Intelligent Systems for Training Damage Control Assistants

Stanley Peters, Elizabeth Owen Bratt, Brady Clark, Heather Pon-Barry, Karl Schultz
Stanford University CSLI
Stanford, CA

peters@csli.stanford.edu, ebratt@csli.stanford.edu, bzack@stanford.edu, ponbarry@csli.stanford.edu, schultzk@csli.stanford.edu

ABSTRACT

The Navy is shifting its training and education from traditional methods, such as on-site instruction, texts, and observing students uring drills, to computer-supported learning such as web-based nstruction and computer simulations in lieu of live drills. This transition presents the challenge of keeping the best parts of traditional methods of instruction while obtaining the advantages that computers afford. The challenge is more difficult because to maximize savings in manpower, money and time, computer-based learning must be able to teach, evaluate and give feedback to students without any instructor in the loop.

A valuable aspect of traditional training methods, in which computers currently fall short, is the 'mentor/student' relationship: an experienced person discussing a novice's performance with him or her. The mentor gives the student direct, personalized feedback in a setting where the student can ask questions and discuss issues. Most computer simulations are lacking in this type of interaction.

We propose that giving computers the ability to debrief and discuss a student's actions using natural language will more closely simulate this relationship and greatly improve the effectiveness of computer-based learning. To assess this hypothesis, we are utilizing natural language technology to (1) allow students to use a damage control trainer for surface ships by speaking with the simulation system, and (2) to support a subsequent spoken discussion with an intelligent tutoring system that provides an after action review of the student's performance. The combined system performs a mentoring function, helping students learn correct actions and avoid 'practicing mistakes'. We are studying the usefulness of this mentoring system for students under training in damage control, and will present results about differences in rate of learning with and without mentoring. An additional benefit of natural language interaction with the computer systems is that students train as they will actually perform on duty.

ABOUT THE AUTHORS

Stanley Peters is Professor of Linguistics at Stanford University. His research is in the areas of meaning and of computation on language. His current research on spoken dialogue systems includes artificially intelligent tutoring systems. His education was in mathematics and linguistics at MIT, and he previously taught at the University of Texas at Austin.

Elizabeth Owen Bratt is a Senior Research Engineer at the Center for Study of Language and Information at Stanford University. Her research is in the areas of spoken language understanding and concept-to-speech generation in dialogue systems. She holds a Ph.D. in Linguistics from Stanford University, and she was previously a Research Linguist at SRI International.

Brady Clark is a Ph.D. student in Linguistics at Stanford University. His research is on approaches to intelligent tutoring, and his dissertation research applies stochastic optimality theory to syntactic change.

Heather Pon-Barry is an A.M. student in Symbolic Systems at Stanford University. Her research is on spoken language tutoring.

Karl Schultz is a Research Engineer at the Center for Study of Language and Information at Stanford University. He holds a B.S. in Computer Science from the University of Illinois, where he worked on the development of the DC-Train simulator.

Intelligent Systems for Training Damage Control Assistants

Stanley Peters, Elizabeth Owen Bratt, Brady Clark, Heather Pon-Barry, Karl Schultz
Stanford University CSLI
Stanford. CA

peters@csli.stanford.edu, ebratt@csli.stanford.edu, bzack@stanford.edu, ponbarry@csli.stanford.edu, schultzk@csli.stanford.edu

INTRODUCTION

We are investigating the hypothesis that giving computers the ability to debrief and discuss a student's actions using natural language will allow an intelligent training system to approximate the mentoring relationship of traditional training methods, in addition to the benefits of computer-supported learning. We are proceeding by developing a training system for Navy damage control, integrating a simulator and a tutor reviewing the student's performance. Our system logs metrics and permits parameterized study of its behavior, to isolate the best characteristics of intelligent training systems.

In this paper, we first discuss the DC-Train simulator, and how the spoken interface works. Next, we discuss the SCoT-DC tutor, which conducts the after-action review and performs related tutoring. Finally, we examine experimental results showing that (1) SCoT-DC improves performance in damage control over using the DC-Train simulator alone and (2) spoken language technology is mature enough to support spoken simulator control and spoken automated tutoring. These results support the utility of the system for mentoring student damage control assistants (DCAs).

VOICE-ENABLED DC-TRAIN SIMULATOR

DC-Train (Bulitko and Wilkins, 1999), the damage control trainer used in our work, was developed as a richer, more flexible and intelligent successor to IDCCT (Integrated Damage Control Training Technology). DC-Train employs artificial intelligence and computer simulation in addition to multimedia and graphical visualization technology to provide student DCAs with an intensive, realistic experience of coordinating ship damage control in a large number and wide range of damage scenarios involving fire, smoke, and flooding. DC-Train intelligently simulates other damage control personnel as well as ship systems and the spread of damage. It supports detailed assessment of student actions as correct, error of commission, error of omission, etc. by comparison to an intelligent DCA agent's problem solution.

