Preliminary Tests of Language Learning in a Speech-Interactive Graphics Microworld

V. Melissa Holland
U.S. Army Research Laboratory

Jonathan D. Kaplan and Mark A. Sabol
U.S. Army Research Institute for the Behavioral and Social Sciences

ABSTRACT

A speech-interactive graphics microworld is described in which learners speak problem-solving directions to an animated agent and in which new scenarios can be authored. A proof-of-concept application is illustrated for sustaining basic speaking skill in Modern Standard Arabic. Preliminary tests of the application are summarized involving learners from both university and military settings. Problems are discussed in predicting and measuring learning gains based on brief exposure to new technologies.

KEYWORDS
Microworld, Speech Recognition, Authoring, Evaluation, User Trials, Graphical Interface

INTRODUCTION

Research initiatives at the Army Research Institute (ARI) and the Army Research Lab seek to integrate advanced methods, such as natural language processing (NLP) and automated speech recognition (ASR), into language tutors and translation systems. One product of the ARI initiative is the Military Language Tutor (MILT), first developed for a military audience but intended generally for adults learning language. Here we describe a version of MILT that couples discrete ASR with animated graphics to give job-relevant communicative practice in selected languages. This system has undergone preliminary tests with both military and university students, as detailed in Kaplan, Sabol, Wisher, and Seidel (in press). In this paper, we describe the design and instructional rationale of the system, highlight selected test results, report new data, and sketch problems involved in measuring learning gains from small prototype systems. We also discuss issues in applying continuous speech recognition to enable unscripted utterances by language learners.

Design Goals

To help adult learners who need to speak another language for their jobs, we sought to develop computer assisted language learning (CALL) that is speech enabling, job relevant, motivating, and portable. These goals were shaped in particular by the needs of our initial audience, soldiers who use (or must be ready to use) languages other than English. Speaking is often an important part of the job for these soldiers, but it is also a skill that declines rapidly after they complete language school, usually from lack of practice (Lett & O'Mara, 1990). To be relevant, and to motivate spare-time practice of the target language, our CALL technology needed to support realistic interactions that are fun as well as typical of work the learner does. We settled on a CALL design that incorporates ASR into an activity known as a microworld. In this activity learners use the target language in an interactive graphics environment that approximates the dialogue games of commercial adventure software and can, for example, simulate a military mission. To stretch the benefits of the microworld, a final design goal was to have an extendible system that could be changed or expanded by instructors with no programming experience.

The Microworld Concept

The benefits of microworlds for language learning have been articulated by Schoelles and Hamburger (1996) and demonstrated in implementations by Hamburger and Hashim (1992), Douglas (1995), and Tomlin (1995). These self-contained, reactive environments support a communicative approach to language instruction and carry out many of the conditions of immersion discussed by Eskenazi (this issue). In the MILT microworld, learners are immersed in a task-relevant environment that they can explore by speaking or typing commands to an animated agent who understands only the target language (much like the TraciTalk agent described by Wachowicz and Scott, this issue). The agent solves problems such as searching a series of rooms for hidden documents (books, letters, maps, etc.) and reading and extracting the information contained.

Learning Principles

Assuming that language learning is part of the general problem of skills acquisition, we have tried to draw our CALL design from basic principles of learning and cognition (Anderson, 1983; Anderson, Kulhavey, & Andre, 1971; Gagne & Briggs, 1979; Newell, 1990; Schmidt & Bjork, 1992; Schmidt, 1990). Three principles were influential: *implicit feedback*, *overlearning* through *intrinsic reward*, and *adaptive sequencing*. The microworld promotes implicit feedback and intrinsic reward, while adaptive sequencing is built into the overall design of MILT.

Implicit feedback comes when learners see the natural consequences of their language use—their utterance is either understood as intended (the microworld agent carries out the right action) or misunderstood (the microworld agent carries out the wrong action or says, "I don't understand") (see Wachowicz & Scott, this issue). It is generally more effective in overall language learning than explicit feedback, which reports on correctness or incorrectness (Robinson, 1991). This learning-by-doing principle applies to acquiring other kinds of skills (e.g., Collins, Brown, & Newman, 1989).

Overlearning comes from repeating training beyond the point of apparent mastery. According to learning theory and research, overlearning builds automaticity, enabling performance under stress and supporting retention of skills over periods of nonuse (Driskell, Willis, & Copper, 1992; Schendel & Hagman, 1982). Repeated practice can be motivated by instruction that is intrinsically rewarding. Intrinsic rewards come from the behavior itself, such as satisfying curiosity or solving puzzles, and have been found to benefit learning more than extrinsic rewards, such as praise or candy (Berlyne, 1968; McClelland, 1961; Malone, 1981). We hypothesized that an engaging problem-solving environment would lead learners to practice language toward overlearning.

