# Using Speech in Complex Teamwork to Support Activity Recognition

**Swathi Jagannath**
College of Computing &
Informatics
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104, USA
swathi.jagannath@drexel.edu

**Aleksandra Sarcevic**
College of Computing &
Informatics
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104, USA
aleksarc@drexel.edu

*CHI'17, Workshop on Interactive Systems in Healthcare (WISH)*
May 7, 2017, Denver, CO, USA.

## Abstract

We describe the results of speech patterns analysis during trauma resuscitation—a complex, error-prone medical process for treating severely injured patients—to inform future development of an activity recognition system that uses different sensors in the environment, including speech. The goal of the system is to support teamwork by detecting process deviations and alerting teams to errors. We contribute to system design and broader literature by performing human-centered analysis of speech and characterizing activities that are challenging to recognize through the speech modality.

## Author Keywords

Activity recognition; speech analysis; speech recognition; emergency medicine.

## ACM Classification Keywords

H.1.2 User/Machine Systems: Human information processing; I.2.7 Natural Language Processing: Speech recognition and synthesis

## Introduction

Trauma resuscitation involves care providers from multiple disciplines working together to treat severely injured patients in a high-risk, fast-paced setting. Although teams follow a standard evaluation protocol, errors and deviations are still observed, even among experienced teams [2]. Our long-term goal is to support this complex medical teamwork by developing a decision-support system that would automatically recognize activities and alert teams to errors in real time. To date, we have equipped the trauma room in our collaborating hospital with sensors, including Radio Frequency Identification (RFID) tags, microphone arrays and Kinect motion sensors. While activities performed with objects (e.g., listening to breath sounds with a stethoscope) can be recognized using RFID tags, activities without objects (e.g., patient palpation) pose a challenge for detection and recognition.

Recent studies have focused on activity recognition in various complex team settings via different modalities including video recordings [3], sensory cues from embedded and RFID sensors [1], and computer vision for identifying body posture, movement and location [5]. Most recently, speech recognition and natural language processing techniques have been combined to develop tools that leverage context-specific grammars and dictation models to interact with computers [6]. While there has been extensive research in activity- and speech recognition, none of these address the need for using speech as a sensor for automatic activity recognition in emergency medical settings.

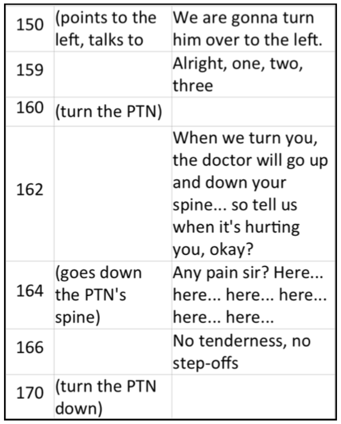
Because trauma teams rely on verbal communication to request or share information, assign tasks, or confirm activity completions, speech can be seen as a useful clue for activity recognition. The challenge, however, is that trauma team members usually do not call out a resuscitation activity while it is being performed. The activities are rather discussed and reported before the execution, during the planning and preparation, or after the completion of the activity to report the results. In this work, we seek to identify speech patterns for all resuscitation activities to leverage speech as a sensor for automatic activity recognition. Knowing these patterns will allow us to identify common keywords and phrases used before, during and after an activity is completed, which can in turn be used for keyword spotting. We conclude by discussing implications and challenges for using speech in designing a system for activity recognition during complex teamwork.

## Methods

We performed secondary-data analysis on 18 detailed transcripts and activity logs of actual trauma resuscitations that were observed over a three-month period at an urban, teaching hospital in the U.S. Northeast region. The transcripts were chronological and included every activity and utterance in the event. For each activity or utterance, we also had timestamps, roles performing the action or talking, and roles towards which the action or speech were directed. For the purposes of our analysis, we only used transcribed communications and activities, blocking out the information about roles or times. Similar to a microphone in the environment that only picks up the speech, we used the minimum amount of information to better understand the constraints and types of clues that speech offers for activity recognition. We also used an activity dictionary—a detailed log of activities developed by medical experts on our team—to ensure that we correctly matched activities from the transcripts to those performed by teams.