DC Train successfully approximates the stressful environment of damage control by bombarding the student with multiple information reports in both audio and video. Trainees using DC-Train reported high levels of effort, anxiety, time pressure and mental demand (Baumann, Sniezek, Donovan, & Wilkins, 1996), which should help prepare DCAs to use their knowledge and skills in an actual crisis.

Each window on the DC-Train screen (see Figure 1) is modeled on a source of information available to a real-life DCA on a ship, including as a detailed drawing of the several hundred compartments on the ship, a record of all communications to and from the DCA, a hazard detection panel showing the locations of alarms which have occurred, and a panel showing the firemain, i.e. the pipes carrying water throughout the ship, and the valves and pumps controlling the flow of the water. The window depicting heads represents the other personnel in the same room as the DCA, who are available to receive and transmit messages.

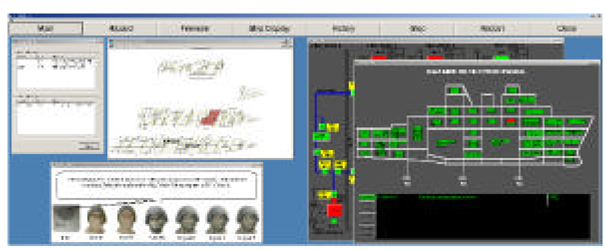

Figure 1. DC-Train screen

In the original version of DC-Train, the DCA's orders and communications to other personnel on the ship took place through a menu system. To enable the Navy student to train in the same manner as they would perform these duties through radio communications on a ship, we have now added a spoken interface, using actual Navy commands.

The user clicks a button to begin speaking, and the speech is recognized by the Nuance commercially available speech recognizer, using a grammar-based language model automatically derived from the Gemini grammar used for parsing and interpretation of the commands (Dowding et al. 1993). A dialogue manager then maps the Gemini logical forms into DC-Train commands. To allow the student to monitor the success of the speech recognizer, the text of the utterance is displayed. Responses from the simulated personnel are spoken by Festival speech synthesis (Black and Taylor, 1997)., and also displayed as text on the screen.

Most spoken interactions with DC-Train involve the student DCA giving single commands without any use of dialogue structure; however, the system will query the student for missing required parameters of commands, such as the repair team who is to perform the action, or the number of the pump to start on the firemain. If the student does not respond to these queries, the system will provide the context of the command missing the parameter as part of a more informative request. The student retains the ability to issue other commands at this time, and need not respond to the system if there is a more pressing crisis elsewhere.

Running a full DC-Train scenario takes 20-40 minutes, and has the flavor of the excerpt in Figure 2.

[buzzing alarm goes off, it is a fire alarm]

DCCO: Fire in compartment 2-78-01-L.
Student: Net80 to repair locker 2, investigate compartment 2-78-01-L.
Repair Locker 2: Reports, fire in compartment 2-78-01-L.
Repair Locker 2: Reports, smoke in compartment 2-78-01-L.
Student: Net80 to repair locker 2, fight the fire in compartment 2-78-01-L.
Student: Net80 to repair locker 2, set fire and smoke boundaries on primary forward 78, primary aft 126, secondary forward 42, secondary aft 174, above 1, below 2.

Figure 2. Excerpt of DC-Train session

At the end of a DC-Train session, the student can then receive customized feedback and tutoring from SCoT-DC, based on a record of the student's actions compared to what an expert DCA would have done at each point, based on rules accounting for the state of the simulation. The goal of the tutorial interaction is to identify and remediate any gaps in the student's understanding of damage control doctrine, and to improve the student's performance in issuing the correct commands without hesitation.

SCOT-DC TUTOR

The SCoT-DC (**S**poken **Co**nverational **T**utor for **D**amage **C**ontrol) tutor, in Socratic style, asks questions rather than giving explanations. The tutor has a repertoire of hinting tactics to deploy in response to student answers to questions, and identifies and discusses repeated mistakes. The student is able to ask "why" questions after certain tutor explanations, and to alter the tutorial plan by requesting that the tutor skip discussion of certain topics.

SCoT-DC uses two instances of the Ship Display from DC-Train, seen in Figure 3, one to give an overall view of the ship and one to zoom in on affected compartments, with color indicating the type of crisis in a compartment and the state of damage control there. The student can click on a compartment in the Ship Display as a way of indicating that compartment to the system. The automated tutor and the student communicate through speech, while the lower window displays the text of both sides of the interaction, and permits the user to scroll back through the entire tutorial session.

