserted into one of the patient's veins to administer fluids directly into the bloodstream. The standard narrative schema for this activity includes seven events (Figure 4): (1) inquiry on patient's pre-hospital IV status, (2) reporting on patient's pre-hospital IV, (3) decision on adequateness of the pre-hospital IV, (4) requesting to get a new IV access, (5) reporting on the progress of getting an IV access, (6) reporting on completion of IV access, and (7) requesting an additional clarification on IV access. An important aspect of this activity is that the patient may already have an IV inserted before arrival to the hospital. The team leader initiates the activity by inquiring about the patient's pre-hospital IV status. A team member checks the IV access and reports the status to the team. If the patient already has an adequate pre-hospital IV, the team leader makes the decision to administer the fluids, which marks the end of the IV

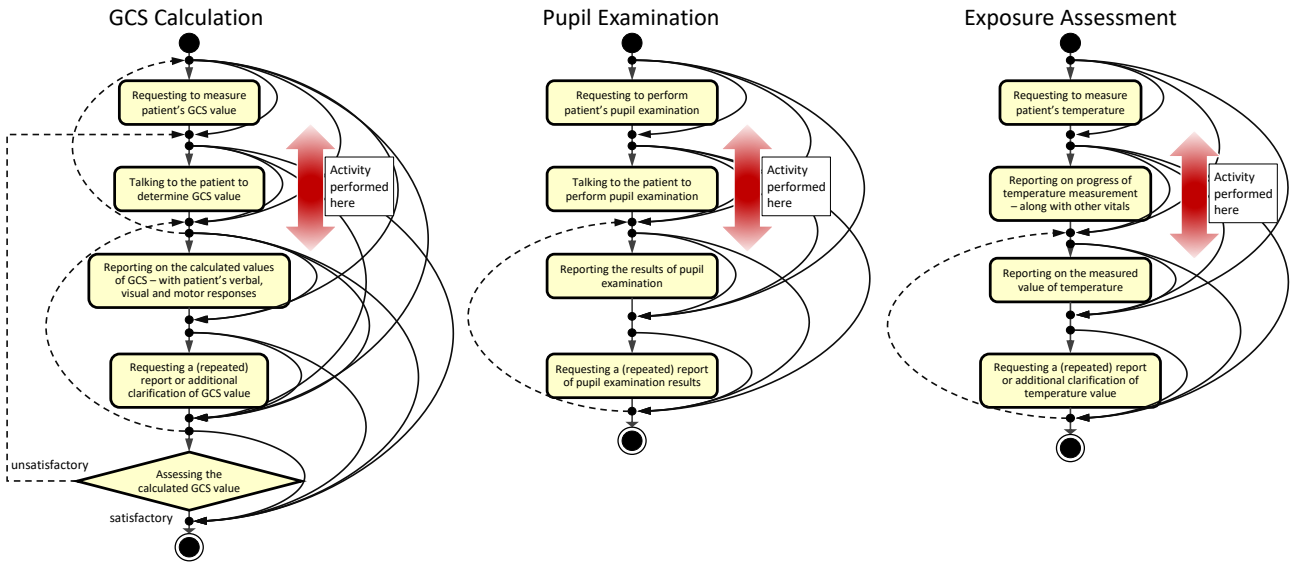

Figure 3: Narrative schemas for GCS Calculation, Pupil Examination and Exposure Assessment activities derived from 20 resuscitations.

Placement activity. If the pre-hospital IV is inadequate, the team leader requests the bedside nurse to establish a new IV access. The nurse continues to report on the progress of the IV placement until the completion is successful. Similar to other activities, other providers in the room may ask for a clarification on the IV access.

Next, we analyzed the structure of sentences using the seven-tuple event notation. For example, the speech-event "Can we get a blood pressure?" for the BP Check activity is represented as *(get, we, blood pressure, ..., ..., ..., ...)* tuple (Figure 2). In this speech-event, the keyword is "blood pressure," which is the object of the sentence, "get" is the verb and "we" is the subject. Some BP Check speech-events also have entries for other tuples, such as "manually" for the adverb and "106 over 92" for the adjective (or a qualifier). Within an activity, we found that the adjective is the key element representing the reported value. For example, a blood pressure report event may take forms:

(..., ..., ..., ..., ..., 106 over 92, ...)