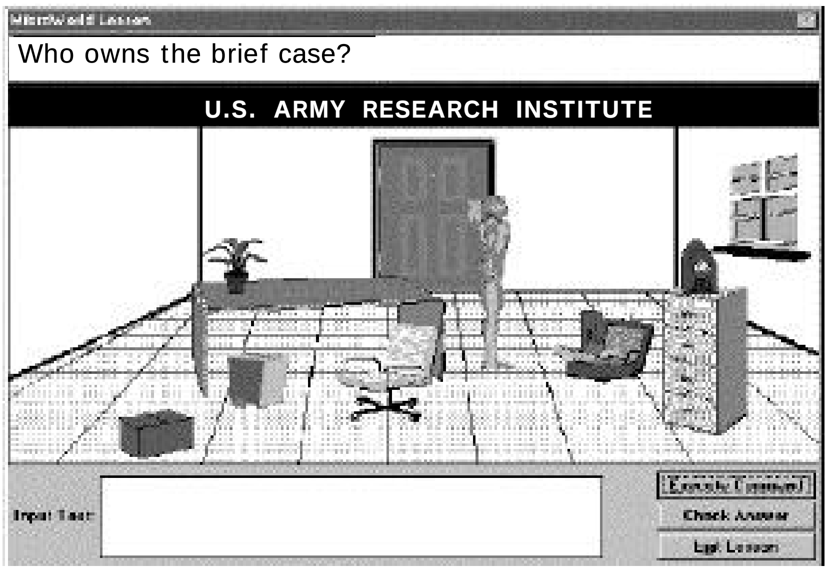
Adaptive sequencing, the individualization of instruction to accommodate particular learners' problems, has been shown to streamline learning (Anderson, Conrad, & Corbett, 1989; Atkinson, 1976; Park & Tennyson, 1983; Schmidt, 1990). In MILT authors can define error thresholds such that making above-threshold errors on an exercise branches the learner to remedial instruction (Kaplan & Holland, 1995).

TYPING TO THE MICROWORLD: FLEXIBLE COMMANDS

Activity Design

The first MILT microworld, sampled in Figure 1, accepted typed rather than spoken input. A scenario-defining problem (e.g., "Who owns the briefcase?" or "Where will the enemy attack?") was given at the top of the screen. Learners constructed their own simple-sentence commands using vocabulary from the scene, for example, "Put the book on the table" or "Read the map." Constraints on what learners could say (the vocabulary understood by the agent) were available in a help window. A proof-of-concept activity was developed that took keyboard input in English, Spanish, and Modern Standard Arabic.

Figure 1
Microworld Exercise: Problem Solving

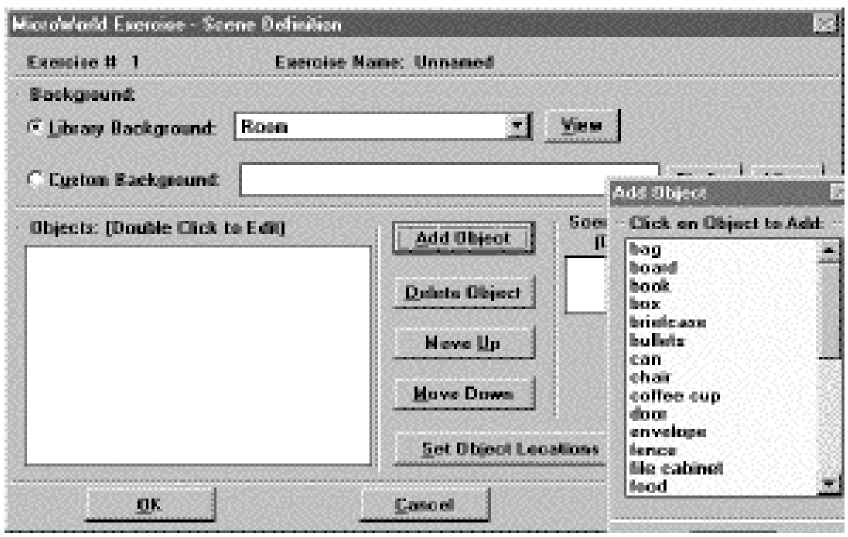


Note: The learner solves a problem (at top of screen) by typing commands into the blank box to direct the on-screen agent. This frame shows a book that has been opened by the agent in response to learner's command.

Authoring

To meet the goal of extendibility, we made the microworld alterable on some dimensions. Underlying the microworld is a fixed pool of actions and objects whose properties are constant and whose locations can be changed in a room (or series of rooms). Instructors can define new scenarios by writing new problem statements, reselecting objects from the pool, rearranging them, inserting some objects inside of others, and customizing the background. Figure 2 shows the interface for selecting objects for a new scenario. The "Set Objects Location" button moves the author to a drag-and-drop screen where these objects can be positioned in a selected background. The readable objects (book, envelope, letter, newspaper) can be rewritten to fit new scenarios. The authoring interface to the microworld is entirely menu- and template-based for easy use by nonprogrammers (detailed by Kaplan et al., in press).

Figure 2
Authoring Interface for Selecting Objects for a Microworld Scenario



Instructors authoring new scenarios did not have to anticipate new input by learners. Variation in learners' commands was to be automatically handled by NLP and by microworld rules for linking verbs and nouns to actions and objects.