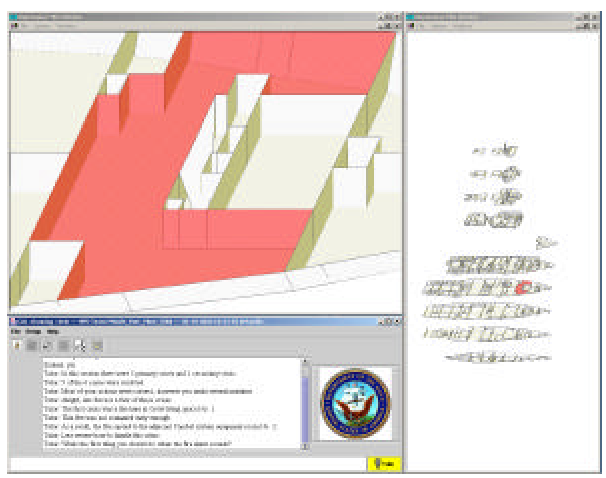

Figure 3. SCoT-DC Screen Shot

Our tutor is developed within the Architecture for Conversational Intelligence (Lemon et al. 2002). We use the Open Agent Architecture (Martin et al. 1999) for communication between agents based on the Nuance speech recognizer, the Gemini natural language system and Festival speech synthesis We used the FestVox tools (Black and Lenzo, 1999) to develop a limited domain voice for the SCoT-DC tutor, to give it a more natural sound. Our tutor adds its own dialogue manager agent, for general principles of conversational intelligence, and a tutor agent, which uses tutoring strategies and tactics to plan out an appropriate review and react to the student's answers to questions and desired topics.

Both DC-Train and SCoT-DC use the same overall Gemini grammar, with distinct top-level grammars producing appropriate subsets for each application. Our Gemini grammar currently has 170 grammar rules and 813 distinct words. In a Nuance language model compiled from the Gemini grammar (Moore 1998), different top-level grammars are used in SCoT-DC to enhance speech recognition based on expected answers. For example, if the system has just asked what action the student should perform next, the recognizer will expect sentence fragments expressing actions, such as *investigate the fire*, but not fragments expressing entities, such as *repair three*.

Interpretation of the Gemini interpreted forms is handled by a more complex dialogue manager in SCoT-DC than in DC-Train, with a structured representation of the dialogue, which is used to guide the system's use of discourse markers, among other things. The dialogue is mainly driven by the tutor agent's strategies, though the student can request to move on to future topics without completing the current discussion, and also ask a "Why" question after some explanations.

Tutorial strategies generally guide the overall path of the conversation, such as choosing which crises to discuss based on the errors made by the student. Tutorial tactics apply at a lower-level throughout the dialogue, for example, when a student gives an incorrect answer, the tutor will give a general hint and repose the question. If the student answers incorrectly a second time the tutor will give a more specific hint and ask the question again. If the student fails a third time the tutor will give the correct answer, and proceed.

SCoT-DC currently covers three different knowledge topic areas within Navy damage control, which we call sequencing, boundaries, and jurisdiction. We illustrate the types of tutoring available in each area in the transcripts below.

Sequencing refers to giving orders for actions in response to crises (e.g. fires, floods) at the correct times, as seen in a constructed example in Figure 4.

Setting boundaries refers to the task of correctly specifying six parameters that determine the location of the bulkheads which need to be cooled or sealed to prevent a crisis from spreading. Figure 5 gives an example from an actual session.

Tutor: The first crisis was a fire here in Crew Living space No. 1. [centers on this compartment]
Tutor: This fire was not contained early enough.
Tutor: As a result, the fire spread to the adjacent Combat system equipment room No. 2. [highlights 2nd compartment]
Tutor: OK, let's review how to handle this crisis. [removes highlight from 2nd compartment]
Tutor: What's the first thing you should do when the fire alarm sounds?
Student: Send repair two to investigate.
Tutor: Yes, that's right.
Tutor: However, you failed to do this during the session.
Tutor: This is a mistake you made more than once.
Tutor: Remember, alarms need to be verified first.
Tutor: OK, suppose that there is a fire.
Tutor: What's the next step?
Student: Set fire and smoke boundaries.
Tutor: Yes. That's what you did during the session.
Tutor: OK, what else needs to be done?
Student: The repair team should fight the fire.
Tutor: No, you missed a step.
Tutor: Remember, the crisis might interfere with other ship systems.
Tutor: Let's try again.
Tutor: What else needs to be done?
Student: I don't know.
Tutor: Well.
Tutor: You should order repair five to mechanically and electrically isolate the compartment.

Figure 4. Sequencing Tutoring

Jurisdiction refers to the task of giving orders to the appropriate agent on the ship—agents are assigned to different regions such as forward, aft, and midship. Figure 6 gives an example from an actual session.