(..., blood pressure, ..., ..., ..., 136 over 60 manual, ...)

Because key speech-events are similar across all activities, we compared the tuples to identify distinguishing speech related to different activities. We found that most verbs were similar, such as *get* or *measure*, and only few were unique for a particular activity type. For example, *look* and *shine* were distinctive for Pupil Examination, and *squeeze* and *move* for GCS Calculation. It appears, however, that verbs can serve to identify the stage of activity performance. Verbs in *requests* were usually used in the imperative mood (Figure 2), while *reports* rarely used a verb. In our domain, the key element that distinguished activity types was the direct object. For example, a request to measure patient data may take forms:

(get, we, blood pressure, ..., ..., ..., ...)

(get, we, temperature, ..., ..., ..., ...)

Both of these speech-events correspond to *requesting to measure patient data* in BP Check and Exposure Assessment. The sentence verbs and subjects are the same, and the only distinguishing element is the object, represented by the keywords for the respective activities.

5.3 Evaluation of Narrative Schemas

To assess how representative are the narrative schemas of conversation during the activity, we evaluated them using 17 new transcripts. We found that the key events and workflow of the standard schemas remained stable for all five activities. The number of tuples, however, changed as new speech sentences were encountered in previously unseen transcripts. We modified the standard schemas derived from the initial 20 resuscitations by adding new tuples (bold text in Figure 2) based on two scenarios:

(a) New keywords occurred in different parts of speech in a tuple. For example, in BP Check activity, we added the verb "recycling" and the adverb "right now:"

(recycling, …, …, …, …, …, right now)

(b) Existing keywords occurred in a new combination within a tuple. For example, in Pupil Examination activity, we added:

(look, …, light, ..., ..., ..., ...)

Although both "look" and "light" already appeared in the schema tuples, they did not occur together in a single tuple.

Most new tuples were added because of new verbs (Table 2). The number of new tuples added to each schema did not correlate with the perceived complexity of the corresponding activity (Table 2). For example, BP Check had more modifications (11 new tuples) than GCS Calculation activity (one new tuple). On average, we added four new tuples, ranging from none for Exposure Assessment to 11 for BP Check.

Table 2: Number of new tuples added to the narrative schemas of activities during evaluation with 17 new resuscitation cases.

Activity type	# of existing tuples	# of new tuples added	# of new verbs added	# of new subjects	# of new direct objects added	# of prepositional relations added	# of new prepositions	# of new adjectives	# of new adverbs
Blood Pressure Check	42	11	8	3	2	1	1	1	1
Pupil Examination	24	5	3	1	0	0	0	1	0
GCS Calculation	50	1	0	0	1	0	0	0	0
IV Placement	50	1	0	0	0	0	0	1	0
Exposure Assessment	35	0	0	0	0	0	0	0	0

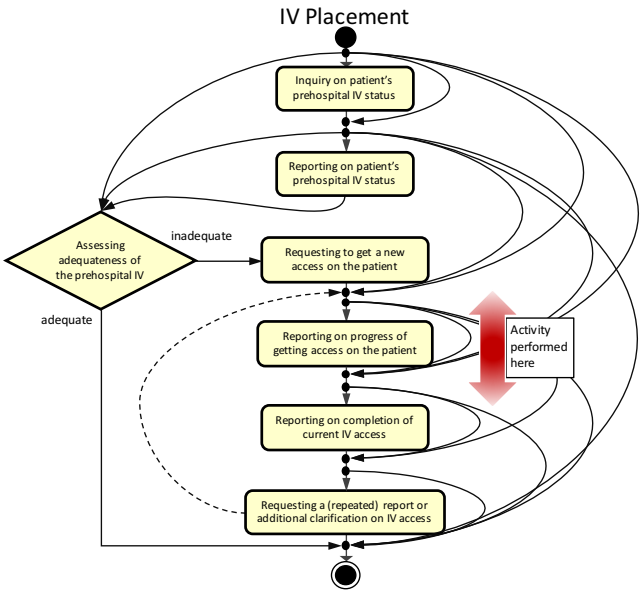

Figure 4: Narrative schema for IV Placement activity.