Simplifying the Computation

Like earlier tutors in the MILT program (Holland, Maisano, Alderks, & Martin, 1993; Criswell, Byrnes, & Pfister, 1992; Kaplan & Holland, 1995), the keyboard entry microworld sought to incorporate NLP both to detect errors and to interpret learners' typed-in sentences produced in the microworld and other exercises. The linguistic mechanisms underlying the NLP have been previously described (Weinberg, Garman, Martin, & Merlo, 1995; Dorr, Hendler, Blanksteen, & Migdaloff, 1995). For a variety of reasons, these tutors did not achieve the robustness needed to go into the hands of students for testing. The NLP was fragile and never succeeded in handling the range of input actually produced by learners in exercises. However, demonstrations with teachers and students in military settings proved the overwhelming popularity of the microworld idea. We therefore implemented a computationally simpler approach—key word matching from the learner's input string—to interpret commands and drive the microworld action. This approach proved effective in our internal tests. It was about as flexible as and more robust than NLP. However, our test users sometimes produced vocabulary not in the key word set and tended to ignore the help option.

While exploring the key word approach, we also looked at the even simpler method of letting users choose from precomposed sets of commands in the microworld. To maintain relevance and motivation, we wanted learners to speak these commands. This plan fit well with prior designs for using ASR in CALL (Bernstein, 1994; Bernstein & Franco, 1996), which, to maintain high recognition accuracy, fix the learner's utterances beforehand as part of the expected, active vocabulary of the recognizer.

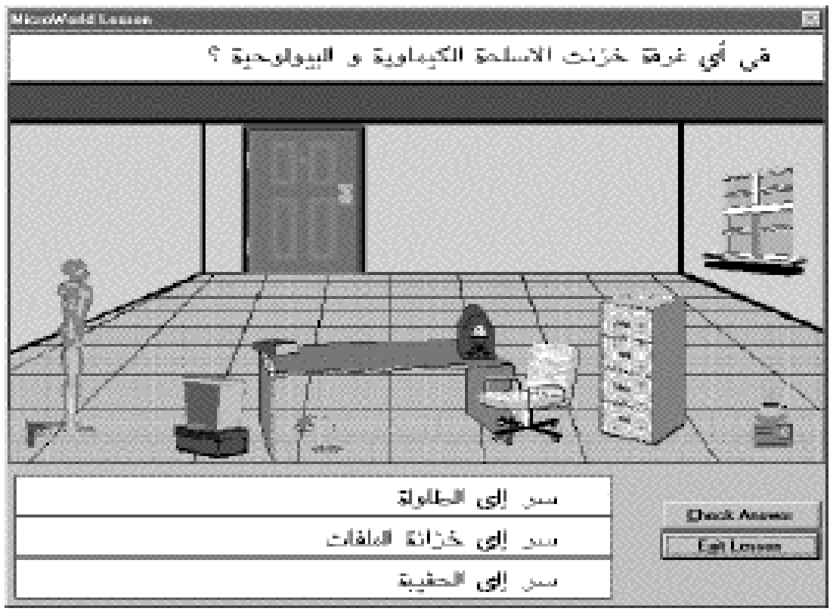
SPEAKING TO THE MICROWORLD: FIXED COMMANDS

Activity Design

We have developed a proof-of-concept speech interface in Modern Standard Arabic to the MILT microworld, as shown in Figure 3. The learner sees a multiple choice set of commands to say to the microworld agent displayed beneath the action frame and reads one of the commands aloud into a microphone. Such an arrangement is typical of most speech-interactive CALL systems, and our particular arrangement is modeled on the multiple choice display in Virtual Conversations (Harless & Zier, this issue), a speech-interactive video-based learning environment.

We created a demonstration scenario much like that for the keyboard microworld: to decipher information from hidden documents and radio messages. We then developed a pool of 72 short Arabic commands based on those used in the keyboard version of the microworld ("Walk to the file cabinet," "Open the drawer," "Turn on the radio," etc.). In storyboarding the scenario, we selected from this pool three commands per frame, any of which are logical things to say and advance the scenario in different ways. Learners could enter the same microworld scenario several times, choosing a different route each time, in keeping with implicit learning principles. (Rypa & Price, this issue, call this practice "branching dialogues.")

Figure 3
Speech-Interactive Microworld Exercise



Note: The learner says one of the three utterance choices in Modern Standard Arabic to direct the on-screen agent.

Speech Recognition

Utterances in the microworld are processed by the discrete speech recognizer from Dragon Systems, trained on the 72 Arabic sentences in the microworld pool. Discrete ASR permits recognition of separate words or of a sentence uttered with pauses between words. To recognize a sentence spoken naturally without pauses between words, discrete ASR must treat the sentence as a single long word already in its recognition vocabulary. This sentence needs to be a fixed entity, hence the practice of writing it on the screen for the speaker, who cannot know ahead of time which sentences are expected.

