Tactical Language Training System: Supporting the Rapid Acquisition of Foreign Language and Cultural Skills

W. Lewis Johnson, Stacy Marsella, Nicolaus Mote, Hannes Viljhálmsson
USC / Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
{johnson, marsella, mote, hannes}@isi.edu

Shrikanth Narayanan, Sunhee Choi
University of Southern California
University Park
Los Angeles, CA 90089
shri@sipi.usc.edu; sunheech@usc.edu

Abstract

The Tactical Language Training System helps learners acquire communicative competence in spoken Arabic and other languages. An intelligent agent coaches learners, assessing their mastery and providing tailored assistance. Learners then perform missions in an interactive story environment, where they communicate with autonomous, animated characters. Our fundamental hypotheses are that such a learning environment will be more engaging and motivating than alternative approaches, and will lead to more rapid skill acquisition.

1 Introduction

The Tactical Language Training System (TLTS) helps people acquire basic foreign language and cultural skills. Learners perform interactive exercises, speaking into the computer and getting feedback on their choice of responses and pronunciation. They then practice in an interactive game in which they must communicate with non-player characters. By combining intelligent feedback and gaming techniques, we hope to make difficult languages accessible to wide range of learners. The project is currently developing training for Levantine and Iraqi Arabic, and trainers for Farsi and other languages are planned. The work is being performed at the Center for Advanced Research in Technology for Education (CARTE) at the University of Southern California's Information Sciences Institute, in cooperation with other laboratories at USC and the U.S. Military Academy.

2 Background and Motivation

The project aims to achieve the following objectives:

- Provide training in less commonly taught languages, such as Arabic. Many such languages have a reputation for being difficult to learn. To make then more accessible to learners, the TLTS uses a task-oriented approach (Doughty and Long, 2003), focusing on the skills needed to accomplish particular communicative tasks. Written language is deemphasized so that training can focus on spoken communication skills.
- Provide training in culture. Trainees acquire cultural sensitivity by interacting with simulated native speakers in a range of social situations. They learn the role of nonverbal communication by selecting gestures for their game character and interpreting the gestures of non-player characters.
- Combine form feedback and authentic practice. The game employs pedagogical drama technology (Marsella et al., 2003) in order to ensure the learner's interactions with the game characters are dramatically engaging and pedagogically appropriate. The interactive exercises are combined with a pedagogical agent (Johnson et al., 2000). This combination makes it possible to evaluate the contributes of intelligent tutoring and game technologies in promoting learning and motivation.

The TLTS builds on ideas developed in previous systems involving microworlds (e.g., FLUENT, MILT) (Hamberger, 1995; Holland et al., 1999), conversation games (e.g., Herr Kommissar) (DeSmedt, 1995), speech pronunciation analysis (Witt & Young, 1998), and simulated encounters with virtual characters (e.g., Subarashii, Virtual Conversations, MRE) (Bernstein et al., 1999; Harless et al., 1999; Swartout et al., 2001). It extends this work by providing rich form feedback as well as simulated dialog, in an implementation that is robust and efficient enough for ongoing testing and use on commodity computers.

3 Examples

The following are examples of the TLTS in use, in the context of training for a civil affairs mission in Lebanon, where the trainee must establish rapport with the local people, meet with the local village leader, and arrange post-war reconstruction.

