The Development and Implementation of Speech Understanding for Medical Handoff Training

Alyssa Tanaka, Ph.D., Brian Stensrud, Ph.D.
Soar Technology, Inc.
Orlando, FL
Alyssa.tanaka@soartech.com, brian.stensrud@soartech.com

Greg Welch Ph.D., Fransisco Guido-Sanz, R.N., Ph.D.
The University of Central Florida
Orlando, FL
Welch@ucf.edu, Frank.Guido-Sanz@ucf.edu

LCDR Lee Sciarini, Ph.D.
Naval Survival Training Institute
Pensacola, FL
lee.w.sciarini.mil@mail.mil

CDR Henry Phillips, Ph.D.
Naval Air Warfare Center Training Systems Division (NAWCTSD)
Orlando, FL
henry.phillips@navy.mil

ABSTRACT

The success of a healthcare team relies on the concerted effort of multiple members of an interdisciplinary team and the failure of the team directly translates to patient outcomes. A medical team can consist of several individuals or potentially dozens of members spanning the spectrum of care. Similarly, military medical teams exist at specific echelons of care and may not work in the same hospital, country, or continent. These complex environments require an emphasis on effective communication and a shared understanding of common goals for team success. The criticality of effective communication can be seen in patient handoffs, during which the responsibility of care and a cognitive off-load of patient information is transferred from one provider to another. During these exchanges, a lapse in information transfer can be detrimental to the patient. These critical moments in patient care are even more complex within the context of combat casualty care, where casualty information is transferred in austere environments and providers may be managing multiple casualties with multiple injuries, over long periods of time.

The successful performance of handoffs in these complex situations require reflectively complex training environments; However, there is currently no standard training or certification for combat casualty handoffs. This paper will describe the design and development of a capability aimed to train and assess medical teams on performing combat casualty handoffs. The core of the system is a speech understanding capability that automatically recognizes, captures, and assesses the verbal components of a handoff. The recognized speech populates two user interfaces in real-time: a digital patient chart to facilitate the scenarios and an instructor dashboard to support trainee assessment and performance review. This paper will describe the development and use of this system as a training tool within the medical field, including domain specific design considerations, limitations, and lessons learned.

ABOUT THE AUTHORS

Dr. Alyssa Tanaka has earned a Ph.D. and M.S. in Modeling and Simulation from the University of Central Florida, Graduate Certificates in Instructional Design and Training Simulations, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida. She also holds a diploma in robotic surgery from the Department of Surgery, University of Nancy, France. Dr. Tanaka currently works at Soar Technology as a Research Scientist. She leads several programs aimed at developing medical training technology and applications by leveraging artificial intelligence and machine learning. Prior to joining SoarTech, she led research initiatives directly related to improving the training and education of robotic surgeons through simulation as a Senior Research Scientist at Florida Hospital Nicholson Center.

Dr. Brian Stensrud is a senior research scientist at Soar Technology and currently serves on the executive management team for its Intelligent Training division. On staff since 2003, Brian has provided scientific and technical leadership for over $25M of DoD research efforts in the areas of interactive human behavior models for simulation,
serious games, adaptive training, intelligent user interfaces and robotic platforms. Brian received a Ph.D. in Computer Engineering from the University of Central Florida (2005) and holds bachelor’s degrees in Electrical Engineering and Mathematics from the University of Florida (2001).

**Dr. Gregory Welch** is a Professor and the *AdventHealth Endowed Chair in Healthcare Simulation* at the University of Central Florida College of Nursing. A computer scientist and engineer, he also has appointments in the College of Engineering and Computer Science and in the Institute for Simulation & Training. Welch earned his B.S. in Electrical Engineering Technology from Purdue University (*Highest Distinction*), and his M.S. and Ph.D. in Computer Science from the University of North Carolina at Chapel Hill (UNC). Previously, he was a research professor at UNC. He also worked at NASA’s Jet Propulsion Laboratory and at Northrop-Grumman’s Defense Systems Division. His research interests include human-computer interaction, human motion tracking, virtual and augmented reality, computer graphics and vision, and training related applications. His awards include the IEEE Virtual Reality Technical Achievement Award in 2018 (VR 2018), and the Long Lasting Impact Paper Award at the 15th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2016).

**Dr. Guido-Sanz** has over 20 years of experience in critical care advanced practice nursing within a Level I Trauma Center (at Ryder Trauma Center) and Surgical Intensive Care (Jackson Memorial Hospital, Miami, FL). His work as Advanced Practice Nurse has been recognized nationally (awarded the Advanced Nurse Practitioner of the Year Award by March of Dimes in Miami, FL) and locally as Nurse Practitioner of the Year at Jackson Memorial Hospital in Miami, FL. Dr. Guido-Sanz also teaches in the Department of Nursing Practice at UCF College of Nursing in the Adult Gerontology Acute Care Nurse Practitioner (AGACNP) Program. He is also a member of the Florida Advanced Surgical & Transport Team (FAST).