We observed that the number of new tuples added because of new object keywords was relatively low: only three (Table 2). Given that in our domain the key element that distinguishes between activities is the direct object, this finding shows that our initial schemas remained stable even with the new cases. Most changes were made because of new verbs, which generally do not serve as differentiators of activities (except for some verbs such as shine or move).

6 DISCUSSION

The results from our analysis of speech patterns, structure and workflow during trauma resuscitation provided several insights into speech as a sensor modality for activity recognition during a complex medical process. Unlike other sensor modalities, such as computer vision and RFID, speech can be used to extract rich and finer-grain information about ongoing activities. Synchronous modalities like RFID and imaging capture only information about the current status of activity performance (e.g., not-performed versus performed). Our results showed that speech can provide information such as activity type, activity stage (preparation, performance, assessment), and even activity content (parameters, outcome and findings). Although we used a single site for this study, our results generalize to most U.S. trauma centers due to similarities in team structures and procedures, as well as the use of the same evaluation protocol [1]. From our prior work [25], we learned that the nature of team communications in an adult trauma center did not differ much from that of a pediatric site. The only notable distinction was that nurses talked more frequently to pediatric patients than adults, to keep them calm and informed. Below we discuss design and research implications for using speech as part of an activity recognition system.

6.1 Design Implications for a Speech-Based Activity Recognition System

Verbal communication associated with activity performance during trauma resuscitation exhibits specific speech patterns and keywords that could facilitate activity recognition. Similar patterns can be expected for other knowledge-based processes. Our findings showed only minor differences between narrative schemas for the four assessment activities. In addition, when comparing the narrative schemas of assessment and control activities, we found that the only major difference is the decision-making event. In assessment activities, the decision is made after the activity performance, while in control activity decision is made early to allow for an action based on the decision. The presence of common key events across activities will allow the activity recognition system to focus on the key differences between each schema, thereby simplifying the activity recognition. We also found that both narrative schemas and their speech-event tuples for all five activities needed only minor modifications when previously unseen cases were analyzed. This finding suggests that the identified standard schemas accurately represent the activity-related speech and can be expected to remain stable when new cases are considered.

We identified keywords specific for both activity type and activity content in three different stages of performance. These keywords can be used to differentiate between the activities. Because speech-event tuples will become available incrementally during process performance, the system will not rely on matching complete schemas. Rather, the current activity type will be predicted from the observed keywords and available tuples (Figure 5). The associated schema for the predicted activity will be instantiated and its tuples will be filled out as new speech is captured. During this process, the correctness of the initial activity prediction can be checked and the activity parameters can be extracted from the speech-event tuples.

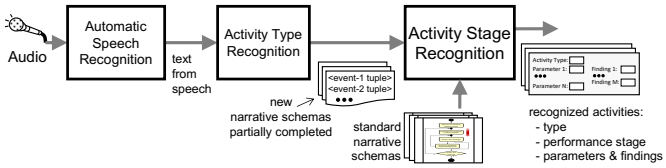

Figure 5: Proposed approach for recognizing activities from speech.

As an activity progresses from the first mention (e.g., planning) through performance to assessment, speech will be automatically captured and each speech-event will be used for constructing narrative schemas (Figure 5). When new speech-event tuples are captured, the narrative schemas for current activities will be continuously evaluated and modified, if necessary. The system will infer the stage of performance for each activity based on the tuples that were recognized from the speech heard so far.

Because resuscitation activities are mostly concurrent and interleaved, the system must also understand the relationships between the narrative schemas of different activities and the transition from one schema to another. To achieve this function, we need to consider dependencies between activities (e.g., performed together or in particular sequence), which may also depend on activity results. For example, a low or high blood pressure might lead to administering fluid or medication. The system, therefore, needs to represent the relationships between narrative schemas of different activities.

A speech-based activity recognition system would need the following information available to successfully recognize the activity type, activity stage, and its parameters: (a) new sentences that become available at random times to extract relevant parameters and construct partially or fully completed activity stories, (b) standard narrative schemas for all activity types, and (c) a representation of dependencies between different activities for appropriate transition between their narrative schemas.