Tutor. Since you had some difficulty ordering boundaries around compartments, lets work through some drills.
Tutor. I will pick a couple compartments, and you answer the questions about boundaries for each one.
Tutor. Lets start some drills with primary boundaries, and then move on to the secondary boundaries.
Tutor. Lets assume there is a fire here, in 3-220-2-Q.
Tutor. What are the primary bulkheads you should use for setting boundaries on compartment 3-220-2-Q?
Student. primary forward two two zero primary aft two five four
Tutor. Great job. Those are both right.
Tutor. Lets assume there is a fire here, in 2-256-2-F.
Tutor. What are the primary bulkheads you should use for setting boundaries on compartment 2-256-2-F?
Student primary forward two five four
Tutor. Good Job. You got the primary forward bulkhead.
Tutor. The primary aft bulkhead is the first bulkhead aft of the crisis.

Figure 5. Boundaries Tutoring

A video clip of an older version of the ScoT-DC system is available at http://www-csli.stanford.edu/semlab/muri/November2002Demo.html

EVALUATION

Because voice-enabled DC-Train and SCoT-DC are designed to enable research into training, tutoring, and dialogue, evaluation of the system with users is important. We conducted a preliminary evaluation of voice-enabled DC-Train with Navy users, and a more rigorous evaluation of both DC-Train and SCoT-DC with a civilian population.

Preliminary Evaluation

In July 2003, students in the Repair Locker Head class at the Navy Fleet Training Center in San Diego ran 12 sessions with voice-enabled DC-Train. Students were interacting with the system to get a feel for it, rather than attempting to manage actual casualties in a amage control scenario. Sessions ranged from 1 to 65 user utterances, with an average of 21. The average utterance length was 7 words. In speech recognition, about 22% of utterances were rejected, and the sentences with a recognition hypothesis had 73% words correct. Detailed error analysis from these Navy sessions helped us improve the grammar for our experiments.

Tutor. Let us practice identifying repair team jurisdictions.
Tutor. Here is where the three different repair teams are located.
Tutor. Here is the location of repair team two.
Tutor. Repair team three is located in this compartment.
Tutor. And finally this is where repair five is located.
Tutor. Here is where one of the repair teams is based.
Tutor. Which repair team is this?
Student repair team three
Tutor No, Repair Team three is not right.
Tutor Repair Team three is located here.
Tutor. We need the repair team located here.
Tutor. Please try again.
Student repair team two
Tutor. No, Repair Team two is also incorrect.
Tutor Repair Team two is located here.
Tutor. In this case, the correct repair team is Repair Team five.
Tutor. Now let us move onto the areas covered by each of these repair teams.
Tutor. This is the jurisdiction of which repair team?
Student repair team five
Tutor. Yes, Repair Team five is correct.
Tutor. Now we are ready to move on to relating repair teams to individual compartments that need damage control.
Tutor. Which repair team has jurisdiction over this compartment?
Student repair team five
Tutor. Yes, that is the right answer.
Tutor. Which repair team has jurisdiction over this compartment?

Figure 6. Jurisdiction Tutoring

Experimental Evaluation

Thirty native English speakers were recruited to participate in an experiment (16 male, 14 female). All subjects were novices in the domain of damage control, twenty-nine had no prior experience in dialogue system studies.

In order to control for any individual differences, such as in spoken language fluency, aptitude for damage control, etc., we designed an experiment in which each subject's performance in one area under one condition could be measured against the same subject's performance in other areas under another condition. Specifically, we modified our tutor to conduct tutoring in only one of its knowledge topic areas per session, while the scenarios practiced on the DC-Train simulator involved actions requiring all three kinds of knowledge. We could thus measure for each area how much practice on the simulator alone improved actions vs. practice plus the after-action review and tutoring session. In every tutoring session, subjects were informed of their performance on each area.

Subjects were randomly assigned to three groups. All groups ran through the same four DC-Train sessions (which increased in difficulty). Between each DC-Train session, all groups received tutoring in one of the three knowledge areas (sequencing, boundaries, and jurisdiction), but at different times. Table 1 shows the layout for each group. Presenting the three knowledge topic areas in different orders to the three different groups allowed us to control for any effects based on the particular area.

Table 1. Experiment Design for Tutoring Topic

Subject Group	**Tutoring Topic for Session 1**	**Tutoring Topic for Session 2**	**Tutoring Topic for Session 3**
1	Sequencing	Boundaries	Jurisdiction
2	Boundaries	Jurisdiction	Sequencing
3	Jurisdiction	Sequencing	Boundaries

Our knowledge base allowed us to specify for every DC-Train session exactly which actions an expert DCA would have taken in the same situation. Thus, we could calculate quantitative performance scores for the three knowledge areas from each subject's DC-Train session, such as what proportion of the student's actions were correct, and what proportion of the expert actions were actually taken by the student. This allowed us to separate learning gains due to the tutorial interaction from learning gains due to practice alone.