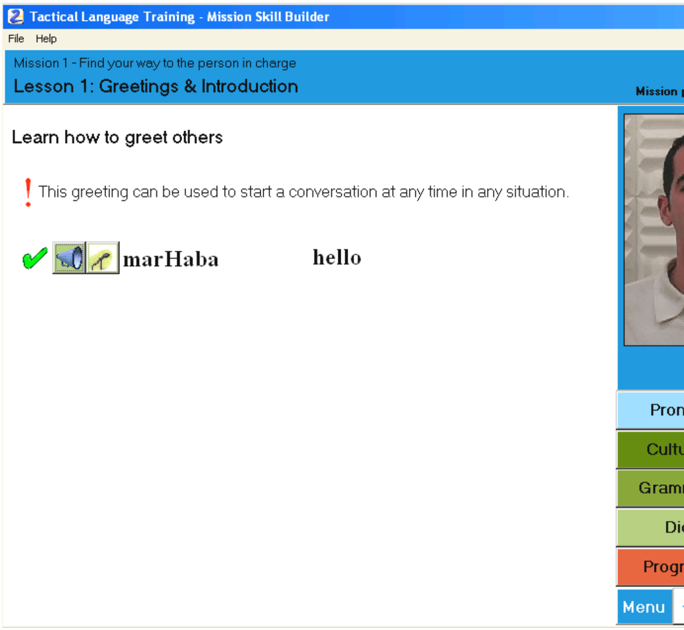

Figure 1: Practicing Arabic speech

In Figure 1 the learner is learning to say “hello” /marHaba/ in Levantine Arabic. The learner hears the pedagogical agent pronounce the phrase, and then attempts to repeat it. The TLTS detects pronunciation errors and provides motivational, corrective feedback. Possible responses include “Remember capital h is ‘H’, and is quite different from lowercase h. ‘H’, ‘h’.”

Figure 2. Greeting a Lebanese man in a café

Figure 2 shows a scene from the video game portion of the TLTS, in which the learner has entered a Lebanese café and is introducing himself. The learner communicates on behalf of his character in the game (shown at left) by speaking into a microphone and selecting from a palette of gestures. In this case the character places his right hand over his heart and bows slightly, a gesture of respect common to the Arab world.

Standing in the background is the pedagogical agent, here in the role of aide, who can assist the learner by translating phrases or offering suggestions of what to say. The agent maintains a learner model that tracks which communicative skills that the learner has mastered. If the learner knows the necessary Arabic vocabulary the aide will offer hints in English, such as “Why don’t you introduce yourself”, otherwise he will prompt the learner with a specific Arabic phrase to say.

If the learner succeeds in establishing rapport with the people in the café, they will give him directions to find the local headman, enabling the learner to proceed to the next scene in the game.

4 Architecture

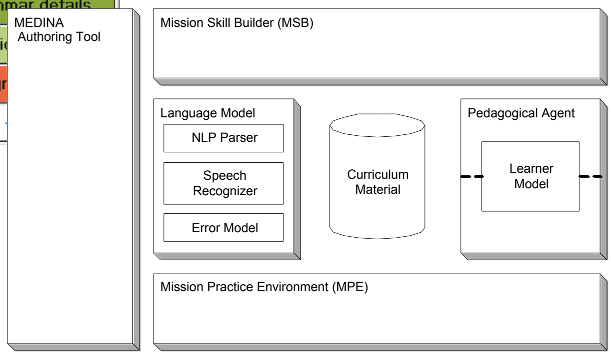

Figure 3: System architecture

Figure 3 shows the overall architecture of the TLTS. The system has three main components that interact with the learner. The Mission Skill Builder (MSB) provides background materials and lessons, such as the lesson shown in Figure 1. The Mission Practice Environment (MPE) is the video game portion, as shown in Figure 2. Micro Analysis & Design is developing an authoring tool, called Medina, which teams of content developers will be able to use to create the curriculum and game content. The MSB, MPE, and Medina rely on a common set of services and content databases, including a curriculum database, a pedagogical agent, a learner model, and a language model.

4.1 Mission Practice Environment

The MPE is responsible for realizing dramatically and visually engaging 3D simulations of social situations. Most of the MPE work is done in two modules: The Mission Engine and the Unreal World. The former controls what happens while the latter renders it on the screen and provides a user interface. The Unreal World uses the Unreal Tournament 2003 game engine.

The Mission Engine represents each character in the story as an agent with its own goals, relationships with other entities (including the learner), private beliefs and mental models of other entities, and a high-level director agent that influences the character agents, to control how the story unfolds and to ensure that pedagogical and dramatic goals are met. A User Manager module handles user requests for nonverbal gestures, invokes the speech recognizer when the learner wishes to speak, and passes the recognized utterances and gesture requests to the Mission Engine which in turn directs the learner's avatar to perform the requested communicative acts and models the responses of the other characters.

