

Serious Use of a Serious Game for Language Learning

W. Lewis Johnson
Alelo, Inc. and the University of Southern California
ljohnson@alelo.com

Abstract. The Tactical Language and Culture Training System (TLCTS) helps learners acquire basic communicative skills in foreign languages and cultures. Learners acquire communication skills through a combination of interactive lessons and serious games. Artificial intelligence plays multiple roles in this learning environment: to process the learner's speech, to interpret and evaluate learner actions, to control the response of non-player characters, to generate hints, and to assess the trainee's mastery of the skills. AI is also used to assist in the authoring process to assist in the generation and validation of lesson content. This paper gives an overview of the current system, and describes the experience to date in transitioning the system from research prototype into a training system that is in regular use by thousands of users in the United States and elsewhere.

Keywords. Game-based learning, pedagogical agents, second language learning

Introduction

The Tactical Language and Culture Training System (TLCTS) is designed to help learners quickly acquire basic communication skills in foreign languages and cultures. Learners acquire knowledge of foreign language and culture through a combination of interactive lessons and interactive games that give trainees concrete contexts in which to develop and apply their skills. It focuses on spoken communication, nonverbal communication, and cultural knowledge relevant to face-to-face communication. TLCTS is an example of a serious game applied to learning [11]: it utilizes game design techniques to promote learning, e.g., by providing learners with missions to achieve, supporting fluid gameplay in the form of simulated conversations with non-player characters, and continual feedback on learner performance within a game scenario context. It utilizes artificial intelligence to engage in speech recognition and dialog with artificially intelligent characters, and to estimate learner mastery of target skills. It also employs artificial intelligence to assist in the creation and validation of instructional content.

This paper provides an overview of the system and its architecture. It then focuses on the process of transitioning TLCTS from a research prototype into a robust learning tool in wide use in the US military and elsewhere.

1. System Overview

Each TLCTS training course includes the following major components. The Skill Builder consists of interactive lessons focusing on task-relevant communication skills.

The Arcade Game and Mission Game are interactive games that give trainees opportunities to develop and practice communication skills. The Web Wizard provides reference material, including glossaries and explanations of the grammatical structure of the phrases used in the lesson material.

Figure 1 shows example screens from the Tactical Iraqi course, designed to help people learn Iraqi Arabic language and culture. The image on the left shows a cultural notes page from the Skill Builder, which illustrates a common greeting gesture in the Muslim world, the palm-over-heart gesture. The image on the right shows the learner's character in the Mission Game greeting an Iraqi non-player character with that gesture.

The Skill Builder and the game experiences are both speech-recognition enabled. In the Skill Builder learners practice vocabulary and phrases, and complete exercises and quiz items that require speaking and understanding spoken language. In the Arcade Game, the learner gives spoken commands in the target foreign language to direct his or her character to move about a game world. In Mission Game, the learner speaks on behalf of his character. This is taking place in the screenshot on the right side of Figure 1, while the character performs a hand gesture that the learner had previously selected from a menu.

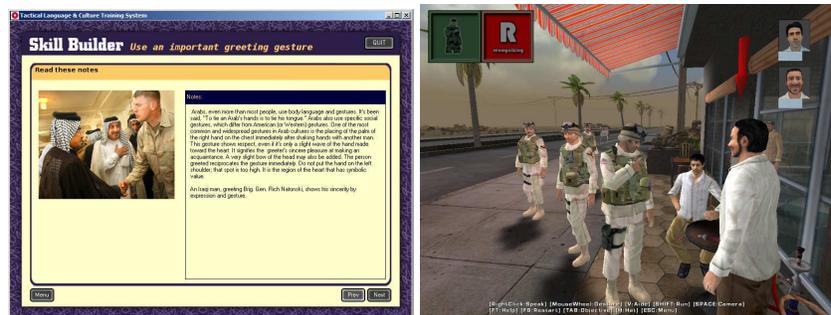


Figure 1. Screenshots from the Tactical Iraqi Skill Builder and Mission Game

The lessons and exercises in the Skill Builder progressively prepare the learner for employing their communication skills in the free-play Mission Game, and ultimately in the real world. Figure 2 shows an intermediate point in this progression, a so-called active dialog. Here the player character (at left) is engaged in a conversation with the head of the household (at right, under the red arrow), in the context of searching a house for weapons and contraband. The screen at bottom shows the most recent phrase that the speech recognizer detected, iftaH il-baab (open the door), and a hint for the next operation to perform (to tell the head of household to put the women and children in a separate room). In active dialogs the learners are guided through the dialog by means of these hints, whereas in the Mission Game the learner does not receive hints unless he or she specifically requests them.