The experimental procedure is illustrated below in Table 2. Steps 4 through 10, the main body of the experiment, correspond to the steps listed in Table 1. In addition to these main steps, all subjects went through an interactive multimedia introduction to (1) familiarize them with DC-Train and basic damage control knowledge, and (2) give them practice using the speech recognition interface. After the multimedia introduction, subjects took a 20 question multiple-choice pre-test, and had one practice DC-Train session. Following the main body of the experiment, subjects took a 20 question post-test (drawn from the same pool of questions as the pre-test) and filled out a questionnaire. The total duration of the experiment was roughly three hours per subject.

Table 2. Experiment Procedure

Step 1	Multimedia Introduction	30-40 min
Step 2	Pre-test	5-10 min
Step 3	Practice DC-Train session	10 min
Step 4	**DC-Train session 1**	15 min
Step 5	**Tutoring Topic 1**	< 15 min
Step 6	**DC-Train session 2**	15 min
Step 7	**Tutoring Topic 2**	< 15 min
Step 8	**DC-Train session 3**	15 min
Step 9	**Tutoring Topic 3**	< 15 min
Step 10	**DC-Train session 4**	15 min
Step 11	Post-test	5-10 min
Step 12	Questionnaire	< 5 min

Tutoring Performance

The first measures are to see if the students did improve overall by using the system.

We measured improvement both in terms of knowledge and performance on DC-Train scenarios. Our system did increase the subjects' knowledge of damage control. All subjects earned higher scores on the post-test than on the pre-test (mean pre-test score = 12.8/20 (64%), standard deviation 3.9; mean post-test score = 16.7 (83%), standard deviation 3.2). See Figure 7, showing a boxplot diagram of the improvement on the 20-question test.

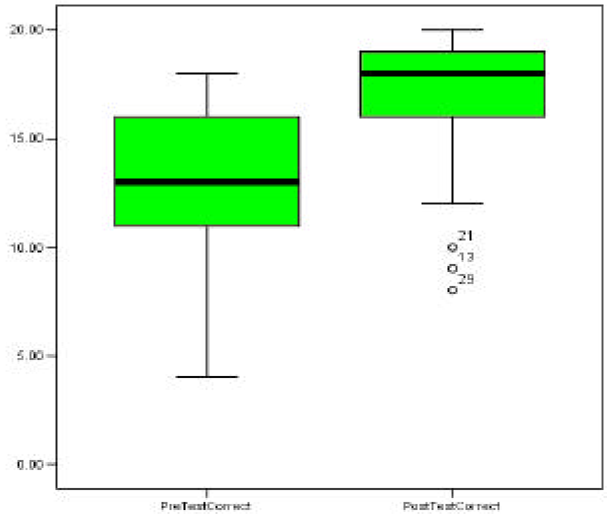

Figure 7. Boxplot of Pre and Post Test Scores

In terms of DC-Train performance, for two of the three knowledge areas, students were performing better in their fourth session with DC-Train than in their first. These performance gains are shown in Figure 8.

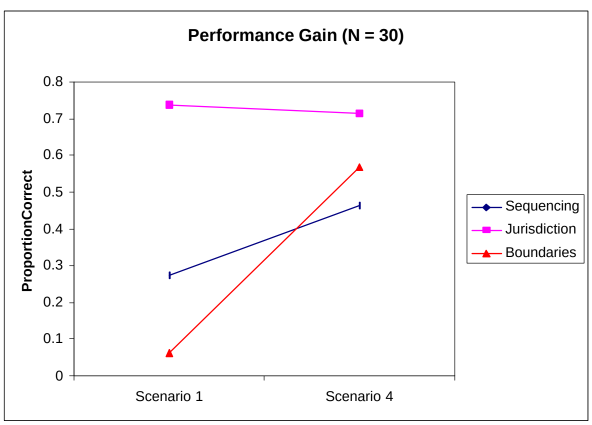

Figure 8. DC-Train Performance Gain by Area

The lack of improvement in the area of jurisdiction may be due to a ceiling effect, because initial performance in jurisdiction was already very high.

The next question is whether the general performance improvement was due to tutoring, or merely to practice on the simulator. For this, we looked at the increase in proportion of actions a subject performed correctly in each area, and separated these score gains by whether the subject had just received tutoring in that area before the session. Here we saw a substantial difference: being tutored in an area improves performance in the following session much more than the controls who were not tutored in that area. Figure 9 shows the combined effect of all of the areas across all of the subjects.