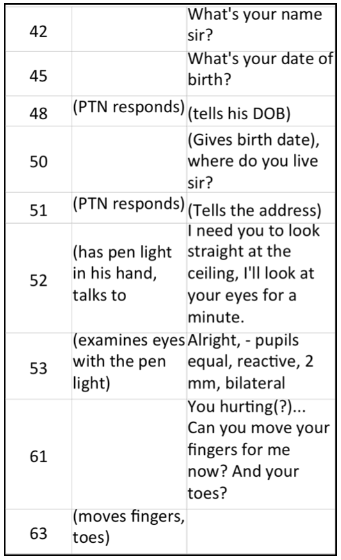
### *Data Analysis*

We began our analysis by categorizing the activities from the activity dictionary into “with objects” and “without objects.” This process of categorizing helped



| | | |
|---|---|---|
| 150 | (points to the left, talks to | We are gonna turn him over to the left. |
| 159 | | Alright, one, two, three |
| 160 | (turn the PTN) | |
| 162 | | When we turn you, the doctor will go up and down your spine... so tell us when it's hurting you, okay? |
| 164 | (goes down the PTN's spine) | Any pain sir? Here... here... here... here... here... here... |
| 166 | | No tenderness, no step-offs |
| 170 | (turn the PTN down) | |

Figure 1: Speech excerpt for back examination (Case 007). PTN denotes the patient.



| | | |
|---|---|---|
| 42 | | What's your name sir? |
| 45 | | What's your date of birth? |
| 48 | (PTN responds) | (tells his DOB) |
| 50 | | (Gives birth date), where do you live sir? |
| 51 | (PTN responds) | (Tells the address) |
| 52 | (has pen light in his hand, talks to | I need you to look straight at the ceiling, I'll look at your eyes for a minute. |
| 53 | (examines eyes with the pen light) | Alright, - pupils equal, reactive, 2 mm, bilateral |
| 61 | | You hurting(?)... Can you move your fingers for me now? And your toes? |
| 63 | (moves fingers, toes) | |

Figure 2: Speech excerpt for neurological exam (Case 001). PTN denotes the patient.

us identify activities that are performed without any RFID-tagged objects and depend on other modalities, like speech, for detection. We then identified the related speech excerpts in the transcripts and mapped them to the activities. In the next step, we identified speech patterns for different activities and extracted most common keywords for these activities. These keywords were classified into spoken before, spoken during, and spoken after the activity was performed. We next present the findings from this analysis.

## Findings

We have identified three types of patterns based on our analysis of speech during trauma resuscitation: activities with a definite speech pattern, activities distinctly reporting the numerical results, and activities reporting the results with specific keywords.

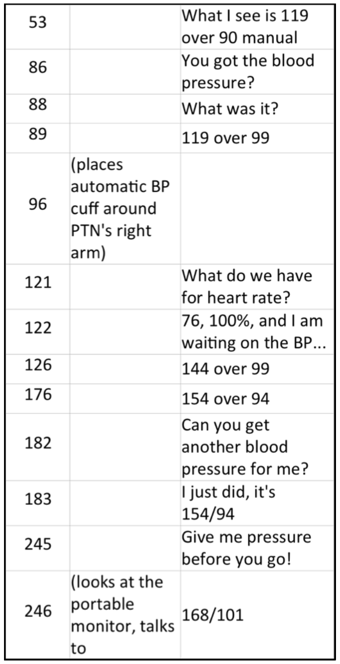
*Activities with a Definite Speech Pattern*

Activities with a definite speech pattern use specific speech attributes that clearly indicate what activity is being performed. We have identified several such activities, including the back examination, neurological exam, pupil examination and breathing assessment. In the interest of space, here we provide details about the first two activities.