All TLCTS content is highly task-based [4], i.e., instruction focuses on what learners need to know in order to accomplish particular tasks. The task-based approach is very appropriate for game-based learning environments such as this. The task-based approach carries over to the Skill Builder lessons as well, since lessons are focused on acquiring particular skills relevant to particular types of situations. The content development method used for TLCTS content explicitly takes this into account. Each Skill Builder lesson is annotated as to the particular skills that it emphasizes, as are the Mission Game scenes. This helps authors to ensure that the Skill Builder adequately prepares learners for the

Mission Game scenes. The scenes and lessons are automatically cross-indexed in terms of skills, making it possible for trainees to focus in on the particular lessons and lesson pages that they need to study in order to complete the game scenes successfully.

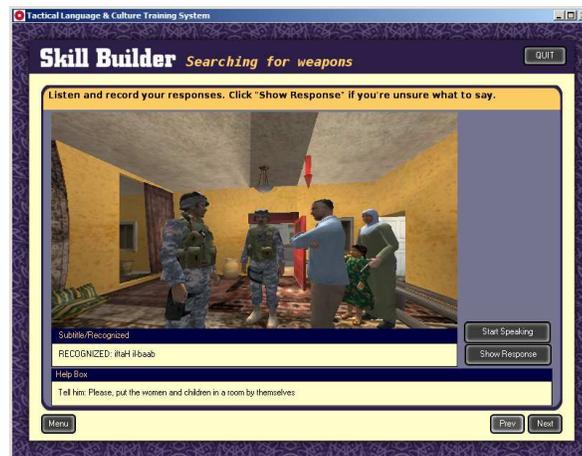


Figure 2. An active dialog in Tactical Iraqi

As the learner works with the software, the software automatically tracks each instance when the learner applies a skill, and uses it as probabilistic evidence of mastery of the skill, akin to knowledge tracing. As Beck and Sison [1] have noted such evidence is inherently uncertain in speech-enabled application where there is a possibility of recognition error, however that can easily be incorporated into a knowledge tracing model by treating recognition errors as just another source of “guesses” (false positives) and “slips” (false negatives). In any case, learners see that the mastery estimates progressively increase with practice, which motivates them to keep practicing until they reach high mastery scores.

While other language learning systems employ speech recognition technology, and support simulated dialogs [2, 3, 6, 7, 8], and employ AIED technology [5], TLCTS is unique in the extent to which it employs these technologies to support the acquisition of face-to-face communication skills. It has been used to develop complete learning environments covering many hours of training.

2. Architecture

The following is a brief summary of the architecture of TLCTS. More detailed descriptions of earlier versions of the architecture are available in other publications [10, 12, 13].

The initial version of the TLCTS prototype was developed using a combination of software tools. The first version of the Skill Builder was authored in ToolBook, and the Mission Game was implemented as an end-user modification (“mod”) of the Unreal Tournament 2003 PC video game. The Web Wizard was implemented in Dynamic HTML and viewed through a Web Browser. This mixed approach was essential in order to develop prototypes quickly that could be presented to stakeholders, however it proved to

be problematic for users. Formative evaluations of TLCTS [9] indicated that this mixed implementation was cumbersome for users, and discouraged them from shifting between components of the learning environment. A new version of the Skill Builder was therefore implemented using Unreal Tournament user interface objects. This transition was accomplished in the following stages. First, an XML specification language was developed for Skill Builder lesson content, and the ToolBook implementation was modified to generate lesson pages from those XML descriptions. Then, a new display generator was created in Unreal that constructs display pages for each page in the XML description. In some cases the same XML description is used to generate multiple display pages: the author specifies a lesson page, and then the page generator automatically generates exercise pages that test the learner's recall of the items on the lesson page.

Once the Skill Builder was integrated into Unreal, learners were more inclined to switch between components of TLCTS as needed. Further steps have since been taken to integrate and simplify the TLCTS architecture, as is described below.

3. Transition into Serious Use

A complete prototype of Tactical Iraqi was completed in June 2005. At this point a new phase of development and evaluation began, leading ultimately to transition into regular use.