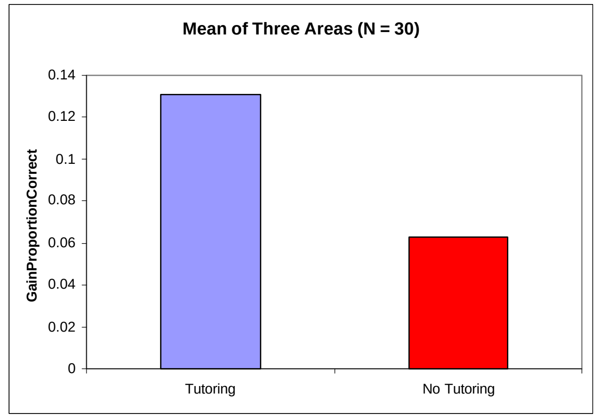

Figure 9. Mean Performance Gains by Tutoring or Not

To look at this result in more detail, Figure 10 presents the mean gains in the tutored vs. not-tutored groups by each topic area, along with standard deviations.

Topic	**Mean (not tutored)**	**Std. Dev (not tutored)**	**Mean - Tut**	**Std Dev Tutord**
Sequencing	.06	.10	.05	.27
Jurisdiction	-.01	.21	.03	.40
Boundaries	.13	.21	.30	.42

Figure 10. Mean Performance Gains by Topic Area

We have already seen that jurisdiction performance did not improve much with use of the system, perhaps because it started at such a high level, so the small difference seen in Figure 10 between the means of -.01 and .03 in jurisdiction performance, when the standard deviation is around .21, may not be as important to consider as the areas in which students did improve, for studying the effects of tutoring on simulation performance vs. simulator practice alone.

In the boundaries area, the mean performance gain of the tutored subjects (.30) is close to a standard deviation (.21) above the performance of the untutored subjects (.13). This statistic gives the clearest evidence of the benefit of tutoring combined with simulator practice.

The fact that subjects in the sequencing area had smaller performance gains after tutoring (.05) than those subjects without tutoring (.06) at first seems to argue against the worth of the tutoring. However, on examining the performance of each subject group separately, an interesting pattern emerges.

Subjects who received sequencing tutoring first performed much better in sequencing in the following DC-Train session than they did in their subsequent sessions, which were not immediately preceded by tutoring, as seen in Figure 11.

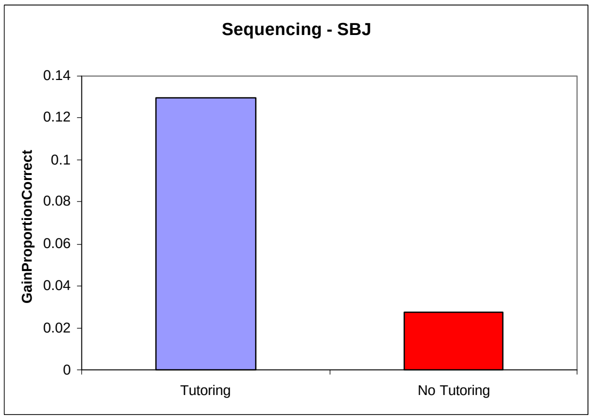

Figure 11. Sequencing in the SBJ Subject Group

However, the other two subject groups, who received sequencing tutoring later, did not appear to benefit from the tutoring, and performed worse in sequencing immediately after being tutored than they did in sessions where their tutoring was in some other area. Figure 12 illustrates this fact for the JSB group, and Figure 13 illustrates this fact for the BJS group.