*Back Examination*: This activity does not involve any objects and is typically performed during the secondary patient survey, once the initial (primary) evaluation is completed and major injuries had been identified and managed. The patient is first rolled on their side and the bedside physician then performs the exam by palpating the patient's back. Figure 1 shows an example of a conversation before, during and after back examination. We have found that back examination always starts with a person asking for patient roll on count "one, two, three." If the patient is conscious, they are alerted that a back examination is being done and the doctor is touching the patient's back. The bedside physician then starts palpating down the patient's spine continuously asking, "Any pain?" or "Does it hurt?" at the location of palpation (Line 164, Figure 1). Sometimes, the physician also mentions what area they are palpating. The result of the back examination is reported after it is completed (Line 166, Figure 1), at which point, the patient is rolled back.

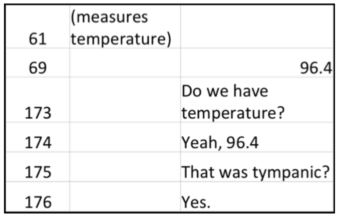
*Neurological Exam*: This is another activity that does not use any objects. Figure 2 shows an excerpt for calculating Glasgow Coma Score (GCS), a value that indicates the patient's neurological status by measuring verbal, eye and motor functions. This activity is part of the initial patient survey to gauge the severity of an acute brain injury, if any. If the patient is conscious and obeying commands, the exam starts with a question directed to the patient, typically asking for their name or the weekday to assess their verbal skills (Line 42, Figure 2). There may be other questions about date of birth or if they remember anything about the accident (Lines 45, 50, Figure 2). These questions are followed by requests to move extremities (e.g., "move your toes" or "squeeze my hand") to assess their motor abilities (Line 61, Figure 2). In several cases, we observed that pupil examination was done along with the neurological exam or immediately following (Lines 52, 53, Figure 2) as a parallel activity with an object. We discuss overlapping activities later in the paper.

As illustrated in the above examples, we have found that the activities in this category use the same lines of speech with a little or no variation before, during and



| | | |
|---|---|---|
| 53 | | What I see is 119 over 90 manual |
| 86 | | You got the blood pressure? |
| 88 | | What was it? |
| 89 | | 119 over 99 |
| 96 | (places automatic BP cuff around PTN's right arm) | |
| 121 | | What do we have for heart rate? |
| 122 | | 76, 100%, and I am waiting on the BP... |
| 126 | | 144 over 99 |
| 176 | | 154 over 94 |
| 182 | | Can you get another blood pressure for me? |
| 183 | | I just did, it's 154/94 |
| 245 | | Give me pressure before you go! |
| 246 | (looks at the portable monitor, talks to | 168/101 |

Figure 3: Speech excerpt for blood pressure check (Case 008). PTN denotes the patient.



| | | |
|---|---|---|
| 61 | (measures temperature) | |
| 69 | | 96.4 |
| 173 | | Do we have temperature? |
| 174 | | Yeah, 96.4 |
| 175 | | That was tympanic? |
| 176 | | Yes. |

Figure 4: Speech excerpt for exposure assessment (Case 002)

after the activity is completed. The common keywords for these activities are shown in Table 1.

*Activities Distinctly Reporting Numerical Results*
Activities distinctly reporting numerical results do not follow a definite pattern of speech, but could be recognized by the specific way the values are reported. We have identified many activities including blood pressure (BP) check, exposure assessment, oxygen saturation, and heart rate/pulse rate check. We use the first two activities to illustrate this pattern.

*Blood Pressure (BP) Check*: This activity is performed multiple times throughout the resuscitation process. Blood pressure is checked to assess circulation and it involves multiple objects including stethoscope and BP monitors. Figure 3 shows a speech excerpt for this activity. We observed that the blood pressure values are sometimes accompanied by identifier words such as “BP” or “blood pressure” or “manual” or “cuff” (Line 53, Figure 3). There are also examples of blood pressure values reported without any identifiers, but these usually follow a request, (e.g., “what is the blood pressure?”), which could instead act as an identifier (Lines 86, 182, 245, Figure 3). Blood pressure is also often reported with other vitals, like heart rate and oxygen saturation (Line 122, Figure 3). The unique feature of the blood pressure value is that it always has a set of two numbers separated by the word “over” (Lines 126, 176, Figure 3).