Achieving transition of Tactical Iraqi would require overcoming a number of technical and nontechnical obstacles. Tactical Iraqi was developed under sponsorship of a research agency (DARPA). No military service had requested it, or had made plans to acquire it. In fact because it was a research prototype, including new and untested technologies, there was little reason to believe that it would be suitable for regular military use. US military already had access to a range of language learning materials, including an Army-wide license for Rosetta Stone. TLCTS required up-to-date videogame computers to run, which most military units do not have for training purposes. The US military places severe security restrictions on software that runs on its networks, so military units that were interested in using Tactical Iraqi would have to purchase entirely new sets of computers that were not linked to the military network. Finally, military units undergoing training have highly packed training schedules; many military commanders felt that their troops did not have the time to commit to a training program such as Tactical Iraqi, or felt that language and culture training was less important than other types of training.

To start this process, DARPA funded an initial evaluation study with the Marine Corps. A series of two pilot two-week training courses were administered at a Marine Corps training facility at Camp Pendleton, CA, on new laptop computers provided expressly for this purpose. Each training course showed promising results, and also identified problems that were collected in subsequent versions of the software and programs of instruction. For example, in the first training course trainees discovered that they could play the Unreal Tournament 2003 game without Tactical Iraqi, and were tempted to do so; therefore we further modified Unreal so that it could not be run separately. The second pilot evaluation started showing significant promise, while still pointing to areas where further improvements are required. Twenty Marines participated in this study. All had been directed by their command to participate in the study (in Marine jargon, they were "voluntold" to participate). The participants were all enlisted personnel, and only one had any appreciable knowledge of Arabic at the beginning of the

course. Of the twenty participants, nine had previously been deployed to Iraq. Overall, the strongest predictor of success with Tactical Iraqi proved to be whether or not they had previously been to Iraq, and therefore understood the importance of the training provided by Tactical Iraqi. Among the participants who had previously been to Iraq, 78% felt at the end of 50 hours of training that they had acquired a functional ability in Arabic within the scope of the missions being trained. Among this group, the trainees all gave the software a subjective evaluation of at least 4 on a scale of 0 to 5, and the participants who gave it a 4 generally had a number of constructive suggestions about how to make the software better. Among the participants who had not previously been to Iraq the percentage that believed that they had obtained functional ability was much less (22%), and the average subjective evaluation was somewhat lower, 3.73 out of a possible five. One reason why this group gave the software a lower score was that the initial prototype was focused on the requirements of Army personnel. One participant reported that because the software did not teach how to say "I am a US Marine" in Arabic, the software was worthless, and he gave it a rating of 0 out of 5.

The training course and delivery platform was subsequently improved based on feedback from these evaluations and other military beta testers. The platform was made to be configurable, so that users can choose whether to operate the system in Army, Marine Corps, or civilian mode. The choice of user classification determines the dress of the characters in the games in the Skill Builder dialogs, as well as the choice of content in the Skill Builder and the choice of dialog in the Mission Game. Customizations for each class of user were necessary in part because soldiers and Marines have different enlisted military ranks, and therefore use different forms of address from each other and from civilians. To support these customizations, the XML specifications for Skill Builder content were augmented so that individual lesson pages can be annotated as being appropriate for particular classes of users.

Although it was convenient to develop the initial prototype of the Mission Game as a mod, this proved to be a barrier to transition. Users did not want to have to install and run multiple programs in order to run TLCTS. It therefore became necessary ultimately to acquire a license to the Unreal Engine game engine underlying Unreal Tournament, and integrate all capabilities into a single executable package. It was also necessary to rework the user interface to include screens that are more readable and conducive to learning, as shown in the figures shown above. The completed learning environment retains very little of the look and feel of the original game environment, except for some of the keyboard controls used to navigate through the simulated game world.

Software testing is particularly critical for learning environment such as TLCTS, which includes complex AI-based interactive characters, speech recognition, learner modeling, and other advanced capabilities. We therefore had to develop a series of analysis tools and test harnesses to validate the learning content. In some cases these can be applied at authoring time, e.g., to check content for misspellings or words that have not yet been introduced. Some testing is performed when the first executable prototypes of the training systems are created, e.g., to test the dialog models in the non-player characters. A testing interface was created that displays turn by turn the possible dialog moves that the non-player character is expecting that the learner might say next, and then shows the character's response. Such testing frequently reveals possible dialog moves that the authors failed to anticipate, or inappropriate actions taken by the non-player characters.

Evaluation of the speech recognizer was particularly important and challenging. One reason for this is that standard measures of speech recognition accuracy (e.g., word error rate) are not very relevant to speech-enabled learning environments such as TLCTS.

