

# Considerations for Dealing with Real-Time Communications in an Intelligent Team Tutoring System Experiment

Anne M. Sinatra<sup>1</sup>, Stephen Gilbert<sup>2</sup>, Michael Dorneich<sup>2</sup>, Eliot Winer<sup>2</sup>, Alec Ostrander<sup>2</sup>, Kaitlyn Ouverson<sup>2</sup>, Joan Johnston<sup>1</sup> & Robert Sottolare<sup>1</sup>

<sup>1</sup> NSRDEC Simulation and Training Technology Center (STTC), USA

<sup>2</sup> Iowa State University, USA

anne.m.sinatra.civ@mail.mil, gilbert@iastate.edu,  
dorneich@iastate.edu, ewiner@iastate.edu,  
alecglen@iastate.edu, kmo@iastate.edu,  
joan.h.johnston.civ@mail.mil, robert.a.sottolare.civ@mail.mil

**Abstract.** In our paper we discuss an intelligent team tutoring system (ITTS) developed for a computer-based surveillance task experiment. In the experiment, two teammates worked together in a shared Virtual Battlespace 2 environment, and tutoring was provided by the Generalized Intelligent Framework for Tutoring. Feedback was triggered by the actions that the team members took, which were largely communications based. Since natural language processing software was not available in the tutor, we devised ways to measure and react to communications in real time. Team members both pushed keyboard buttons associated with the message they intended to send (which was processed by the computer and used to drive feedback) and were asked to verbally speak the same information to their teammates. In this paper, we discuss the reasoning behind the process used, challenges associated with using real-time communication in an ITTS, and initial analysis approaches that were done using the audio recordings of teammates' sessions. Emphasis is placed on how to reconcile real-time inputs to the computer system with audio recordings that occurred during the session and were later used for analysis. We discuss the challenges we encountered engaging with a real-time ITTS, which relies on communications between team members, and provide suggestions on addressing these challenges for future experiments.

**Keywords:** Team Tutoring, Intelligent Tutoring Systems, Generalized Intelligent Framework for Tutoring, Communication

## 1 Introduction

Intelligent tutoring systems (ITSs) are computer-based training systems that adapt to a learner based on criteria such as performance in a scenario or individual difference characteristics. It has been shown that ITSs can be as effective as a human tutor [1]. Additionally, ITSs can be used in an educational setting in many ways, such as part of lessons in a computer-lab, as review prior to exams, and as a supplement to in-class

activities [2]. ITSs have been developed in many different domains, some of which are straightforward/computational (e.g., math and physics), and other that are less defined (e.g., solving puzzles) [3]. While developing an individual ITS is a challenging and time-intensive task, adjusting an ITS for use by teams is even more difficult.

There is a great deal of research on teams and team performance [4]. However, due to the technological challenges, as well as the authoring challenges, there has not yet been much research done in the area associated with creating ITSs for teams [5, 6]. Sottolare et al. [5] identified behavioral markers that can be used in an Intelligent Team Tutoring system (ITTS) to monitor performance and determine how team members interact with each other. However, many of the behavioral markers identified were focused on communications-based information that is difficult for a computer to decipher in real-time and use for adaptation in an ITTS. One of the first steps toward being able to quantify team behavioral markers, and to adapt them for grading in a team situation, is to examine the communication that naturally occurs in a team tutoring situation and determine how it relates to the performance of the team. Through this method, the level of granularity and the impact on performance that is needed to assess team communication can be determined in specific tutoring domains. While data is being collected about the type of communications that needs to be tracked, it is also important to have the ITTS respond based on the actions of the individual in the current system. We developed an ITTS experiment that relied predominantly on team communication, and also demonstrated the capabilities of the Generalized Intelligent Framework for Tutoring (GIFT) to be used for teams of two.

### **1.1 The Experiment**

In the experiment, the research team set out to create an ITTS that required two teammates to interact with each other through both key pushes on their individual computers, and verbal communications. While the system recorded the key pushes, audio recordings were made of the verbal communications.

There were many technological challenges that were overcome to ensure the computer participants used were able to communicate with each other and engage in a simultaneous scenario. Additionally, work was done to ensure that feedback could be provided to the teams based on the actual performance on the task. However, due to the need to have the system respond dynamically and in-real time to communication, verbal communication between the team members was not taken into account during the initial scenario in the experiment. Feedback and grading of performance was based on the key presses that the participants engaged in during the activities.