PREPARATION AND REMEDIATION

Available before and during the microworld is a familiarization lesson in which all 72 sentences from the microworld pool are displayed. Learners can hear a native speaker pronounce each sentence, hear and compare their own attempts at pronunciation, and get a translation of each sentence into English. This lesson also serves as remediation for learners who cannot solve the microworld problem or who give up. In an adaptive tutoring sequence, these learners branch from the microworld to the remedial lesson. They can try the microworld again at any time.

AUTHORING

Like the keyboard entry microworld, the speech-interactive version is authorable through a menu- and template-based interface. This interface gives authors the same choices as for the keyboard entry microworld (see Figure 2), with the addition of an utterance selection facility, shown in Figure 4. In the keyboard entry microworld, the author of a new scenario does not have to prescribe what learners can type in because either key word matching or NLP automatically interprets learners' commands. In the speech-interactive microworld, the author must predefine the utterance choices to be displayed to the learner at each frame. At the simplest level, the author can draw from a fixed pool of utterances. Adding new utterances is more complex, requiring some familiarity with a speech development kit. We opted for simplicity and put in place a facility to select from the pool of 72 Arabic commands on which the recognizer was trained for the demonstration scenario. An author can combine these commands as desired, using the screen in Figure 4 to select utterances in sets of three to be displayed at the bottom of the student's screen. The author uses a branching tree to define the sequences of utterance choices to be shown to the student and activated within the recognizer.

This procedure can produce lots of practice from few resources. Moreover, limiting to three the number of utterances to be distinguished at any one time improves the performance of the recognizer.

Figure 4
Authoring Interface for Selecting Utterance Choices for a Microworld Frame

Note: The frames are defined by the numbering system shown in the branching tree.

PRELIMINARY USER TRIALS OF THE SPEECH-INTERACTIVE MICROWORLD

University Students: A Pilot Study

PURPOSE

We conducted a pilot study of the proof-of-concept microworld to look at the robustness and overall acceptance level of the activity. Would students tolerate the inevitable ASR errors? Would they accept the forced choice of utterances in solving the microworld problem? Would they enter the world created by the graphics environment and want to stay there through rounds of sometimes repetitious utterances?

METHOD

Six students from Georgetown University participated. Each had had two years of Arabic, putting them at about the proficiency level we expected of military users—level 1 to 1+ on the 5-point Interagency Language Roundtable scale.

To estimate the overall acceptance level for the microworld, we gave the last four students an attitude questionnaire after they had tried the microworld, including the following key questions:

- "Was the speech-interactive microworld enjoyable?"
- "Is the microworld a good idea for language learning?"
- "Would you choose to use a longer, more advanced version of the microworld as part of your regular language learning [sustainment training]?"

Responses were to be marked on a 4-point scale, with positive and negative anchor points (e.g., "yes," "no"). We also observed each student perform and solicited his or her informal comments afterwards.

RESULTS

All six students responded favorably to the microworld, despite the fact that the recognizer sometimes failed to recognize their utterances, forcing them to repeat. Those who received the questionnaire rated the microworld as an "excellent idea," "very good idea," or "good idea"; none rated it as "not a good idea." Questionnaire responses indicated that students would like to see the microworld approach used in language instruction "as often as possible," "much more often," or "sometimes," with no "not at all" responses. All students responded, "It [the microworld game] forced me to practice much more," suggesting that one of our original goals—promoting practice toward overlearning—might be met. It appeared that the frustration of being misrecognized and having to repeat was more than compensated for by the game, whether due to its intrinsic interest or its temporary novelty.

Military Users: A Preliminary Field Trial

PURPOSE

We conducted a further study, preliminary to a formal evaluation, to observe acceptance levels among a military audience and to try out measures of learning based on the microworld experience. How would soldiers like using the Arabic microworld? What aspects of language skill were likely to improve? How could we measure them?

METHOD

SUBJECTS

Sixteen soldiers in the Fifth Special Operations Group at Fort Campbell, KY, participated in the study. Each of the 16 had studied Modern Standard Arabic and was required to maintain skill in MSA for possible future missions. Based on yearly tests of reading and listening by the Defense Language Institute, we selected subjects such that half were rated 1+ or higher (averaged between reading and listening) on their last Defense Language Proficiency Test (DLPT) and half were rated 0+ to 1. Thus, we could observe interactions between learners' tested language proficiency levels and their attitudes and learning gains.

MEASURES OF ATTITUDE AND LEARNING

Measures were taken of (a) attitude toward the microworld activity and (b) effect of the activity on language learning. Attitude was measured using the questionnaire from the pilot study. Effect on language learning was measured through pretest-posttest differences in subjects' basic sentence-making skill. We reasoned that learners would acquire basic elements of sentence making through speaking to the microworld:

- vocabulary for the component objects and actions;
- syntactic structure for imperatives, with attendant agreement rules;
- pronunciation on both phonetic and prosodic dimensions (see Eskenazi, this issue);
- fluency as the leaner gets repeated practice in putting these elements together.

