International Journal of Artificial Intelligence in Education 20 (2010) 175-195
DOI 10.3233/JAI-2010-0006
IOS Press

Serious Use of a Serious Game for Language Learning[1]

W. Lewis Johnson
Alelo Inc., Los Angeles, CA 90066, USA
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 in the authoring process to assist in the generation and validation of lesson content. This article gives an overview of the system, and describes the experience to date in transitioning the system from research prototype into a training system that is in regular use by tens of thousands of users in the United States and elsewhere. Experimental results from field studies are presented, relating learning outcomes, motivational effects, and the role of game-based learning and intelligent tutoring in achieving the learning outcomes.

Keywords. Game-based learning, computer-aided language learning, technology transition

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 (Johnson et al., 2005): 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 providing continual feedback on learner performance within a game scenario context. It utilizes artificial intelligence to enable learners to engage in spoken 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.

Multiple TLCTS training systems have been developed to date. The first prototype, Tactical Levantine, focused on spoken Levantine Arabic. A second learning system, Tactical Iraqi, teaches Iraqi Arabic language and culture. These were prototyped initially at the University of Southern California, and development has since transitioned to Alelo Inc. Alelo has developed additional military courses: Tactical Pashto teaches the language and culture of the Pashtuns in Afghanistan, and Tactical French teaches French as it is spoken in sub-Saharan Africa, along with the culture of Sahel region. Meanwhile Alelo has

[1] This paper is an expanded and updated version of a paper that was originally presented at the 2007 International Conference on Artificial Intelligence in Education.

ISSN 1560-4292/10/$27.50

developed Mission to Iraq, aimed at business people travelling to Iraq, e.g., to aid in reconstruction. Additional courses are being developed for Modern Standard Arabic, Chinese, Cherokee, Dari, and French, for a variety of purposes, including high-school and college instruction and language preservation.

TLCTS is the result of, and continues to be a platform for, adoption-based research in AI in Education. Adoption-based research is a term used by Tak-Wai Chan in his keynote lecture at the 2007 AIED conference, citing Alelo as an example. In the context of this work we give it the following definition: research leading to, and predicated upon, the successful adoption of effective learning systems. A key assumption of adoption-based research is that it is not enough for AIED technologies and methods to prove themselves in the laboratory; they must also be tested in field conditions, used by learners and instructors, and ultimately adopted by them as part of their educational practice. Our goals in adoption-based research are similar to those of the Pittsburgh Science of Learning Center and its *in vivo* classroom experiments, with the additional concern of making sure that the learning system has a life after the *in vivo* experiments are complete.

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. Results from a field evaluation of Tactical Iraqi are presented that relate to the following adoption-related research questions: (1) Is the learning environment effective? (2) Does it enhance motivation, and encourage learners to continue learning? (3) Does the game-based learning approach contribute to the overall learning effect?

SYSTEM OVERVIEW AND PEDAGOGICAL APPROACH

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 Reference 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 military users learn Iraqi Arabic language and culture. The image on the top 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 bottom shows the learner's character in the Mission Game greeting an Iraqi non-player character with that gesture. This pair of images illustrates a curriculum design strategy that is employed throughout TLCTS courses: the Skill Builder introduces knowledge about the foreign language and culture and basic communicative skills, and the game environments give learners opportunities to practice and integrate those skills, especially in simulated conversational settings.

The TLCTS approach is based upon, and builds upon, research on motivation in learning and game-based learning. The first priority, particularly for beginning learners, is to overcome motivational barriers to language learning. Many adult language learners have low self-confidence in their ability to learn a foreign language, particularly languages that are very much unlike their native language. They find traditional drill-and-practice exercises to be boring. TLCTS employs a game-based approach in order to promote and maintain learner motivation.