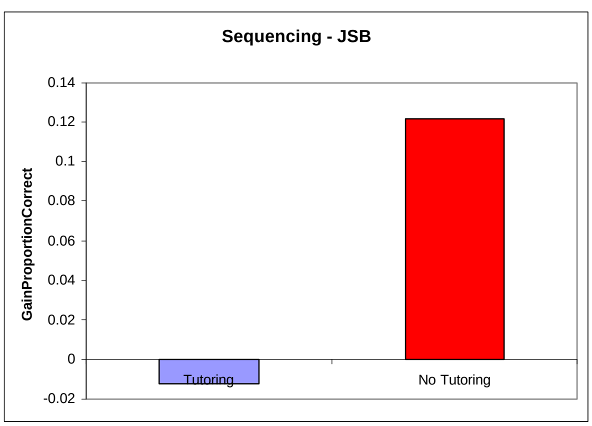

Figure 12. Sequencing in the JSB Subject Group

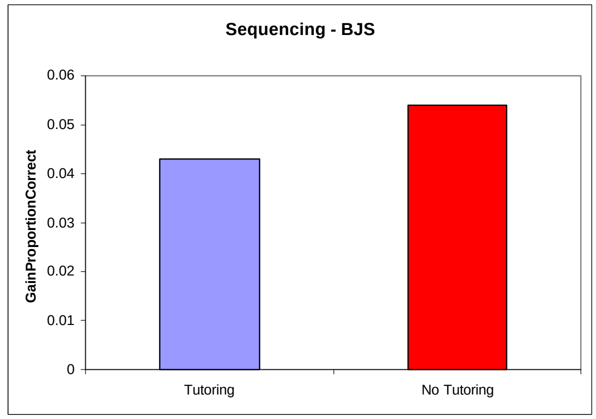

Figure 13. Sequencing in the BJS Group

An intriguing possibility is that the subject matter of sequencing, which refers to which actions should be taken to address a fire, smoke or flood crisis, is so fundamental to performance on DC-Train that it is critical to be tutored on it early on. Allowing students to “practice their mistakes” on multiple DC-Train sessions before reviewing the correct actions with them may lead them into habits that are harder to unlearn. We hope that future experiments may clarify whether this explanation is in fact a visible phenomenon in using simulation plus tutoring systems.

Speech Interface Performance

For the speech interface, we were interested in the question of whether the speech recognition would be accurate enough to allow the students to complete their task, and also whether individual differences in speech recognition performance would be so great that they would outweigh tutoring factors in determining a student’s performance.

We collected the following statistics on speech recognition performance:

- percentage of words correctly recognized
- percentage of sentences recognized with no word errors
- percentage of sentences rejected by the speech recognizer

For tutoring sessions, twenty speakers have been analyzed so far, though for several statistics only sixteen cases ended up meeting all the conditions necessary for the analysis. For voice-enabled DC-Train sessions, eleven speakers have been analyzed so far.

Overall performance rates were good enough for the subjects to complete all the sessions, with the percentage of words correct falling around 80% for SCoT tutoring (Figure 14) and around 90% for voice-enabled DC-Train (Figure 15). (DC-Train speech recognition performance appears to have improved since the 2003 Fleet Training Center sessions (73% words correct), perhaps due to the resulting grammar refinement.) Figures 14 and 15 present the percentage of words correct for each of the tutoring and simulator sessions, respectively. Overall, performance stayed relatively steady across the sessions, and did not show any effect of subjects having better speech recognition as they became accustomed to the speech interface.

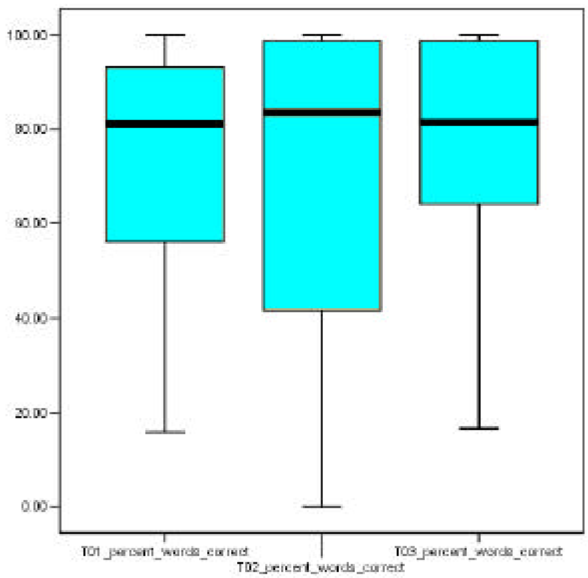

Figure 14. Boxplot of Words Correct for SCoT

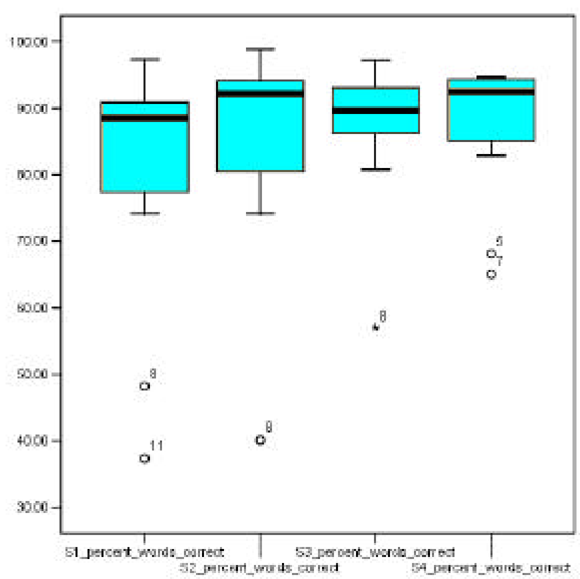

Figure 15. Boxplot of Words Correct for VE-DCT

Correlating the speech recognition performance of these subjects with their knowledge gains as measured by the post-test shows that speech technology is not having a significant effect on subject performance. No speech recognition performance metric correlated with any knowledge or performance gain metrics. Table 3 gives several sample correlations.