These are distinct elements of oral production skill in the psycholinguistic literature. We estimated that the simplest way to elicit these elements was through a translation task, which calls for knowledge of target language vocabulary, grammar, and pronunciation as well as enough automatic control of these elements to support fluent (unfaltering) speech. We took the 72 Arabic utterances that compose the microworld pool as a baseline and reasoned that if our measures were meaningful, they would at a minimum reveal improvement on those utterances. We tested sentence-making skill in regard to these utterances by giving subjects a written English version of each command and asking them to speak the commands in Arabic (i.e., oral translation).

Scoring Procedure

Each of the four elements of sentence-making skill was rated separately for each attempted Arabic utterance: vocabulary, grammar, pronunciation, and overall fluency. Each element was rated on a 5-point scale, with 0 being “entirely negative or no attempt” and 5 being “excellent, perfect.” The rater was a native speaker of Arabic with a Ph.D. in Arabic studies and linguistics and expertise in second language learning. He followed a blind rating procedure, not knowing whether utterances came from pre or posttest

Testing Procedure

Each student was given 1 hour of practice with the microworld, preceded by the translation pretest and followed by the translation posttest and attitude questionnaire. In an effort to counterbalance sentence presentation order between pre and posttest, two versions of the 72 sentences were prepared: one version consisted of the sentence order 1-36, 37-72; the other version, the order 37-72, 1-36. Half the subjects received one version and half the other version as pretest; then each subject received the opposite version as posttest. Utterances were tape recorded.

Caveats on the Method

Clearly, we would expect sentence-making skill to improve between pre and posttest for the specific utterances on which a learner is trained. This first set of measures establishes a baseline to see whether learners improve on at least the sentences in the lesson. This baseline also indicates whether the sentence-making measures in a translation task are consistent and interpretable for later application in testing learning gains.

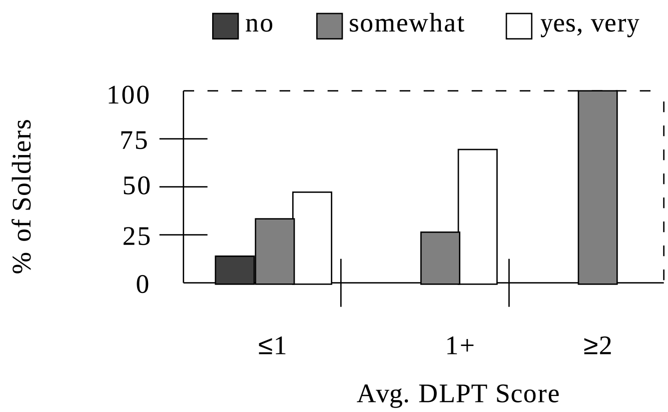
Results

Attitudes

Like the university students, the soldiers in our sample uniformly liked the microworld lesson and indicated they wanted to use that approach for future language sustainment. Subjects who had low to moderate DLPT scores were more positive than those with higher scores. Figure 5 shows how subjects in three DLPT-level groupings answered the question "Was the microworld enjoyable?" Subjects with DLPT scores < 2 called it "very enjoyable" at rates of 75% (subjects at 1+ level) and 50% (subjects at ≤1 level), whereas those with scores ≥ 2 called it "somewhat enjoyable" (see Figure 5).

Figure 5
Relationship between Soldier's DLPT Scores and their Attitude toward the Speech-Interactive Microworld

Question: Was using the speech microworld enjoyable?

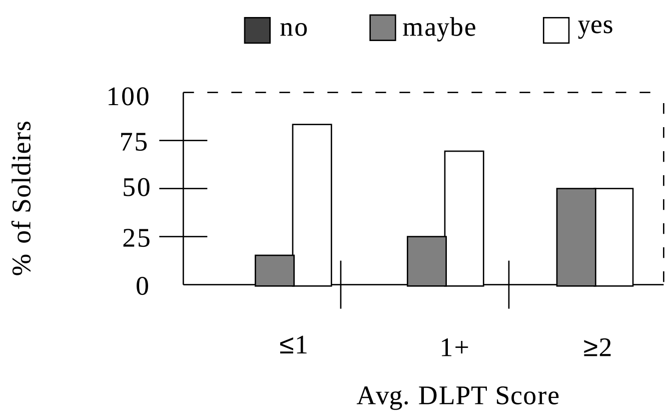


Note: The DLPT scores are an average of listening and reading subscores.

Responses to other attitude questions showed a similar trend. To the question of whether the speech-interactive microworld was a good idea, 75% of subjects at DLPT < 2 called it an "excellent idea" whereas 100% of those at DLPT ≥ 2 called it "good" to "very good." To the question, "Would you use longer, more advanced versions of the microworld as part of your regular sustainment training," 80% of subjects at DLPT levels < 2 responded "yes" whereas 50% of subjects at DLPT ≥ 2 responded "yes." The remaining respondents said "maybe" to this question, and no one said "no." These responses are summarized in Figure 6.

Figure 6
Relationship between DLPT Scores and Anticipated Use of Microworld

Question: Would you use advanced microworld for sustainment training?



The differences in attitude by respondents' proficiency level are not surprising. Our implementation of the microworld with its simple-sentence commands, basic vocabulary, and many repeated utterances was aimed toward users with lower proficiency, whereas users with higher proficiency are more likely to tire of games at this level. Like the university students, the soldiers felt the microworld led them to practice utterances much more than they normally would have practiced.