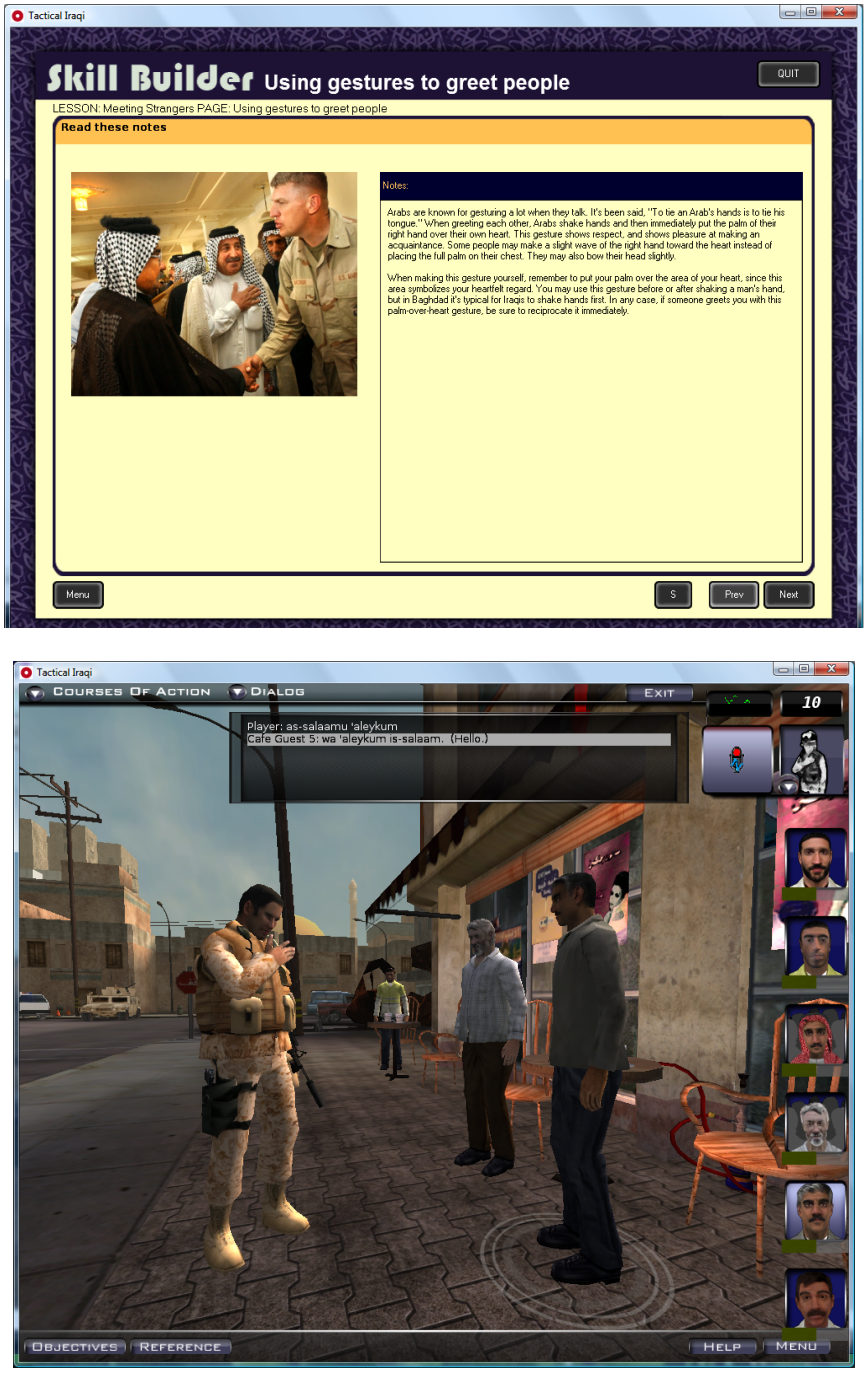

Fig. 1. The Tactical Iraqi Skill Builder and Mission Game.