### **1.2 Communication**

In the current paper, we discuss the types of communication that individuals engaged in, describe the challenges associated with dealing with communication in a team tutor, and do an initial examination of the verbal communication that occurred between team members during performance. While it would be ideal to process the verbal data in real-time during an ITTS performance, there is still utility in capturing auditory data that at the time may not be used in driving real-time assessment but can provide important

insights into the task in which the team was engaged. Even though it is difficult to deal with team-communication data in a computer-based tutoring environment in real-time, it may be advantageous to spend effort on finding ways to capture the data and semantically analyze it such that it can help to determine appropriate feedback. As these capabilities are not yet implemented in the ITS framework that was used for our study, we captured verbal data while relying on key presses to prompt feedback.

## **2 Method**

### **2.1 Participants**

Fifty participants were initially recruited from a large state university. Due to various technical issues, some participant data was incomplete or lost. After removing incomplete data, there were 32 participants. Of those participants, there were 20 males, 11 females, and 1 individual that preferred not to specify. Two participants signed up for each time slot, and upon arrival they were paired as a team. In total, there were 16 teams run in the experiment. As part of the procedure, audio was recorded for all participants. However, in some cases there were errors in the files or partial recordings. When removing these sessions, there were a total of 11 teams that had full audio recordings for all sessions.

### **2.2 Experimental Design**

This was a between-subjects design with repeated measures. Each participant only engaged in one condition, and each team engaged in four experimental sessions, with each session lasting five minutes. The three conditions were: no feedback, individual feedback, and team feedback. The no feedback condition served as a baseline, the individual feedback condition only provided feedback to the teammate that made an error, and the team feedback condition provided feedback to both teammates based on all errors that occurred.

### **2.3 Task**

Participants engaged with Virtual Battlespace 2 (VBS2) software and were asked to monitor a 180-degree sector for enemy forces (OPFOR) who were running. They were told to alert their teammate of crossings from one section of the area to the other (transfer), and to acknowledge when their teammate had passed individuals to their own sector (acknowledge). They were also asked to identify when new enemies were visible (identify).

Feedback based on performance was provided to participants through GIFT on the left side of the screen in the Individual and Team Feedback conditions. Individual feedback was specific to the errors that the individual was making and was only viewed by the individual. Team feedback was triggered by errors that an individual was making but was displayed and addressed to the entire team.

## **2.4 Materials and Apparatus**

Each team consisted of two participants who sat at desktop computers in separate rooms. There was a speaker phone next to each computer that the participants used to communicate to their teammate, as well as an audio recorder used to capture team verbal communication for later analysis. Each computer had the GIFT 2014-3X software installed on it, as well as VBS2. During the trials, participants pushed specific keys on the keyboard to indicate to the software the information that they were passing verbally to their teammates. These key pushes were recorded in GIFT's logs so they could be used to infer behaviors and trigger prompt, relevant feedback. Surveys were given to participants after the completion of the sessions using Qualtrics on a separate laptop computer.

## **2.5 Procedure**

When participants arrived, they were given an informed consent form and provided an opportunity to ask questions. After completing the form, participants sat at a computer, and watched a video that explained the task that they would be engaging in. Participants were told that they should press the keys on their keyboard that were associated with the actions they were to take, as well as verbally tell their teammate the command that they were saying. Teammates were able to communicate via the speakerphone next to them on their desk. There were four consecutive trials of five minutes each. Participants completed two surveys between each session. After each interaction session the scenario was reset, and a new 5-minute scenario began. At the end of the four sessions participants were asked to answer one final survey, and they participated in a verbal forum discussion where they talked about their assessment of their performance, the task, and the feedback they received from the ITTS.

## **3 Approaches used to Process Audio Data**

As this was the first team tutor developed with the GIFT software, the team task itself was relatively simple (two players) and effort was also spent on ensuring that the computers could communicate information to GIFT for assessment during the sessions. In the traditional individual version of GIFT, there is a single Domain Knowledge File (DKF) that determines the feedback that will be presented to a participant based on his or her actions or performance. For the team version, each individual participant had a DKF, and there was an additional one that assessed the performance of the team and provided feedback. Determining how to monitor the performance of the participants was important, and there were additional challenges such as ensuring that participants were not overwhelmed by too much feedback.