**LCDR Lee Sciarini, PhD** is US Naval Aerospace Experimental Psychologist # 141 and is currently serving as Director of the Training Technologies Directorate at the Naval Survival Training Institute. Dr. Sciarini completed his PhD in Modeling and Simulation at the University of Central Florida with a specialization in human systems interaction. Lcdr Sciarini has conducted and overseen the execution of research in a variety of domains including unmanned systems, simulation training system design and effectiveness, decision making, neurophysiological assessment of performance and physiological episodes, as well as live, virtual, and constructive training. Previously, Lcdr Sciarini served as an assistant professor in the Operations Research Department and the MOVES Institute at the Naval Post Graduate School. Other past assignments include: Project Officer at the Naval Air Warfare Center Training Systems Division, and Technical Direction Agent for the Office of Naval Research, Code 30 HPT&E, Decision Making and Expertise Development program.

**CDR Henry L. Phillips IV, PhD** holds a PhD in Industrial/Organizational Psychology from the University of Houston, and is a winged Naval Aerospace Experimental Psychologist. He has over 40 publications and presentations in the areas of training design, personnel selection, enterprise analytics, psychometrics, and mixed reality. He is a previous recipient of the Army Human Systems Integration Award, the Team Orlando Innovation Award, and the Admiral Jeremy M. Boorda Award for Outstanding Integration of Analysis and Policy.
INTRODUCTION
The success of a healthcare team relies on the concerted effort of the multiple members of an interdisciplinary team, and the failure of the team directly translates to patient outcomes. A medical team can consist of several individuals (e.g., a surgeon, an anesthesiologist, and a nurse) or potentially dozens of members that span various departments (e.g., emergency room, operating rooms, and intensive care unit staff). Similarly, military medical teams are interdisciplinary and spread across the continuum of care. These teams, however, particularly in operational environments, exist at specific echelons of care and may not work in the same hospital, country, or continent. These complex environments require an emphasis on effective communication and a shared understanding of common goals for overall team success.

An example of the criticality of team communication can be seen in medical handoffs. During these interactions, the responsibility of care for a patient is being transferred from one provider to another, as is a cognitive off-load of patient information. A lapse in information transfer can be detrimental to the well-being of the patient. In fact, 80 percent of medical errors occur during handoffs (Derlet, Richards, & Kravitz, 2001; Alavardo et al., 2006; Horwitz et al., 2006; Joint Commission, 2011). Many experts, both within and outside the medical field, have become increasingly aware of the issues related to transitions of care and have demanded that healthcare systems measure and guarantee patient safety (Accreditation Council, 2010; Kemp et al., 2008). In response, the Joint Commission made the standardization of handoff communications a key focus of the 2009 National Patient Safety Goals (Joint Commission, 2010). Despite the focus on increasing patient safety through handoff techniques, these critical processes are still often inadequate and error-prone. The Joint Commission suggests that eighty percent of severe medical errors can be attributed to miscommunication specifically during handoffs (Center for Transforming Healthcare, 2011).

These critical moments in a patient’s care are even more complex and important within the context of combat casualty care, where casualty information is being transferred between providers in austere and stressful environments and the providers may be managing multiple casualties that have multiple injuries, over long periods of time. Complex operational environments require reflectively complex training environments. There are many resources, ranging from live to virtual, available at Medical Simulation Training Centers (MSTC) for training these skills; however, there is currently no standard for training, certifying, and recertifying providers on combat casualty handoffs. Thus, there is current interest in developing capabilities that can leverage these resources and allow for standardized training and assessment of patient handoffs. To address these training needs, we designed and developed a proof-of-concept training tool, the Combat Casualty Handoff Automated Trainer (C-CHAT), that uses speech recognition and analysis software to assist with the recording and assessment of trainees in medical handoff simulations.

The prototype C-CHAT system leverages a speech understanding capability that automatically recognizes the contextual underpinnings of verbal utterances. This recognized speech is generated into text that is populates specific domains of a medical handoff form. This speech is also provided to an instructor, to review captured speech data along with an automatic assessment of an utterances correctness. This paper describes development of a speech-based
training capability for medical providers to practice and receive feedback on the performance of a patient handoff. In this paper, we will discuss domain specific design considerations, including limitations and lessons learned. Finally, we will discuss future work.