Informal observations during the field trial revealed that, as with the university students, the soldiers experienced some frustration with the tutor's failure to recognize their utterances. Recognition failures were more common at the beginning of the activity. This may be because users' pronunciation improved in the course of using the tutor or because they found ways to adapt their voices in specific ways to the recognizer. The fact that limitations in ASR technology did not diminish learners' enthusiasm for the activity is reminiscent of an early study of a rudimentary ASR-based language tutor developed for soldiers at Fort Bragg, NC (EER, 1991). Despite significant misrecognitions, soldiers said they liked the speech interactivity and wanted more of it in CALL.

Learning Effects

The mean ratings of Arabic utterances produced in the posttest (after instruction in the microworld) were nearly double those of the pretest on each of the four dimensions rated: vocabulary, grammar, pronunciation, and fluency. (Means were averaged across all 72 sentences for all 16 participants). The pretest-posttest learning gains for 14 of the 16 subjects were statistically significant according to analyses outlined by Kaplan et al. (in press), with the greatest gains occurring in those participants whose prior proficiency was at the 1+ level.

Whether from the CALL intervention or simply from the pretest exposure, some learning occurred for the specific sentences tested (and trained). Although these gains tell us little about the effectiveness of the CALL intervention, they do suggest that ASR-based instruction did not impede learning and that the rating measures from the translation task are viable for further development.

Summary and Discussion

The microworld was viewed favorably by language learners in both the university and the military. All respondents, whether above or below the median DLPT score, said that the microworld was a good idea for language learning and that they wanted to use it more. Many noted that the game format led them to practice more, supporting our hypothesis that the microworld by its intrinsic motivation could promote overlearning. Data on sentence making were easy to obtain from a translation task and simple to score. In addition, the scores gave expected results: Learners improved on the sentences that were trained and these improvements were consistent across rated elements. We propose to refine these measures for use in more controlled tests of instructional effectiveness.

The trial reported here is preliminary to a full evaluation of the speech-interactive microworld. First, that evaluation requires control conditions, comparing the microworld to other kinds of instruction (the microworld without speech, speech with nonmicroworld activities, etc.) and to no instruction (learning gains may be due merely to the repeated test). Second, reliability and validity must be established for the sentence-making measures. We need correlations between raters as well as between elements being rated (e.g., are pronunciation and fluency independent?). We also need to validate ratings against standard measures of proficiency, such as DLPT scores, course grades, or specific linguistic features of utterances. Finally, the test set must be expanded to include sentences not in the training set to see how the instruction generalizes, and pretest and posttest sets must differ. Generalization can be further assessed by measuring utterances elicited through a task other than translation, such as picture description.

It is obvious that a full evaluation of speech-interactive microworlds should be based on more instruction covering more content. To produce gains on instruments like the DLPT or end-of-course exams takes exposure to and practice with massive amounts of language. We therefore need to find other measures, narrower but predictive, to assess the effects of prototype systems and methods before trying to implement them on a large scale. The quest for metrics is, in fact, central to our experimental work in language learning and language technology in general (e.g., Voss & Reeder, 1998).

SPEAKING TO THE MICROWORLD: FLEXIBLE COMMANDS THROUGH CONTINUOUS SPEECH RECOGNITION

Advantages of fixed-utterance recognition include relatively high certainty of recognition and relatively low development expense. We are still a long way from supporting practice of natural spoken dialogue. To enable flexible utterances that do not involve reading from the screen calls for judicious application of continuous speech recognition.

ARI is exploring continuous speech recognition in collaboration with the U.S. Military Academy at West Point, using Entropic's HTK (Hidden Markov Model Tool Kit), an ASR development environment for building speaker independent, continuous ASR systems (LaRocca et al., this issue). An acoustic model for MSA was developed by training on a large sample of native speakers. A language model was also built based on all possible commands from the microworld vocabulary. We first explored the utility of the HTK recognizer in providing measures for some key design comparisons. In the preparation exercise that precedes the MILT microworld, learners can click on a written sentence to hear a native speaker say it and can also record and hear themselves. Which works better for shaping learners' pronunciation: to hear the native speaker without hearing themselves or to hear themselves played back for comparison with the native speaker? In the playback condition, after the first time hearing the native speaker, are repeated attempts more effective if the learner goes first or the native speaker goes first?

Sabol (in press) studied these questions with students who had no prior knowledge of MSA. Confidence scores output by the HTK speech recognizer (reflecting the probability that it accurately recognized a particular string) were taken as a measure of goodness (nativeness) of pronunciation. In Sabol's analysis of variance on this measure, playback condition did not reach significance but sentence length did ($p < .01$), as did the interaction of length with practice ($p < .05$). That is, confidence scores were not affected by whether learners listened to themselves, nor whether they did this before or after hearing the native speaker. However, scores did increase over repeated attempts, an apparent learning effect, for the shorter sentences only. It would appear that beginners could get sufficient grip on shorter utterances to improve their pronunciation with repeated imitations of a native speaker. Before ASR confidence scores can serve as an index of pronunciation quality, however, they must be validated against scores for native speakers and against expert judgments of pronunciation quality (in this issue see Dalby & Kewley-Port; Rypa & Price): a next step in our research.