Table 3. Correlating Speech and Performance Metrics

Speech Metric	Performance Metric	Correlation	Significance
Rejection rate	Test score gain	.210	.435
% Words Correct	Test score gain	-.038	.890
% Words Correct	Boundary Performance Gain	-.303	.254
% Words Correct	Jurisdiction Performance Gain	.216	.440
% Words Correct	Sequencing Performance Gain	.016	.953

In addition, students who had only 60% of their words recognized correctly showed learning gains comparable to students who had 95% of their words recognized correctly. Even though speech recognition is far from perfect, students interacting with SCoT-DC improved regardless of the number of misrecognitions they encountered.

RELATED WORK

Our tutor benefits from other work on intelligent tutoring with natural language interfaces (Evens et al., 2001; Graesser et al, 2000). Spoken input to tutoring systems is rarer (Aist and Mostow, 1997; Litman, 2002).

CONCLUSION

Voice-enabled DC-Train and SCoT-DC have succeeded in training novices in Navy damage control, and tutoring in an area furthers learning more than practice alone. There are potential timing effects of tutoring to avoid the entrenchment of bad habits; this area needs more study. The speech interface supports the tutoring, and speech performance does not significantly affect learning results.

Other training systems (e.g. McDowell and Darken, 2004) could also pair with a spoken tutor such as SCoT.

ACKNOWLEDGEMENTS

This work is supported by the Department of the Navy under research grant N000140010660, a multidisciplinary university research initiative on natural language interaction with intelligent tutoring systems. We also wish to thank Perry McDowell of Naval Postgraduate School for consultation on Navy doctrine and language, as well as system design.

REFERENCES

Aist, G., and Mostow, J. (1997). A time to be silent and a time to speak: Time-sensitive communicative actions in a reading tutor that listens. *AAAI Fall Symposium on Communicative Actions in Humans and Machines*, Boston, MA.

Baumann, M. R., Sniezek, J. A., Donovan, M. A. and Wilkins, D. C. (1996). Training Effectiveness of an Immersive Multimedia Trainer for Acute Stress Domains: Ship Damage Control. *University of Illinois Technical Report UIUC-BI-KBS-96008*, Urbana, IL: Beckman Institute.

Black, A. and Lenzo, K. (1999). Building Voices in the Festival Speech Synthesis System (DRAFT) Retrieved June 25, 2004, from http://www.cstr.ed.ac.uk/projects/festival/papers.html

Black, A. and Taylor, P. (1997). Festival speech synthesis system: system documentation (1.1.1). Technical Report Technical Report HCRC/TR-83, University of Edinburgh Human Communication Research Centre.

Bulitko, V. V. and Wilkins, D.C. (1999). Automated instructor assistant for ship damage control. In *Proceedings of AAAI-99*.

Dowding, J., Gawron, M, Appelt, D., Cherny, L., Moore, R. and Moran, D. (1993). Gemini: A natural language system for spoken language understanding. In *Proceedings of ACL 31*.

Evens, M., Brandle, S., Chang, R., Freedman, R., *et al*. (2001). CIRCSIM-Tutor: An Intelligent Tutoring System Using Natural Language Dialogue. In *Proceedings of the Twelfth Midwest AI and Cognitive Science Conference, MAICS 2001*.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., and the Tutoring Research Group. (2000). AutoTutor: a simulation of a human tutor. *Journal of Cognitive Systems Research*, *1*, 35-51.

Heffernan, N., & Koedinger, K. (2002). An Intelligent Tutoring System Incorporating a Model of an Experienced Human Tutor. In *Proceedings of the 6th International Conference, ITS 2002*.

Lemon, O., Gruenstein, A., and Peters, S. (2002). Collaborative Activities and Multi-tasking in Dialogue Systems , *Traitement Automatique des Langues* (TAL), 43(2):131-154, special issue on dialogue.

Litman, D. (2002). Adding Spoken Dialogue to a Text-Based Tutorial Dialogue System. *ITS 2002 Workshop on Empirical Methods for Tutorial Dialogue Systems*, San Sebastian, Spain.

Martin, D., Cheyer, A. and Moran, D. (1999). ``The open agent architecture: A framework for building distributed software systems," *Applied Artificial Intelligence*, v.13:91-128.

McDowell, P and Darken, R. (2004) . Using a Open Source Game Engine to Build a Compelling Training Simulation. Submitted to I/ITSEC 2004.

Moore, R. (1998). Using Natural Language Knowledge Sources in Speech Recognition." *Proceedings of the NATO Advanced Study Institute.*