As Lepper et al. (1993) have observed, motivation in learning is influenced by the 4 Cs: Curiosity, Challenge, Confidence, and Control. Good game-based learning systems optimize learner motivation via the 4 Cs (Gee, 2003). They present learners with game experiences that they are interested in and curious

about. They present a series of game activities that are challenging enough not to be boring, but not so challenging as to be frustrating and discouraging. They build learner confidence, as learners master each game level and progress to more advanced levels. The learner is in control throughout, deciding what actions to take from moment to moment during gameplay. By following these principles, TLCTS courses seek to promote and maintain learner motivation. Game scenarios emphasize situations that are similar to those learners are likely to encounter in the foreign country; this promotes learner interest in the material, and also promotes learner confidence because learners can sense that they are making rapid progress in learning to handle common situations. Learners are presented with progressively more challenging situations, which they can master in turn. And most important, they are in control of their own learning process; they can decide what actions to take in the game experiences, and even how they divide their time between structured Skill Builder activities and less structured game activities.

The design of TLCTS was also inspired by the work of Squire and Jenkins (2003) on game-based learning, but goes beyond their work in significant ways. Squire and Jenkins observed that learners who played their games were motivated to acquire background knowledge so that they could play the game better. In a similar vein, TLCTS game scenarios motivate learners to develop the communicative skills that they need to play the game scenarios. The key difference from the games that Squire and Jenkins studied is that the relevant learning materials are integrated into the learning environment, in the Skill Builder and Reference Wizard. The resulting learning environments make use of game design techniques, but are substantially more than just a game. This combination of learning activities gives learners additional options as to how they organize their learning activities, consistent with the philosophy of promoting learner control. Evaluations of TLCTS courses (e.g., [Johnson & Wu, 2008]) show that there are wide individual differences in how learners prefer to learn with TLCTS courses. Some like to work first in the Skill Builder to develop their communication skills, while others like to jump into the Mission Game, discover what skills they lack in the course of gameplay, and then focus their Skill Builder training on those skills. Either learning strategy is acceptable, as long as the learners succeed in mastering the target skills.

The Skill Builder and the game experiences both make heavy use of automated speech recognition. 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 foreign language to direct his or her character to move about a game world. In the Mission Game, the learner speaks on behalf of his character. This is taking place in the screenshot on the bottom of Figure 1, while the character performs a hand gesture that the learner had previously selected from a menu.

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 within the Skill Builder, a so-called Active Dialog. Here the learner character (at left) is engaged in conversation with a man on the street named Anwar (at right). The window at the top shows the history of the conversation up that point, in which the learner said /aani John/ (I'm John) and Anwar said /ismi anwar/ (My name is Anwar). The number in the top right corner, 20, is the learner's score up to that point (10 points for each conversational turn completed without hints). Active Dialogs are somewhat less difficult than Mission Game scenes because they are typically shorter in duration, are more focused in terms of dialog topic, and make greater use of hints. In contrast Mission Game scenes may involve conversations with multiple non-player characters on multiple topics, and in advanced modes the learner does not receive any hints unless he or she specifically requests them.

Note that even this simple dialog illustrates some variability in the use of language. The learner and Anwar each introduced themselves, but expressed themselves in slightly different ways: Anwar used the word /ismi/ (my name), while the learner used the word /aani/ (I). Handling linguistic variability is one of the key technical challenges that TLCTS had to overcome to enable learners to engage in conversations with non-player characters.

All TLCTS content is highly task-based (Doughty & Long, 2003), 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.

Fig. 2. An active dialog in Tactical Iraqi.

Games in TLCTS vary in terms of the complexity of the language tasks that they require learners to perform. Compared to the Mission Game, the Arcade Game is much lower in complexity. It focuses on

mainly on spoken directions, e.g., /ruuH yimne 'end il-beet il-ebyaZ/ (turn right at the white house). Game levels focus on listening skills or speaking skills, but not at the same time. The Arcade Game is designed to enable learners to get into a flow of gameplay (Csikszentmihalhi, 1990) without extensive knowledge of the target language. For example, in the Arcade Game level shown in Figure 3, the learner gives repeated commands to his or her avatar to collect objects quickly and gather points before the avatar is destroyed by one of the colored enemies travelling through the maze. This provides learners extensive opportunities to practice their language skills in a fast-paced mode, thereby building fluency. At the same time, the engaging, fluent gameplay provides offers learners a change of pace, and a break from the more cognitively demanding activities in the Skill Builder and Mission Game.