**SPEECH UNDERSTANDING DEVELOPMENT**

C-CHAT’s Automatic Speech recognition (ASR) needed to go beyond those provided by traditional speech to text recognition systems. COTS systems such as Dragon NaturallySpeaking or Google’s Text-to-Speech engine are open recognizers which, while impressive with their recognition capability, do not understand any of the content they recognize. Text without meaning is not actionable for training purposes without an underlying understanding of semantic meaning. The speech recognition capability developed in this research was needed to monitor the trainee’s speech throughout a training scenario and identify utterances that correspond to what should be spoken during a handoff, as specified by the speech protocol.

**Protocol Identification**

An initial step to developing the speech understanding was to identify the handoff protocol that would be used for recognition. There are several protocols that exist for supporting patient handoffs in hospital settings; however, it was imperative for the protocol used in C-CHAT to specifically identify patient domains relevant to a combat setting. The currently available standardized encourage teamwork communication for handoffs, but do not necessarily capture the needs of military medicine and, more specifically, combat casualty care. These complex combat circumstances require specific protocols and techniques that address the unique needs of the patients, providers, and the austere environments to allow for effective communication and information transfer.

To address this, we developed a C-CHAT specific handoff protocol that defines the specific information that must be transferred during a handoff (the “what”), how this information should be shared from one caregiver to another (the “how”), and in what order the information is shared (the “when”). The protocol describes the interactions between providers during handoff situations and was used to inform the design and development of the speech understanding described below. In order to identify the content and correct methodology, we researched and evaluated current civilian and military handoff protocols and procedures, and the capabilities of tactical combat casualty care (TCCC) for Point of Injury/Role 0 to Role 3. Concept maps were then used to organize the material and synthesize the combat casualty handoff content and methodology. Throughout this process, a subject matter expert reviewed our procedures and provided feedback and recommendations to guide development.

![Figure 1. Depiction of the protocol definition pathway](image)

Our research surrounding handoff protocols led to the identification of specific domains of a handoff protocol that must be included within the recognition. The domains, summarized as I-BIDS, include: Identification, Background, Illness Severity, Duties, Synthesis). Table 1 provides an overview of the handoff protocol domains that were identified as critical to a casualty handoff.
### Table 1. Patient handoff domains

<table>
<thead>
<tr>
<th>C-CHAT Handoff Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identification/Information</strong></td>
</tr>
<tr>
<td><strong>Background (Situation)</strong></td>
</tr>
<tr>
<td><strong>Illness Severity</strong></td>
</tr>
<tr>
<td><strong>Duties</strong></td>
</tr>
<tr>
<td><strong>Synthesis</strong></td>
</tr>
</tbody>
</table>

### Intent and grammar development

Using the handoff domains identified, we began developing the intent model that drives the recognition system. The intent model defines the words and phrases that are captured and recognized within the system and is comprised of many variations on the phrasing needed to identify the specific injuries, patient information, and caregiver duties. So, it was important that the repository represented the many utterances that may be spoken by a trainee when performing a handoff.

To do this, the team first developed a repository of phrases and words that may be spoken under each of the protocol domains. The repository consisted of two categories: specific utterances that should be spoken and various words and phrases that may be used within those utterances. For example, the utterance “I gave the patient…” may be combined with a variety of words or phrases that may represent treatments or medications (e.g., morphine).

The domains were expanded to the lowest level possible, including identifying any values associated with a certain utterances. For example, numerical statements that can be given as a range of acceptable responses such as heart beats per minute, time, blood oxygen saturation percentage level, etc. The domains that included utterances without an expected value or a numerical range (e.g., mechanism of injury, injury type, injury location) were slightly more challenging to capture because of the large variety of phrases that may be spoken.

These utterances and variations were encoded using a speech recognition, generation and parsing middleware, developed internally, that facilitates the development and integration of speech-enabled training capabilities. These tools enabled the team to develop a grammar and define the intent of a speaker, using a simple user interface instead of creating grammar and intent definition files by hand.

### USER INTERFACES (UIs)

Medical handoffs can be frequent and contain large amounts of patient information. It can be time- and resource-consumptive to perform the assessment of these techniques manually. This most likely requires that the information be recorded and analyzed after the training is complete, which does not facilitate easy in-situ evaluation for the instructor or allow for immediate feedback to the learner. To accompany the speech understanding, two UIs were developed for the real-time and post-exercise review of the performance: the trainee UI and the instructor UI.

### Instructor UI
The primary function of the instructor UI is to support data capture and review of trainee performance. The instructor UI is designed around a mobile platform (i.e. Android) and provides interface elements that allow an instructor to schedule a training session, view the session by trainee role, and view the individual form fields and utterances captured during the training session. The UI enables the instructor to conduct AAR by providing both the recorded audio of the trainees and the parsed text from the individual trainees/roles.