Speech recognition performance needs to vary based on the context in which recognition is performed, and the skills of the learner. In the Mission Game speech recognition is used to recognize the learner's intent (i.e., category of communicative act), whereas the Skill Builder lessons tend to focus more on detecting and correcting common language errors. If a learner is working on an advanced lesson but his or her pronunciation is poor we actually prefer the word error rate to be high, so that learners will be motivated to improve their pronunciation. We therefore collected data sets (recorded by TLCTS) of users using the system in different contexts, and used these both to evaluate and retrain the performance of the speech recognizer. This enabled us to improve speech recognition performance so that it performs acceptably well as needed, with high reliability.

The learning environment itself is just one component of the set of training materials that needed to be developed. User manuals and staff development seminars (known as "train-the-trainer" sessions in the military) needed to be developed. Although educators are familiar with the importance of staff training, such notions are not common in the videogame community. As a consequence many prospective users supposed that a software CD is all that they needed to start training. We have continued to add tutorials and background materials to make the training systems more self-explanatory, and we expect that users can gain training benefit without guidance or orientation. However in a learning environment as complex as TLCTS it is common for learners to overlook important features of the system, or use the system in sub-optimal ways.

Programs of instruction (i.e., curricula) needed to be developed, that provided trainers with guidance as to how trainees should make most effective use of the learning environments. These were developed mainly through collaborations between TLCTS project staff (mainly this author) and early adopters of the learning environment. Early adopters would sometimes develop their programs of instruction, of varying quality. This author then took examples of the better programs of instruction, generalized and extended them and incorporated them into a trainer guide which is made available to all users. The programs of instruction recommend that learners alternate between Skill Builder study and practice in the game components as they progress through the course. Although this might seem obvious, many learners and trainers fail to recognize the importance of this approach.

Summative evaluation of learning outcomes will also be an enabler of transition of TLCTS into use. However such evaluations are feasible only once a commitment has been made to make use of TLCTS in training. In the case of Tactical Iraqi, some military training centers, such as the Expeditionary Warfare School and the Battle Simulation Center at 29 Palms, CA, chose to become early evaluators and trial users of the learning environment. 29 Palms became a particularly important site – it has a laboratory of fifty computers installed with Tactical Iraqi, used both for training and demonstration. Marine units started using the Center in an organized fashion for training starting in the summer of 2006. We have recently collected log data from several hundred Marines who have used the Center, and are currently analyzing their learning gains and performance.

4. Status and Future Work

Up to the present time Alelo and USC have distributed over 6000 copies of Tactical Iraqi and Tactical Pashto. The total number of copies in use is likely to be substantially greater than that, since a number of training centers copy the software and distribute copies to their users. A number of US military posts and training have set up computer labs for training, in the United States, Europe, and in Iraq. Some of these centers have made a

heavy commitment to using the software. For example, the 3rd Infantry Division trained 2,000 soldiers with Tactical Iraqi prior to their next deployment to Iraq. The Marine Corps Expeditionary Warfare School (EWS) has integrated Tactical Iraqi as a required part of its curriculum, and has made the software available on computers throughout the School. Students also receive copies of the software which they can take home with them.

Use of Tactical Iraqi continues to expand. This is happening because learners stakeholders at all levels evaluate it positively and advocate its use. Learners at all ranks who have worked with it are convinced that it is effective in acquiring relevant communication skills. Instructors such as those EWS are convinced of its value, and also report that the current version is relatively free of bugs and technical glitches. Commanding officers have become convinced of the value of this training, and in more than one case general officers have taken it upon themselves to train with Tactical Iraqi, in part so as to set an example for their troops.

Military servicemembers with .mil accounts are authorized to download free copies of Tactical Iraqi; over a hundred copies are downloaded each month, and this number steadily increases. The Marine Corps plans to set up additional training labs for use with Tactical Iraqi. As soon as Tactical Iraqi is approved for use on military networks, it will be distributed across those networks as well.

Meanwhile Tactical Pashto is starting to be used. Extensive use by the 3rd Marine Expeditionary Force in Okinawa and the Marine Corps Mountain Warfare Training Center in Bridgeport, CA in June, 2007.

Additional courses are currently under development. One course, Tactical French, is designed to prepare military personnel to train local military forces in Sahel Africa, with Chad at the notional focus. Another prototype course, Mission to France, is designed for businessmen to help them conduct business in France; this course takes place a slightly fictionalized version of Nice, France. This is the first course developed without a military focus. However, we anticipate that the same task-based approach can be used effectively in this non-military application as well. We are engaged in a collaborative project with McKinley Technology High School in Washington DC to develop the Tactical French. The students in this school are interested in learning about Francophone Africa, and are also interested in acquiring videogame development skills. We therefore plan to engage these high school students initially to test early versions of the training system, and then contribute to the development of the 3D virtual world in the game.