We have not at this point applied continuous ASR models to supporting flexible input to the microworld. Such application involves a range of difficulties. Continuous ASR works on models trained on massive amounts of speech data. To recognize learners' utterances for the sake of dialogue, and to recognize their mistakes for the sake of correction, would mean acquiring models of learner talk. These models could be acquired from large-scale samples of learners' speech at varying levels of proficiency, which would have to be collected at considerable expense (see Egan, this issue). Alternatively, a preexisting native-speaker language model could be perturbed to represent expected errors. Knowledge about how to perturb could come from questioning instructors, bootstrapping from limited data, or mining NLP engines developed to process typed input by learners (e.g., the multilingual NLP suites developed for MILT by Weinberg et al. (1995) as well as those described in Bailin (1991)). However, even with well honed models, permitting flexible utterances dramatically reduces the accuracy of speech recognition. The issue then becomes how to use imperfect recognition in the service of learning without frustrating, confusing, or misleading learners. We look to the interface designs described in this volume as a source of ideas.

ACKNOWLEDGEMENTS

MILT was developed with support from the Defense Advanced Research Projects Agency and the DoD Office of Special Technology, and sponsorship by the U.S. Special Operations Command and the U.S. Army Intelligence Center, Fort Huachuca. ARI's prime contractor for the MILT project is Micro Analysis & Design of Boulder, CO. The discrete speech Arabic language model was developed by Interactive Drama, Inc. The views, opinions, and conclusions contained in this paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., Conrad, F., & Corbett, A. (1989). Skill acquisition and the LISP tutor. *Cognitive Science, 8*, 87-129.

Anderson, R.C., Kulhavey, R., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Research, 62*, 148-156.

Atkinson, R.C. (1976). Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, NJ: Lawrence Erlbaum.

Bailin, A. (Ed.). (1991). Intelligent Computer Assisted Language Instruction [Special Issue]. *CALICO Journal, 9*.

Berlyne, D. (1968). Curiosity and exploration. *Science, 153*, 25-33.

Bernstein, J. (1994). Speech recognition in language education. In F. L. Borchardt & E. Johnson (Eds.), *Proceedings of the 1994 CALICO annual symposium: Human factors* (pp. 37-41). Durham, NC: CALICO.

Bernstein, J., & Franco, H. (1996). Speech recognition by computer. In N. Lass (Ed.), *Principles of experimental phonetics*. St. Louis: Mosby.

Collins, A., Brown, J. S., & Newman, S. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum.

Criswell, E., Byrnes, H., & Pfister, G. (1992). Intelligent automated strategies of teaching foreign language in context. In M. L. Swartz & M.Yazdani (Eds.), *Intelligent tutoring systems for foreign language learning: The bridge to international communication* (pp. 321-335). New York: Springer-Verlag.

Dorr, B., Hendler, J., Blanksteen, S., & Migdaloff, B. (1995). On beyond syntax: Use of lexical conceptual structure for intelligent tutoring. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum.

Douglas, S. A. (1995). LingWorlds: An intelligent object-oriented environment for second language tutoring. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum.

Driskell, J. E., Willis, R., & Cooper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*, 615-622.

EER Systems (1991). *Final report: Foreign language training for Special Forces Functional Language Course*. Prepared for U.S. Army Training and Doctrine Command. Newport News, VA: Author.

Gagne, R., & Briggs, L. (1979). *Principles of instructional design*. New York: Holt, Reinhart & Winston.

Hamburger, H., & Hashim, R. (1992). Foreign language learning and tutoring environment. In M. L. Swartz & M.Yazdani (Eds.), *Intelligent tutoring systems for foreign language learning: The bridge to international communication*. New York: Springer-Verlag.

Harless, W., & Zier, M. (1998). *Foreign language training with Virtual Conversations™ technology*. Paper presented at the meeting of the Annual Computer Assisted Language Instruction Consortium. San Diego, CA.

Holland, M., Maisano, R., Alderks, C., & Martin, J. (1993). Parsers in tutors: What are they good for? *The CALICO Journal, 11*(1), 28-46.

Kaplan, J. D., & Holland, V. M. (1995). Application of learning principle to the design of a second-language tutor. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum.

Kaplan, J. D., Sabol, M., Wisher, R., & Seidel, R. (in press). The Military Language Tutor (MILT) program: An advanced authoring system. *International Journal of Computer Assisted Language Learning.*

Lett, J., & O'Mara, F. (1990). Predictors of success in an intensive foreign language context: Correlates of language learning at Defense Language Institute Foreign Language Center. In T. Perry & C. Stansfield (Eds.), *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice Hall.

Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4*, 333-369.

McClelland, D. C. (1961). *The achieving society*. New York: Van Nostrand.

Newell, A. (1990). *A unified theory of cognition*. Cambridge, MA: Harvard University Press.