4.2 Mission Skill Builder

When the Mission Skill Builder presents lessons and exercises, it invokes the speech recognizer to process learner utterances, and invokes the Pedagogical Agent to evaluate the utterances and choose responses. Lessons include a mixture of images, video clips, and sound clips, and are generated automatically from lesson descriptions in the Curriculum Database. The agent's responses are also selected from a library of sound and video clips. We have also incorporated into the MSB visualizations of speech articulation created using the Baldi toolkit (Massaro, 2004). The MSB's interface is implemented in ToolBook.

4.3 Shared Services and Databases

Curriculum and tutorial materials are stored in the Curriculum Database. It includes exercises that are organized in a recommended sequence, grammar explanations that are automatically linked with the phrases that employ the grammatical constructs, as well as instances of form feedback that are chosen dynamically.

The Language Model includes a speech recognizer that both applications can use, a natural language parser that can annotate phrases with structural information and refer to relevant grammatical explanations, and an Error Model which detects and analyzes syntactic and phonological mistakes.

The Pedagogical Agent determines how the virtual tutor in the MSB should respond to learner actions. It includes a Learner Model, summarizing the skills that the learner has mastered, that other system components such as the MPE can reference as needed.

4.4 Speech Processing and Error Detection

Like most applications of speech technology for language learning, the TLTS needs to b to process highly variable learner speech in a robust fashion. Initial efforts focused on acoustic modeling for robust speech recognition especially in light of limited domain data availability (Srinivasamurthy and Narayanan, 2003). In this case, we bootstrapped data from English and Modern Standard Arabic and adapted it to Levantine Arabic speech and lexicon. Dynamic switching of recognition grammars was also implemented, allowing the recognizer to focus on recognizing the words and phrases that are likely to occur in a given learning context.

Speech recognition in the MPE currently focuses on recognizing the most likely utterance from among a set of utterances that are appropriate in a given scene. This enables the system to simulate dialog with non-player characters. If the recognizer recognizes a phrase that is inappropriate in the current dialog context, the non-player character indicates that he does not understand, allowing the learner to try again. If the recognizer repeatedly fails to recognize an appropriate utterance, the aide intervenes with a suggestion. This enables the dialog to move forward even if microphone problems interfere with recognition.

The MSB, in contrast, uses speech recognition to detect learner errors, so that the pedagogical agent can provide form feedback. For each lesson or exercise, a recognition grammar is loaded that detects correct responses for that context as well as likely learner errors. The learner response, the recognition confidence score, and the history of learner errors determine the type of form feedback that the learner receives. More detail about the speech recognition and pronunciation error detection methods may be found in a companion paper in this volume (Mote et al., 2004).

5 Evaluations

System and content evaluation is being conducted systematically, in stages. Usability and second language learning experts have evaluated and critiqued the learner interface, content organization, and instructional methods. Learner speech data are being collected, to inform and train the speech recognition models. Learners at the US Military Academy and at USC have worked through the first set of lessons and scenes and provided feedback. Formative evaluation studies with learners of varying abilities are in progress in Spring 2004. Learners complete a learning session, after which they comment on the utility of the system and its components and are assessed on the skills that they have acquired. The first evaluations were conducted with learners who claimed that they had limited language learning skills. The subjects indicated that they found the system to be more effective than classroom language instruction that they had received in the past, and some of them commented that they would be happy to take the system home and spend time with it instead of playing video games or other recreation. The speech recognizer was falsely detecting pronunciation errors in some cases, which limited its effectiveness as a training aid. An improved version of the speech recognizer has since been incorporated, and further formative evaluations are currently in progress.

6 Conclusion

The Tactical Language Training System is still in development, yet it has already made rapid progress in combining intelligent tutoring, pedagogical drama, and speech recognition technologies into an integrated language learning system. Formative evaluations have demonstrated that this approach is effective at promoting communication skills, even with learners who profess to have limited foreign language aptitude.