The trainer application allows users to schedule new exercises, modify scheduled exercises, and review completed exercises by navigating the interface using the “Hamburger” icon in the top left of the screen. Within the Scheduled Exercises interface, the instructor can schedule and provide information relevant to an exercise. The instructor can provide a name for the exercise, which is used to visually identify the exercise to users, as well as a start date and time. The instructor can also list the roles that will participate in the exercise and the order in which they will perform the handoff. This is an important feature within the system because it can be used to identify specific trainees within the exercise (Figure 2). Once all of the information has been provided and the exercises is scheduled, it will be displayed within the Scheduled exercise functionality.

When viewing the completed exercises, the interface displays the list of completed exercises, listed by name and date. By selecting a completed exercise, the instructor can review that specific exercise. The Completed Exercise interface displays properties of the exercise at the top. More importantly, this interface details the individual’s handoffs that took place during the training, listed by individual conversations. Each conversation lists what was said between two roles (Figure 3).

When selected, each conversation (i.e. handoff) instance provides a list of the individual intents that were recognized. Within this listing of recognized and unrecognized intents, the instructor has several capabilities. The instructor can review which role was speaking and which role was listening, as well as the date and time the audio was recorded. The instructor is provided with a transcription of the intent if it was recognized. If the intent was not recognized, the interface will also specify this to the instructor. In these instances, the instructor can play a recording of the audio. This functionality is particularly helpful for the instructor to determine why the intent was not recognized and provide feedback as necessary.
Within this interface, the instructor can also view the patient form that was created as a result of the conversation. The system is able to populate the form via the mapping of the individual intents to the protocol items. The form view of the handoff is beneficial for the training scenario and casualty information (Figure 4).

### Trainee UI

The primary functionality of the trainee UI is to allow the user to join the scheduled training exercise, capture their handoff speech in a form, and handoff the patient to the next trainee. In the initial interface, the trainee can view a list of scheduled exercises available to join (Figure 5). Once the trainee joins the exercise, the handoff form will be displayed.

On the handoff form interface, the trainee can view the exercise name and the provider role details of who should be speaking and who should be listening. These details are updated as the exercise progresses and handoffs occur. Below these exercise details are form fields which will populate with intents as the trainee speaks the handoff items.

After speaking all necessary casualty information, the trainee can select the “Handoff” button to indicate that the handoff is completed and transition the handoff responsibility to the next provider in the exercise. After this selection is made, a blank form will be displayed with updated values for Role Speaking and Role Listening. If this is the final handoff, the “Handoff” button will transition into a “Finish” button that will allow for completion of the exercise. After completing the exercise, all data is sent to and stored for review by the instructor.

### Figure 5. Trainee interface

**Figure 4. Handoff form review interface**

As the provider speaks the casualty information, the audio is converted to text and mapped to the intents within the intent model. As the intents are recognized, the matching casualty form items are populated accordingly. The handoff form populates in real-time with the patient information.

After speaking all necessary casualty information, the trainee can select the “Handoff” button to indicate that the handoff is completed and transition the handoff responsibility to the next provider in the exercise. After this selection is made, a blank form will be displayed with updated values for Role Speaking and Role Listening. If this is the final handoff, the “Handoff” button will transition into a “Finish” button that will allow for completion of the exercise. After completing the exercise, all data is sent to and stored for review by the instructor.

**FUTURE ENHANCEMENTS NEEDED**

The prototyping work described above helped to demonstrate the feasibility of using speech recognition and intent mapping to capture the verbal components of a combat casualty care handoff. The C-CHAT served as a valuable testbed to identify more system requirements and design enhancements that should be made. Specifically, the
environment of use was a large factor for many of these considerations. Future work will focus on developing enhancements to the system to ensure that the capability aligns with training needs of the end users and training environments that the system will be used in.

**Natural Speech Interaction**

*Expanded intent model.*

To support a more natural interaction, we want the intent model to encompass the entirety of the phrases that may be spoken in regards to the combat casualty care domain. The more phrases that we include, the more conditions that we can support. This includes more medical terminology and phrases that are generally more difficult to recognize (e.g., specific medication names). We will also expand the

Another expansion of the intent model is the capability to extract multiple intent mappings from a single phrase. So, instead of a single utterance mapping to a single intent, multiple intents can be extracted from a single utterance. This would allow the trainee to perform the handoff in a more natural manner and still be assessed appropriately. Figure 6 provides an example utterance that contains three different intents.