6.2 Challenges in Speech Pattern Modeling

Our analyses uncovered several challenges associated with speech modeling and using verbal communication as a cue for automatic activity recognition. A major barrier for domains such as trauma resuscitation is the need to acquire large datasets of stories for different activities to allow for system training. Existing research on script modeling uses carefully crafted stories with clear beginning and ending, and with all content relevant to the core story [4,18,21,22]. Large numbers of such stories are usually acquired by crowdsourcing using Amazon Mechanical Turk [18] or using Wikipedia pages broken into paragraphs [21,22]. In contrast, acquiring such large number of stories for complex medical domains presents three challenges:

Incomplete (lacking or nonverbal) or unintelligible communications: Due to the constant movement of team members in the trauma bay, it is difficult to position the microphones to capture high-quality audio recording [2]. Recordings captured from the overhead microphones are noisy and contain overlapping conversations ("cocktail party effect"), which makes the transcribing process tedious, time-consuming and often requires domain expertise for correct transcription of medical terms.

Concurrent activities and intertwined stories: The "stories" are often interleaved due to concurrent activities, requiring manual processing to separate the stories and determine the beginning and ending of each story. A model of the process workflow could support predicting the flow of activities, but it would still require manual separation of activities. The Stanford Dependency Parser [26] could be used to parse the input text into speech-event tuples.

Planned but abandoned activities or repeated, possibly unsuccessful performances: During any resuscitation, the trauma team may consider an activity and discuss it, but decide not to perform the activity because it was deemed unnecessary. Sometimes, the trauma team may also perform an activity multiple times because the results were unsuccessful. The system may not be able to correctly identify such scenarios, either because of nonverbally communicated decisions or poor audio recordings resulting in inaccurate speech-to-text mapping.

7 CONCLUSION

This work described findings from speech analysis and modeling to assess the feasibility of using speech as a sensor for automatic activity recognition in a complex medical setting. We identified speech patterns for verbalizing activities during trauma resuscitation along with commonly occurring keywords. Using this knowledge, we developed narrative schemas (speech models) for five resuscitation activities and applied a linguistic modeling approach to understand the structure of sentences. We found similarities between the narrative schemas of different activities, even with a relatively small sample of resuscitation events. These findings suggest that both the flow of key events and sentence structure of standard narrative schemas are stable and adequately represent the ongoing activity, including its stage and content. Using and modeling speech for the purposes of activity recognition poses several challenges, including incomplete or unintelligible verbal communications, concurrent activities, planned but abandoned activities or repeated, possibly unsuccessful performances. Through continued research, we aim to explore how speech could be combined with other modalities like RFID, computer vision, and other sensors to support the future development of a decision-support system for complex medical teamwork.

ACKNOWLEDGMENTS

This research is supported by the National Science Foundation under Award No. 1253285 and by the National

Library of Medicine of the National Institutes of Health under Award No. R01LM011834. We thank Dr. R.S. Burd, Dr. O. Ahmed, and Y. Gu for their feedback and help.