Meanwhile, content development for Tactical Iraqi continues. New lessons and scenarios are being created for new missions of current relevance in Iraq, such as the vocabulary needed to train Iraqi defense forces and conduct patrols. Plans are underway to extend the Tactical Iraqi curriculum up to the point where trainees who complete the course will be able to pass a basic spoken proficiency test in Iraqi Arabic, which will entitle Marines to pay bonuses.

Through this and other steps we anticipate that Tactical Iraqi will become more tightly integrated into the practice of language and culture training in the US military. Military forces in other countries are also interested in Tactical Iraqi and Tactical Pashto, and so we plan to make these courses available to military forces in other countries in the near future.

Research continues to be conducted in a number of areas. A new multilingual content authoring tool named Kona is currently under development, which allows authors specify the content of lesson pages. Specifications may include the source language (e.g., English), the target written language (e.g., written Arabic), phonetic transcriptions used for speech recognition, and pronunciation glosses designed to aid the learner. We automate the generation of some of these transcriptions, e.g., so that pronunciation glosses are generated

automatically following rules specified by the instructor or content author. Prototype Skill Builder implementations have been developed for handheld game players and Pocket PCs, these could provide valuable just-in-time reinforcement for skills acquired in the learning environment. Meanwhile work continues on improving the provision of hints (companion submission to this conference), tracking learner activities in the learning environment against training plans, providing added capabilities for detecting pronunciation errors and providing feedback, and assisting the authoring process.

5. Acknowledgments

The author wishes to acknowledge the members of the development teams at Alelo and USC for their contributions to the work described here. The author also wishes to acknowledge the assistance of the early adopters of TLCTS products throughout the military services. This work was sponsored by DARPA, USSOCOM, and USMC PM TRASY. Its opinions expressed in the paper are those of the authors' and do not reflect those of the sponsors or the US Government.

6. References

- [1] J. Beck, & J. Sison, Using knowledge tracing to measure student reading proficiencies. Proceedings of the 7th International Conference on Intelligent Tutoring Systems, Springer-Verlag, 2004.
- [2] J. Bernstein, A. Najmi, & F. Ehsani, F. Subarashii: Encounters in Japanese Spoken Language Education. CALICO Journal **16** (3) (1999), 361-384.
- [3] W.J. DeSmedt, Herr Kommissar: An ICALL conversation simulator for intermediate German. In V.M. Holland, J.D. Kaplan, & M.R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*, 153-174. Lawrence Erlbaum, Mahwah, NJ, 1995.
- [4] C.J. Doughty & M.J. Long, Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology* **7**(3), (2003), 50-80.
- [5] G. Gamper, G. & Knapp, J.: A review of CALL systems in foreign language instruction. In J.D. Moore et al. (Eds.), *Artificial Intelligence in Education*, 377-388. IOS Press, Amsterdam, 2001.
- [6] H. Hamberger, Tutorial tools for language learning by two-medium dialogue. In V.M. Holland, J.D. Kaplan, & M.R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*, 183-199. Lawrence Erlbaum, Mahwah, NJ, 1995.
- [7] W.G. Harless, M.A. Zier, and R.C. Duncan, Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method. CALICO Journal **16** (3) (1999) 313-337.
- [8] V.M. Holland, J.D. Kaplan, & M.A. Sabol, Preliminary Tests of Language Learning in a Speech-Interactive Graphics Microworld. CALICO Journal **16** (3) (1999) 339-359.
- [9] W.L. Johnson, W.L. & C. Beal, Iterative evaluation of an intelligent game for language learning, AIED 2005, IOS Press, Amsterdam, 2005.
- [10] W.L. Johnson, C. Beal, A. Fowles-Winkler, U. Lauper, S. Marsella, S. Narayanan, D. Papachristou, and Vilhjalmsón, H., Tactical Language Training System: An Interim Report. ITS 2004. Berlin: Springer-Verlag, 2004.
- [11] W.L. Johnson, H. Vilhjalmsón, & S. Marsella, Serious games for language learning : How much game, how much AI ? AIED 2005. IOS, Amsterdam, 2005.
- [12] W.L. Johnson, S. Marsella, & H. Vilhjalmsón, H. The DARWARS Tactical Language Training System. Proceedings of I/ITSEC 2004.
- [13] H. Vilhjalmsón, & P. Samtani, P., MissionEngine : Multi-system integration using Python in the Tactical Language Project. PyCon 2005, 2005.