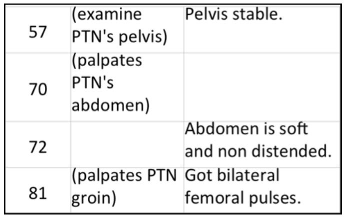
*Exposure Assessment*: Similar to BP check, this activity is performed multiple times throughout the process using thermometer. During this activity, the patient’s temperature is measured and environmental exposures are reviewed. Figure 4 shows a speech excerpt for this activity. Temperature values are sometimes accompanied by “temperature” or “temp” or preceded by a request for temperature (Line 173, Figure 4). We also observed that the temperature was the only numerical value that could be a decimal number. Although these activities do not use a definite speech pattern or specific keywords, it would still be possible to recognize these activities by their unique way of reporting numbers.

*Activities Reporting the Results with Specific Keywords*
Activities with specific keywords for reporting the results typically use less speech to no speech before or during the activity. Several activities fall into this category, including abdomen and pelvis examination, ear examination, chest examination, peripheral pulse check and central pulse check. We use the first two activities to exemplify this pattern.

*Abdomen and Pelvis Examination:* This activity does not involve any objects. The patient’s abdomen and pelvic

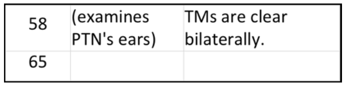
| | **Back Examination** | **Neurological Examination** |
|---|---|---|
| *Before* | turn, roll, on my count, one two three, back, exam, hurts | no speech |
| *During* | does it hurt, squeeze, spine, press, up and down, lower thoracic | open your eyes, what is your name, what is your date of birth, squeeze my hand, wiggle your toes, move your hand, lift your leg |
| *After* | roll back, count, one two three, gross, ecchymosis, step offs, vault, prostate, tenderness | GCS, glasgow coma score, values |

Table 1: Common keywords/phrases for back examination and neurological examination.



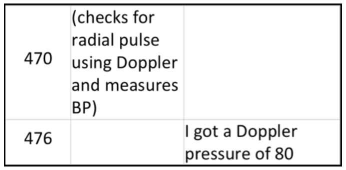
| | | |
|---|---|---|
| 57 | (examine PTN's pelvis) | Pelvis stable. |
| 70 | (palpates PTN's abdomen) | |
| 72 | | Abdomen is soft and non distended. |
| 81 | (palpates PTN groin) | Got bilateral femoral pulses. |

Figure 5: Speech excerpt for abdomen and pelvis examination (Case 003). PTN denotes the patient.



| | | |
|---|---|---|
| 58 | (examines PTN's ears) | TMs are clear bilaterally. |
| 65 | | |

Figure 6: Speech excerpt for ear examination (Case 012). PTN denotes the patient.



| | | |
|---|---|---|
| 470 | (checks for radial pulse using Doppler and measures BP) | |
| 476 | | I got a Doppler pressure of 80 |

Figure 7: Speech excerpt for blood pressure check using Doppler (Case 017)

area are visually inspected and palpated to identify any injuries, and the results are reported once the activity is completed (Lines 57, 70, 72, Figure 5).

*Ear Examination*: This activity is performed to check injuries in patient's ears using an otoscope. As shown in Figure 6, when the patient's ears are examined, the results are reported using specific keywords, e.g., "TMs [tympanic membranes] clear" or "clear bilaterally."

As illustrated in the examples, these activities use less speech, but they can be recognized using the specific keywords found in verbal result reports. Common keywords for these activities are shown in Table 2.

## Design Implications and Challenges

This paper describes the preliminary findings from speech analysis of trauma resuscitation transcripts to leverage speech as a sensor for automatic activity recognition. We have found three types of speech patterns that could be incorporated into the design of an activity recognition system. Here we discuss a few design implications.