Fig. 3. Arcade Game example.

Complete TLCTS courses include learning materials that employ a variety of media, not just computer-based media. We provide paper-based study supplements, which include example dialogs that learners can practice with other learners, in order to get face-to-face (as opposed to face-to-computer) conversational practice. We have also developed an iPod-based supplement that contains material from the Skill Builder (Figure 4). This allows learners to continue to review their language knowledge whenever they have a free moment, and do not have access to computer. This can help both with learning and with retention.

While other language learning systems employ speech recognition technology and support simulated dialogs (Bernstein et al., 1999; DeSmedt, 1995; Hamberger, 1995; Harless et al., 1999; Holland et al., 1999), and employ AIED technology (Gamper & Knapp, 2001), 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 well over 100 hours of training, that can take learners from no knowledge of the foreign language toward an Intermediate level of spoken proficiency on the ACTFL scale (ACTFL, 1983), with a focus on job-related performance.

Fig. 4. iPod supplement to Tactical Iraqi.

ARCHITECTURE

The following is a brief summary of the current architecture of TLCTS, and of other Alelo language and culture courses. This architecture has been progressively revised and developed based on experience with evaluating and fielding earlier versions of the system. More detailed descriptions of the architecture are available in other publications (Johnson et al., 2004a; Johnson et al., 2004b; Vilhjalmsson & Samtani, 2005; Johnson & Valente, 2009).

All content in Alelo courses, including lesson and exercise pages, scene dialogs, and language resources, are specified in XML and stored in a content repository. A Web-compatible server portal named Kona[2] maintains the repository and provides shared access to the development teams (authors, artists, media editors, and programmers) working on Alelo courses. Kona is built on top of Subversion, a version control system. The content for each Alelo course is stored as a "book" in Kona, comprised of "chapters" that specify individual sections of content. Each Skill Builder lesson and Mission Game episode is specified in separate Kona chapter, as is the set of phrases used in the Arcade Game. Individual team members can check out an individual chapter (i.e., lock it for editing), edit it using our suite of authoring tools, and then check it back in. The underlying revision control framework makes it possible to roll back to earlier versions of content, either to correct errors introduced during the editing process or to reconstruct old released versions of Alelo systems.

[2] Many of the names of system components in the Alelo architecture are of Hawaiian origin. Hawaii is a cultural crossroads, and is emblematic of Alelo's mission: to promote cross-cultural communication through the use of learning technologies. A glossary of Hawaiian terms is included at the end of the article.