### **3.1 Capturing Communication Data in the Experiment**

As GIFT is not equipped with real-time speech analysis capabilities, the decision was made to capture communication in two ways: (1) through speech recording, and (2)

through button pressing behaviors. A button press served the purpose of alerting the system to what the participant was trying to communicate. However, after discussion the researchers decided that it was important to not only capture the button presses, but to also have access to the natural communication during the experimental session. There were many reasons behind this decision, including the belief that in a high workload situation, a participant may forget to either press the button or verbally communicate their action, and this would allow for data to be examined after the fact to understand the intention behind the actions that were taken during the session. Given the current technical state of the ITTS, it was impractical to base real-time feedback based on spoken words. Therefore, there was a reliance on input or button-pushes into the keyboard to assess performance and prompt feedback.

Whereas a human coach gives feedback based on their student's overall behavior, the ITTS could only make feedback decisions based on each isolated action. To enable this higher-level reasoning, and to reduce the amount of feedback given, a feedback controller was designed and implemented for GIFT that adjusted the performance model based on the recent history of actions in addition to the current one. This was especially relevant in the team condition when the performance of multiple individuals could impact the performance state. The user model adjustments were based on when the accurate button was pushed in relation to the state of a corresponding OPFOR.

### **3.2 Extracting Data for Analysis after the Experiment**

Additional analysis was conducted after the experiment, which required the extraction of data from the ITTS logs. Performance measures were calculated, and included individual and team transfer rates, acknowledge rate, identify rate, and identify timing. Additional information about these data analyses can be found in [7]. A visualization and coding scheme was created such that it could be established that transfer and acknowledge events were connected with the appropriate OPFOR and could be assessed by a human coder. The post-processing data measures were then calculated based on the log data provided by the ITTS.

### **3.3 Initial Analysis Process for Verbal Data**

While the system needed to rely on manual button-press data for feedback, the post-processing analysis examined the content of the verbal data. There were a number of challenges involved with the initial audio data processing for the experiment. Among these were determining how the data would be of most use.

Rather than going into the content of the data, one approach considered was a simple count of the number of utterances made by each teammate during performance. If communication increased or decreased over time and trials, it would be considered relevant information. One of the largest challenges here includes matching up the log file data to audio data. Initial approaches that were taken included creating a transcript for each of the participants in a team that required timestamps for utterances. The timestamps within the transcripts would then be lined up with the records of button pushing and visualizations created for performance analysis. See Figure 1 for an example of mock data in the transcription format.

This approach was challenging because the recordings included both participants' voices, which at times could get confusing to a transcriber if they sound similar. Further, ensuring that the timestamps of the audio lines up with the timestamps of the log file could be difficult since the audio recordings began earlier than the sessions did.

After an initial examination of the auditory data it was also found that participants often repeatedly said similar phrases multiple times, as opposed to having conversations with each other during the sessions. Due to this and that the voices sometimes sounded similar, there was a reduced utility in making simple transcripts of everything that was said during a session.

While the verbal communication is relevant, it is difficult data to work with, and it takes a large amount of time to ensure that transcripts are done in a way that will be useful for analysis purposes. For an initial analysis, as opposed to examining the content of the verbal interactions, it may be helpful to focus on the number of spoken interactions that occurred between team members.