We have identified commonly used keywords for activities before, during and after their completion. These keywords could be used for constructing narrative schemas for resuscitation activities [4]. First, for each of the resuscitation activity, all possible "standard" narrative schemas could be constructed and built into the system. The system would then dynamically construct narrative schemas at runtime and continuously compare them to the "standard" schemas to recognize the activity. Furthermore, once the activity has been recognized, the system could also detect potential process deviations. For some of the resuscitation activities, we also found common phrases that are used before the activity is started. The narrative schemas along with the commonly used phrases could be used to indicate the beginning of the activity, completion status of activities, and time to complete the activities, thus supporting better decision making process. Also, the pattern with a distinct way of reporting numerical results not only indicates the completion of the activity, but also the outcome of the activity.

Next, we turn to the design challenges for using speech as a sensor for automatic activity recognition. First, several of these activities are almost always overlapping. It is challenging to recognize the activity when there are interleaved requests without a clear speech attribute. Second, in our analysis, we used transcripts that were previously transcribed by a human. However, in a real-world scenario, there will be microphones in the trauma room and on people performing activities. Occasionally, important parts of the conversation could be missed for various reasons, like technical difficulties, people talking in a low voice, or too many people talking at the same time. It is difficult to recognize the activity if the missed part of

| | **Abdomen and Pelvis Examination** | **Ear Examination** |
|---|---|---|
| *Before* | pelvis stability, belly, hurt, press | no speech |
| *During* | no speech | left ear, right ear |
| *After* | abdomen, distended, non distended, soft, belly, pelvis, stable, seatbelt, pulse, abrasion, tender, DP, fracture, intact | bilateral TM, clear, blood, bleeding, wax |

Table 2: Common keywords for abdomen and pelvis examination and ear examination.

the conversation is the sole indicator of the activity. For example, if the word "over" in "154 over 94" (Line 176, Figure 3) is skipped, there is no other indicator of this statement being a BP value. Third, at times, different devices may be used for examination. Because reporting values changes based on the device, it will be challenging to recognize the activity since the pattern to report the result varies. For example, in Figure 7, we see that the blood pressure was reported as a single value rather than the standard, value-over-value when a Doppler was used. Although these cases do not happen often, it is important to consider these anomalies for the design purposes.

## Conclusion

This work describes our initial findings from a speech analysis to leverage speech as one of many sensors for automatic activity recognition in an emergency medical setting. We have identified three types of patterns for verbalizing the resuscitation activities, and discussed implications and challenges for designing a speech-based, automatic activity recognition system. Through ongoing research, we continue to explore how speech could be combined with additional modalities like video, computer vision, and other sensors to support the future development of a decision-support system for a complex medical setting such as trauma resuscitation.

## Acknowledgements

This research is supported by the National Science Foundation under Award Number 1253285, and partially supported by the National Library Of Medicine of the National Institutes of Health under Award Number R01LM011834. We thank Ivan Marsic, Randall. S. Burd, Omar Ahmed, Richard Farneth, and Yue Gu for offering their expertise during the analysis.

## References

1. Jakob E. Bardram, Afsaneh Doryab, Rune M. Jensen, Poul M. Lange, Kristian LG Nielsen, and Søren T. Petersen. 2011. Phase recognition during surgical procedures using embedded and body-worn sensors. In *Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on*, IEEE, 45-53.
2. 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, no.1, 66-71.
3. Ishani Chakraborty, Ahmed Elgammal, and Randall S. Burd. 2013. Video based activity recognition in trauma resuscitation. In *Automatic Face and Gesture Recognition (FG), 10th IEEE International Conference and Workshops*, IEEE, 1-8.
4. Nathanael Chambers, and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, 602-610. Association for Computational Linguistics.
5. Hilde Kuehne, Juergen Gall, and Thomas Serre. 2016. An end-to-end generative framework for video segmentation and recognition. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 1-8.
6. Dean Weber. 2002. Object interactive user interface using speech recognition and natural language processing. U.S. Patent 6,434,524, Filed October 5, 1999, issued August 13, 2002.