REFERENCES

1. American College of Surgeons, Advanced Trauma Life Support® (ATLS®), 7th Edition, Chicago, IL, 2005.
2. 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 (Aug. 2005), 905-911. DOI: https://doi.org/10.1016/j.injury.2004.12.047
3. Elizabeth A. Carter, Lauren J. Waterhouse, Mark L. Kovler, Jennifer Fritzeen, and Randall S. Burd. 2013. Adherence to ATLS primary and secondary surveys during pediatric trauma resuscitation. *Resuscitation* 84, 1 (Jan. 2013), 66-71. DOI: https://doi.org/10.1016/j.resuscitation.2011.10.032
4. Nathanael Chambers and Daniel Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proc. Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 602-610.
5. John R. Clarke, Bonnie L. Webber, Abigail Gertner, Jonathan Kaye, and Ron Rymon. 1994. On-line decision support for emergency trauma management. In *Proc. Annual Symposium on Computer Application in Medical Care*, American Medical Informatics Association, 1028.
6. John R. Clarke, Stanley Z. Trooskin, Prashant J. Doshi, Lloyd Greenwald, and Charles J. Mode. 2002. Time to laparotomy for intra-abdominal bleeding from trauma does affect survival for delays up to 90 minutes. *J. Trauma and Acute Care Surgery* 52, 3 (Mar. 2002), 420-425.
7. Demetrios Demetriades, Brian Kimbrell, Ali Salim, George Velmahos, Peter Rhee, Christy Preston, Ginger Gruzinski, and Linda Chan. 2005. Trauma deaths in a mature urban trauma system: is "trimodal" distribution a valid concept? *J. Amer College of Surgeons* 201, 3 (Sep. 2005), 343-348. DOI: https://doi.org/10.1016/j.jamcollsurg.2005.05.003
8. Mark Fitzgerald, Peter Cameron, Colin Mackenzie, Nathan Farrow, Pamela Scicluna, Robert Gocentas, Adam Bystrzycki, Geraldine Lee, Gerard O'Reilly, Nick Andrianopoulos, Linas Dziukas, Jamie D. Cooper, Andrew Silvers, Alfredo Mori, Angela Murray, Susan Smith, Yan Xiao, Frank T. McDermott, Jeffrey V. Rosenfeld. 2011. Trauma resuscitation errors and computer-assisted decision support. *Archives of Surgery* 146, 2 (Feb. 2011), 218-225. DOI: https://doi.org/10.1001/archsurg.2010.333
9. Germain Forestier, Florent Lalys, Laurent Riffaud, Brivael Trelhu, and Pierre Jannin. 2012. Classification of surgical processes using dynamic time warping. *Journal of Biomedical Informatics* 45, 2 (Apr. 2012), 255-264. DOI: https://doi.org/10.1016/j.jbi.2011.11.002
10. Lea Frermann, Ivan Titov, and Manfred Pinkal. 2014. A hierarchical Bayesian model for unsupervised induction of script knowledge. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL-14)*, 49-57.
11. Theodoros Giannakopoulos and Georgios Siantikos. 2016. A ROS framework for audio-based activity recognition. In *Proc. 9th ACM Int'l Conf. Pervasive Technologies Related to Assistive Environments*. ACM Press, New York, 41. DOI: https://doi.org/10.1145/2910674.2935858
12. Russell L. Gruen, Gregory J. Jurkovich, Lisa K. McIntyre, Hugh M. Foy, and Ronald V. Maier. 2006. Patterns of errors contributing to trauma mortality: lessons learned from 2594 deaths. *Annals of Surgery* 244, 3 (Sep. 2006), 371-380. DOI: https://doi.org/10.1097/01.sla.0000234655.83517.56
13. Yue Gu, Xinyu Li, Shuhong Chen, Jianyu Zhang, and Ivan Marsic. 2017. Speech Intention Classification with Multimodal Deep Learning. Mouhoub M., Langlais P. (eds) *Advances in Artificial Intelligence. AI 2017*. Lecture Notes in Computer Science, Springer, Cham, vol 10233, 260-271. DOI: https://doi.org/10.1007/978-3-319-57351-9_30
14. Yue Gu, Xinyu Li, Shuhong Chen, Hunagcan Li, Richard A. Farneth, Ivan Marsic, Randall S. Burd. 2017. Language-based process phase detection in trauma resuscitation. In *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*, IEEE, 239-247. DOI: https://doi.org/10.1109/ICHI.2017.50
15. Lars Hertel, Huy Phan, and Alfred Mertins. 2015. Comparing time and frequency domain for audio event recognition using deep learning. In *Neural Networks (IJCNN), 2016 International Joint Conference on*, IEEE, 3407-3411. DOI: https://doi.org/10.1109/IJCNN.2016.7727635
16. Xinyu Li, Dongyang Yao, Xuechao Pan, Jonathan Johannaman, JaeWon Yang, Rachel Webman, Aleksandra Sarcevic, Ivan Marsic, and Randall S. Burd. 2016. Activity recognition for medical teamwork based on passive RFID. In *RFID (RFID), 2016 IEEE International Conference on*, IEEE, 1-9. DOI: https://doi.org/10.1109/RFID.2016.7488002