“Patient **blood pressure** 120 over 75 and **52 pulse** but currently **unresponsive.**”

**Figure 6. Example of multiple intent recognition within one utterance.**

Similarly, the intent model should be enhanced to allow for multiple utterances to be captured for a single intent. For example, if the patient has multiple injuries, it will be critical to capture all utterance spoken by the trainee and map them to the injury intent.

*Two way dialogue capture*

The system currently only captures spoken phrases from a single individual at a time (the “send” of patient information). While this is one part of a patient handoff, a very important component of handoffs is for the receiving provider to synthesize and restate the information. For the system to capture this in the future, we will need to implement the capability to capture two way dialogue.

This can be done in several ways. One way is to develop the capability into the intent model itself to know that the sender has completed the handoff by mapping out the dialogue that would occur in a very specific way. This would also include modeling the receiving dialogue to represent what the computer would expect for the receiver to say. The issue with this is that it creates a very rigid template for both the computer and thus the trainee to follow. A more desired, yet challenging solution, is the capability of distinguishing between multiple speech patterns automatically. This is a difficult problem to address because of the variations of speech patterns and other voice qualities that would have to be modeled.

*Noise reduction*

Another enhancement needed is the capability to reduce the environmental noise that may affect the audio capture and recognition. Voice-based input often arrives with environmental sound artifacts (i.e. background chatter, traffic noise, bioacoustics, mechanical noise) that can interfere with the recognition. C-CHAT will likely be used in straining environment that simulates operational environment conditions, meaning that there may be significant environmental noise such as, gunfire, explosions, and yelling patients that may interfere with the audio capture and subsequently intent recognition.

Thus, we will explore the implementation of denoising capabilities. There are both hardware and software approaches to this problem and that we explore. In terms of hardware, it is possible to integrate noise cancelling microphones into the C-CHAT system design. This is not the desired solution because it will increase the physical components that the trainee will have to use. This can make the system more cumbersome and potentially reduce the trainees’ desire to use the system. Also, noise cancelling microphones are more rigid in their noise reducing capabilities as they remove more general aspects of the audio and will not be as reliable as desired. Another possible solution is a software-based denoising solution to isolate the intended input voice signal. While this is a more desired solution, it is also a significantly more challenging option.
Automatic capture capabilities
The initial prototype relies on push-to-talk as the method for initiating the audio capture for the handoff. While this provides a more structured and reliable method for initiating the system, it is does not allow for as natural interaction as is desired for the training capability. Specifically, within the context of a live training exercise, we expect that the users will often use their hands to attend to the casualty and may not be able to press a button on the trainee application. Thus, another enhancement that we explore is to allow the system to capture the audio in a hands-free manner.

There are several possible implementations to address this capability. Just as many technology devices currently use specific utterances to initiate voice recognition, a similar implementation can be made for the C-CHAT system. This implementation can be made by designating a specific trigger word or phrase for the system to “listen” for (e.g. “Ready to receive?”). Another possible solution is to have the system continuously listening to the audio input. In this implementation, there would be a sliding window of a specified amount of time that the system is listening and capturing audio. During this window, the system would continuously attempt to match the recognized speech to an intent. Once a recognized intent is heard, the system would recognize that the handoff is initiated. While this method provides less interference to the experience, it is more computational work for the system and would affect other system components (e.g. battery life). Also, if the trainee says something incorrect during the handoff, this could indicate to the system that the handoff is over and cause performance captures issues.

CONCLUSIONS
The success of a healthcare team relies on the concerted effort of multiple members of an interdisciplinary team and the failure of the team directly translates to patient outcomes. Patient handoffs are perfect examples of moments in care when team coordination and communication are critical. The described research aimed to design and develop a capability to train and assess medical teams on performing handoffs.

The developed C-CHAT system allows for caregivers to practice and master combat casualty handoffs, from the point of injury through the continuum of care. The system captures trainee utterances during a patient handoff and matches the utterances to a standardized handoff protocol. User interfaces for the trainee and trainer applications of the system were also developed and implemented on Android and Microsoft devices. The trainee interface populates in real-time with these recognized utterances and mapped intents. The trainer interface also populates with this data for assessment and feedback.

During development, specific considerations were made to provide a design compatible to the unique and complex environments of use. With these considerations in mind, we were able to develop a relevant speech-driven technology prototype. Future work will continue to focus on these specific design considerations for system enhancement, as well as conduct an assessment of usability and validity.

ACKNOWLEDGEMENTS
This work is supported by the Defense Health Agency under Contract No. W81XWH-18-C-0028. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views of the U.S. Government or Department of Defense.

REFERENCES