Park, O. K., & Tennyson, R. D. (1983). Computer-based instructional systems for adaptive education: A review. *Contemporary Education Review, 2*, 121-134.

Robinson, G. (1991). Effective feedback strategies in CALL: Learning theory and empirical research. In P. Dunkel (Ed.), *Computer-assisted language learning and testing. Research issues and practice* (pp. 155-167). New York: Newbury House.

Sabol, M. (in press). *The voice recognition experiment: First results* (ARI Working Paper). Alexandria, VA: U.S. Army Research Institute.

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology, 67*, 605-610.

Schmidt, R. A. (1990). *Feedback for skill acquisition: Preliminaries to a theory of feedback* (Final Report, U.S. Army Research Institute Contract MDA903-85-K-0225). Los Angeles, CA: University of California at Los Angeles, Dept. of Psychology Motor Control Laboratory.

Schmidt, R. A., & Bjork, R. (1992). New conceptualizations of practice. *Psychological Science, 3*, 207-217.

Schoelles, M., & Hamburger, H. (1996). Cognitive tools for language pedagogy. *International Journal of Computer Assisted Language Learning, 9*, 213-234.

Tomlin, R. (1995). Modeling individual tutorial interactions: Theoretical and empirical bases of ICALL. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum.

Voss, C., & Reeder, F. (1998). *Embedded machine translation: Design, construction, and evaluation of systems with an MT component*. Workshop presented at the annual meeting of the Association for Machine Translation in the Americas, Langhorne, PA.

Weinberg, A., Garman, J., Martin, J., & Merlo, P. (1995). A principle-based parser for foreign language tutoring in German and Arabic. In V. M. Holland, J. D. Kaplan, & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum.

AUTHORS' BIODATA

V. Melissa Holland leads the Multilingual Systems Research Team at the U.S. Army Research Lab, including applications of machine translation. Formerly, she led the Language Learning Technology Team at the U.S. Army Research Institute, which developed the Military Language Tutor, and she studied reading and writing processes at the American Institutes for Research, Document Design Center. Her Ph.D. in experimental psychology from Pennsylvania State University focused on language acquisition and psycholinguistics.

Jonathan D. Kaplan is a Senior Psychologist at the U.S. Army Research Institute, where he manages the Dialogue Tutor Projects, aimed at applying advanced language technologies to computer assisted instruction, including the Military Language Tutor. He formerly managed the MANPRINT Methods Development Program, aimed at modeling the application of human factors to complex army systems. For the latter, he received the 1985 Army Research and Development Award. His Ph.D. is from the University of California, Santa Barbara.

Mark A. Sabol is a Senior Psychologist at the U.S. Army Research Institute in Alexandria, VA, where his research has centered on predicting the ability of soldiers to acquire and retain specialized knowledge and skills. He received his Ph.D. in experimental psychology from Bowling Green State University, with concentrations in human memory, psycholinguistics, and sensation and perception, and he taught college-level courses in these subjects for several years.

AUTHORS' ADDRESSES

V. Melissa Holland
U.S. Army Research Lab (AMSRL-IS-CI)
2800 Powder Mill Rd.
Adelphi, MD 20783-1145
Phone: 301/394-3001
Fax: 301/394-2682
E-Mail: mholland@arl.mil

Jonathan D. Kaplan
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave.
Alexandria, VA 22333-5600
Phone: 703/617-8828
E-Mail: kaplan@alexandria-emh2.army.mil

Mark. A. Sabol
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave.
Alexandria, VA 22333-5600
Phone: 703/617-5779
E-Mail: sabol@alexandria-emh2.army.mil

CALICO ‘99

1-5 June 1999

MIAMI UNIVERSITY
OXFORD OHIO

Tuesday-Wednesday	**1-2 June**	**Preconference Workshops**
Thursday-Saturday	**3-5 June**	**Opening Plenary, Sessions, Exhibits, Luncheon, Courseware Showcase, SIG Meetings, Banquet, Closing Plenary**

Plenary speakers
G. Richard Tucker, Carnegie Mellon University
Diane Birckbichler, Ohio State University
Gary Strong, National Science Foundation.

Register online at http://calico.org/calico99.html

Early (before 1 May)	with luncheon & banquet		no luncheon or banquet
Member	$175		$165
Nonmember	$200		$190
K-12 or Community College	$125		$115
Saturday only		$50	
Regular (after 1 May)			
Member	$200		$190
Nonmember	$225		$215
K-12 or Community College	$150		$140
Saturday only		$55	
On-site			
Member	$225		$215
Nonmember	$250		$240
K-12 or Community College	$175		$165
Saturday only		$60	

This year’s conference does not have a designated conference hotel. Lodging is available in residence halls on campus and at motels in the area. For more information, visit CALICO’s web site.

Ascot Travel is the official travel agency for CALICO ‘99 and offers special discount fares on Delta Airlines. Visit CALICO’s web site or call Ascot Travel at 800/460-2471. Be sure to mention you are part of the CALICO group.

For more information, contact CALICO
512/245-1417, info@calico.org, http://www.calico.org.