Player 1 Zone		Time	Player 2 Zone	
Keystroke	Spoken Words		Keystroke	Spoken Words
	(OPFOR Appears)	18.00		
		25.50		(OPFOR Appears)
		33.00		
	(OPFOR Appears)			
Transfer to Partner	You have two coming in from side two.	40.50		
Transfer to Partner		47.86		
		48.00	Acknowledge Transfer	(no speech)
		49.51	(OPFOR Appears)	
		52.70		
		53.64	OPFOR Identified	(no speech)
	(OPFOR Appears)	55.05		
		55.50	OPFOR Identified	(no speech)
		60.62	Transfer to Partner	(no speech)
Acknowledge Transfer	(no speech)	63.00		
		64.37		(OPFOR Appears)
Transfer to Partner	You have one coming from side two.	65.97		
		66.97	Acknowledge Transfer	(no speech)
OPFOR Identified	(no speech)	69.46		
	(OPFOR Appears)	69.99		
		70.43	OPFOR Identified	(no speech)
		70.50	Transfer to Partner	There is two coming from side two.
Acknowledge Transfer	(no speech)	72.24		
Acknowledge Transfer	(no speech)	73.27		
		78.00	Transfer to Partner	There is one coming from side one
		78.28		(OPFOR Appears)
Transfer to Partner	You have one coming in on side one.	79.56		
		85.50	Acknowledge Transfer	(no speech)
		88.36	Transfer to Partner	(no speech)
		88.53	OPFOR Identified	(no speech)
Acknowledge Transfer	Okay	89.99		
		93.00		(OPFOR Appears)
		93.04	Transfer to Partner	(no speech)
		93.66		(OPFOR Appears)
Acknowledge Transfer	(no speech)	95.26		
OPFOR Identified	(no speech)	97.18		

**Fig. 1:** Mockup of a representative log of keystrokes and spoken words for two players. The ideal sequence is: One player transfers an OPFOR to the other player (e.g., at 93.04); the other player acknowledges the communication (e.g., at 95.26) and then Identifies the OPFOR when it comes into view (e.g., at 97.18). It is apparent that sometimes players enter a keystroke without speaking. Also, the two keystrokes at 40.50 and 47.86, which align with one spoken utterance at 40.50, illustrate the potential difficulty of aligning speech with keystrokes.

## 4 Recommendations for Future Approaches Based on our Experience

In an ideal ITTS, the system would be able to convert the spoken information into typed words and do a semantic analysis based on the content. However, in the current state of many ITTSs this is not practical. This approach would require speech recognition software to transcribe what participants are saying in real-time, and then process it for semantic analysis. While speech recognition software is widely available, it is not always reliable and typically requires an initial training process to be run before it can accurately transcribe an individual. It may be helpful in the current state to have the speech software as a starting point for a transcript, which would then have a human check the material to ensure that the transcription makes sense. In regard to lining up audio data with entered data logs, it would be helpful to have a plan in place ahead of time so that timestamps could easily be created from the audio files, or a signal provided to the transcriber that could be stated on the recording to make the process easier to determine when the session started. Additionally, creating templates that transcribers could type into and cutting audio files down such that they start at the beginning of the session would be exceedingly helpful.

## 5 Conclusions

In our recent experiment, we addressed communication between team members through button presses and recording verbal communication. While button presses were necessary so that actions could be tracked by an ITTS, verbal information was not processed in real-time. Verbal information has utility for checking the accuracy of the intention of the button pushes, as well as providing relevant information about the communication content of the team members. There are a number of challenges that should be considered when dealing with audio information, especially when it comes from a team performing a high workload task. Therefore, it is important to carefully consider the approach that would be used when recording and analyzing verbal data during an ITTS interaction.

**Acknowledgements.** The work was completed with support from a cooperative agreement with the US Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED)/NSRDEC Simulation and Training Technology Center (STTC). Statements and opinions expressed in this text do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

## References

1. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, **46**(4), 197-221 (2011).
2. Sinatra, A. M., Ososky, S., Sottolare, R., & Moss, J. Recommendations for use of adaptive tutoring systems in the classroom and in educational research. In *International Conference on Augmented Cognition* (pp. 223-236). Springer, Cham (2017).
3. Sinatra, A. M., Sims, V. K., & Sottolare, R. A. The Impact of Need for Cognition and Self-Reference on Tutoring a Deductive Reasoning Skill (No. ARL-TR-6961). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD (2014).
4. Salas, E. *Team training essentials: A research-based guide*. Routledge (2015).
5. Sottolare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*, 1-40 (2017).
6. Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., et al. (2017). Creating a team tutor using GIFT. *International Journal of Artificial Intelligence in Education*, 1-28 (2017).
7. Gilbert, S., Sinatra, A. M., MacAllister, A., Kohl, A., Winer, E., Dorneich, M., et al. Analyzing Team Training Data: Aspirations for a GIFT Data Analytics Engine. In R. Sottolare (Ed.), *Proceedings of 5th Annual GIFT Users Symposium (GIFTSym5)*: U.S. Army Research Laboratory (2017).