17. Xinyu Li, Yanyi Zhang, Jianyu Zhang, Moliang Zhou, Shuhong Chen, Yue Gu, Yueyang Chen, Ivan Marsic, Richard A. Farneth, and Randall S. Burd. 2017. Progress Estimation and Phase Detection for Sequential Processes. *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sep. 2017), 73. DOI: https://doi.org/10.1145/3130936
18. Ashutosh Modi, Ivan Titov, Vera Demberg, Asad Sayeed, and Manfred Pinkal. 2017. Modeling Semantic Expectation: Using Script Knowledge for Referent Prediction. Retrieved from: arXiv:1702.03121
19. Thomas Neumuth, and Christian Meißner. 2012. Online recognition of surgical instruments by information fusion. *International journal of computer assisted radiology and surgery* 7, 2 (Mar. 2012), 297-304. DOI: https://doi.org/10.1007/s11548-011-0662-5
20. John Walker Orr, Prasad Tadepalli, Janardhan Rao Doppa, Xiaoli Fern, and Thomas G. Dietterich. 2014. Learning Scripts as Hidden Markov Models. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (*AAAI-14)*. 1565-1571.
21. Karl Pichotta and Raymond J. Mooney. 2016. Learning Statistical Scripts with LSTM Recurrent Neural Networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*. 2800 – 2806.
22. Karl Pichotta and Raymond J. Mooney. 2016. Using sentence-level LSTM language models for script inference. In *Proc. 54th Annual Meeting of the Assoc. Computational Linguistics (ACL 2016)*. 279-289.
23. Michaela Regneri, Alexander Koller, and Manfred Pinkal. 2010. Learning script knowledge with web experiments. In *Proc. 48th Annual Meeting of the Assoc. for Computational Linguistics (ACL-10)*. 979-988.
24. Nicole K. Roberts, Reed G. Williams, Cathy J. Schwind, John A. Sutyak, Christopher McDowell, David Griffen, Jarrod Wall, Hilary Sanfey, Audra Chestnut, Andreas H. Meier, Christopher Wohltmann, Ted R. Clark, Nathan Wetter. 2014. The impact of brief team communication, leadership and team behavior training on ad hoc team performance in trauma care settings. *The American Journal of Surgery* 207, 2 (Feb. 2014), 170-178. DOI: https://doi.org/10.1016/j.amjsurg.2013.06.016
25. Aleksandra Sarcevic, Ivan Marsic, Michael E. Lesk, and Randall S. Burd. 2008. Transactive memory in trauma resuscitation. In *Proc. 2008 ACM Conference on Computer Supported Cooperative Work*. ACM, New York, NY, 215-224. DOI: https://doi.org/10.1145/1460563.1460597
26. Richard Socher, John Bauer, and Christopher D. Manning. 2013. Parsing with compositional vector grammars. In *Proc. 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*. vol. 1, 455- 465.
27. Johannes A. Stork, Luciano Spinello, Jens Silva, and Kai O. Arras. 2012. Audio-based human activity recognition using non-Markovian ensemble voting. In *RO-MAN 2012 IEEE*, IEEE, 509-514. DOI: https://doi.org/10.1109/ROMAN.2012.6343802
28. Homer C.N. Tien, Vincent Jung, Ruxandra Pinto, Todd Mainprize, Damon C. Scales, and Sandro B. Rizoli. 2011. Reducing time-to-treatment decreases mortality of trauma patients with acute subdural hematoma. *Annals of surgery* 253, 6 (Jun. 2011), 1178-1183. DOI: https://doi.org/10.1097/SLA.0b013e318217e339
29. Achyut Mani Tripathi, Diganta Baruah, and Rashmi Dutta Baruah. 2015. Acoustic sensor based activity recognition using ensemble of one-class classifiers. In *Evolving and Adaptive Intelligent Systems (EAIS), 2015 IEEE International Conference on*, IEEE, 1-7. DOI: https://doi.org/10.1109/EAIS.2015.7368798
30. Mithra Vankipuram, Kanav Kahol, Trevor Cohen, and Vimla L. Patel. 2011. Toward automated workflow analysis and visualization in clinical environments. *Journal of Biomedical Informatics* 44, 3 (Jun. 2011), 432-440. DOI: https://doi.org/10.1016/j.jbi.2010.05.015
31. Terry Winograd, and Fernando Flores. 1986. *Understanding Computers and Cognition.* Ablex Publishing Corp, Norwood, NJ.
32. Zhan Zhang and Aleksandra Sarcevic. 2015. Constructing awareness through speech, gesture, gaze and movement during a time-critical medical task. In *Proceedings of the European Conference on Computer-Supported Cooperative Work* (ECSCW '15). Springer, Cham, 163-182. DOI: https://doi.org/10.1007/978-3-319-20499-4_9