Game scenes and lessons continue to be developed. The US Special Operations Command plans to start putting the Lebanese Arabic training system into use in the classroom starting in late 2004. Meanwhile a version for Iraqi Arabic is under development. Other planned work includes integration of the authoring tools, automated tracking of learner focus of attention to help diagnose learner difficulties, summative evaluations of instructional effectiveness, and tools for multi-lingual authoring.

7 Acknowledgements

The project team includes, in addition to the authors, CARTE members Carole Beal, Ulf Hermjakob, Catherine M. LaBore, Andrew Marshall, Yousef Nouhi, Dimitra Papachristou, David V. Pynadath, Mei Si, Gladys Saroyan, Tania Baban, Nadim Daher, Chirag Merchant and Brett Rutland. From the US Military Academy COL Stephen Larocca, John Morgan and Sherri Bellinger. From the USC Viterbi School of Engineering Naveen Srinivasamurthy, Abhinav Sethy, Jorge Silva, Joe Tepperman and Larry Kite. From Micro Analysis and Design Ursula Lauper and Anna Fowles-Winkler. From the USC School of Education Harold O'Neil, and from UCLA CRESST Eva Baker. This project is part of the DARWARS initiative sponsored by the US Defense Advanced Research Projects Agency.

References

J. Bernstein, A. Najmi, and F. Ehsani. 1999. Subarashii: Encounters in Japanese Spoken Language Education. *CALICO Journal*, 16 (3): 361-384.

W.H. DeSmedt. 1995. Herr Kommissar: An ICALL conversation simulator for intermediate German. In "Intelligent language tutors: Theory shaping technology", V.M. Holland, J.D. Kaplan, & M.R. Sams, eds., pages 153-174. Lawrence Erlbaum, Mahwah, NJ, USA.

C.J. Doughty and M.H. Long. 2003. Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology,* 7(3): 50-80.

H. Hamberger. 1995. Tutorial tools for language learning by two-medium dialogue. In "Intelligent language tutors: Theory shaping technology", V.M. Holland, J.D. Kaplan, & M.R. Sams, eds., pages 183-199. Lawrence Erlbaum, Mahwah, NJ, USA.

W.G. Harless, M.A. Zier, and Duncan, R.C. 1999. Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method. *CALICO Journal* 16 (3): 313-337.

V.M. Holland, J.D. Kaplan, and M.A. Sabol. 1999. Preliminary Tests of Language Learning in a Speech-Interactive Graphics Microworld. *CALICO Journal* 16 (3): 339-359.

W.L. Johnson, J. Rickel, abd J. Lester. 2000. Animated pedagogical agents: face-to-face interaction in interactive learning environments. *IJAIED* 11, 47-78.

S. Marsella, W.L. Johnson, and C.M. LaBore. 2003. An interactive pedagogical drama for health interventions. In "Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies", U. Hoppe & F. Verdejo, eds., pages 341-348, IOS Press, Amsterdam.

D. Massaro. 2004. Symbiotic value of an embodied agent in language learning. Proceedings of the HICCS Conference.

N. Mote, A. Sethy, J. Silva, S. Narayanan, and L. Johnson. 2004. Detection and modeling of learner speech errors: The case of Arabic tactical language training for American English speakers. In Proceedings of the InSTIL Symposium on NLP and Speech Technologies in Advanced Language Learning Systems.

S. Srinivasamurthy, N. and Narayanan, S. 2003. Language-adaptive Persian speech recognition. in Proc. Eurospeech 2003.

W. Swartout, R. Hill, et al. 2001. Towards the holodeck. In "Proceedings of the Intl. Conf. on Autonomous Agents", J. Müller, E André, & S. Sen, eds., pages 409-416. ACM Press, New York.

S. Witt, and S. Young. 1998. Computer-aided pronunciation teaching based on automatic speech recognition. In "Language teaching and language technology", S. Jager, J.A. Nerbonne, & A.J. van Essen, eds., pages 25-35. Swets & Zeitlinger, Lisse.