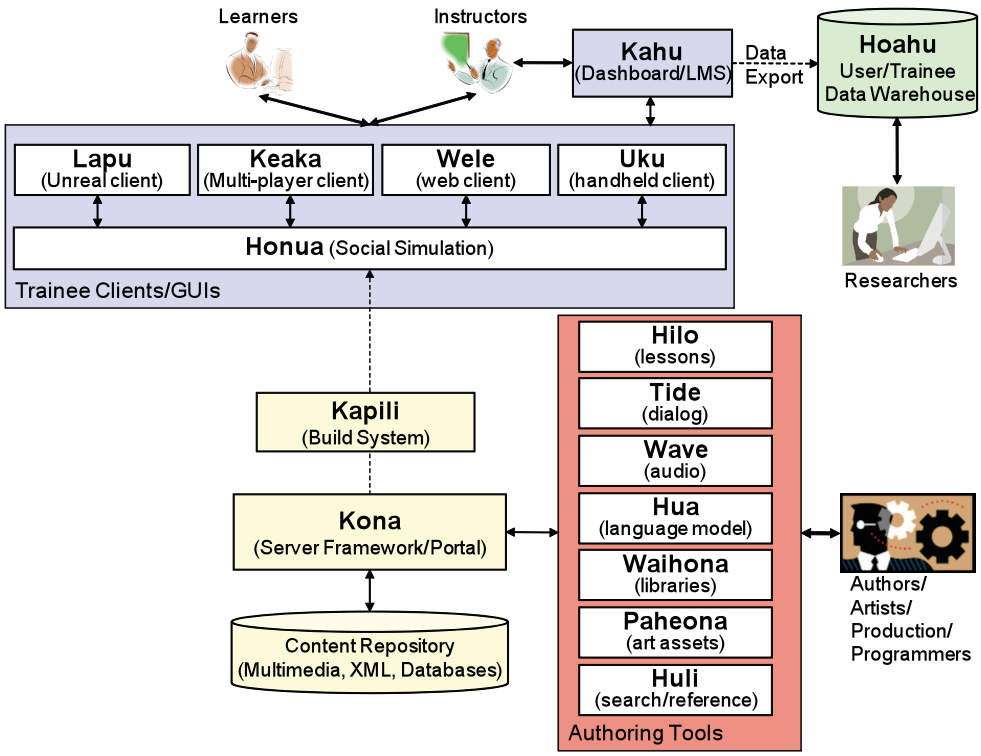

Fig. 5. The overall Alelo architecture.

The authoring process is supported by a collection of authoring tools (Meron et al., 2007; Johnson & Valente, 2008). Each focuses on editing a particular subset of the content data in the Kona repository; however they also can access other content types as needed, and provide authors with an integrated view of the underlying content. The authoring tool interfaces employ a combination of dynamically generated HTML pages, Java, and Flash. The following authoring tools are included in the tool suite: Hilo, which authors lesson and exercise pages; Tide, which authors dialog specifications; Wave, for editing and managing the many audio files that are used in creating Alelo courses; Hua, for managing the lexicons and other language resources associated with the various languages that Alelo courses teach; and Waihona and Paheona, which manage the libraries of content and art assets that Kona books refer to.

This collection of authoring tools has made it possible to create courses with large amounts of curriculum materials. For example, the current version of Tactical Iraqi includes 63 Skill Builder lessons comprising 3984 lesson and exercise pages, 16 Mission Game scenes and 44 Active Dialogs, and a vocabulary of 3194 Arabic words (where distinct inflected forms are counted as separate words).

To create Alelo learning systems, content is automatically converted to a form that can be used by the target delivery platform, and exported. One commonly used delivery platform employs the Unreal 2.5 game engine from Epic Games. All learning resources, including game levels, lessons, and reference materials, are delivered on this same platform as an integrated package. A display generator was created in Unreal that constructs one or more lesson and exercise pages from each XML page specification. Content for the Unreal Tournament game (e.g., space aliens, futuristic weapons) was removed, so that

learners are not tempted to play Unreal Tournament instead of practicing their foreign language and culture skills.

Although Unreal 2.5 is serviceable as a delivery platform, and tens of thousands of trainees have used it successfully, there has been an increasing need to support other platforms as well. Our military users want a training capability that can be integrated with other types of military training. We have therefore developed a multi-player version called Keaka that allows teams of users to conduct mission rehearsal training together, and use their language and culture skills to communicate with non-player characters that they meet, in the target foreign language. Keaka is designed to support multiple simulation-based training systems of interest to the military, such as RealWorld and Virtual Battlespace 2 (VBS2). This mission rehearsal training does not take the place of the language training that learners currently undergo using the Unreal-based Lapu client; rather, it is expected that trainees will first develop their language and culture skills using our other training delivery platforms, and then practice employing their skills in simulated military missions.

Some users want a learning capability that can be delivered via the Web, and are willing to sacrifice some interactivity and 3D graphics capability to obtain Web-based delivery. We therefore have developed a Web-based delivery client, named Wele. There is also a need for a supplemental training capability that is easily accessible when learners are not in front of a computer. We therefore have experimented with delivering content on a variety of handheld devices. These handheld devices typically have much less computing power, but the power that they do have is sufficient to enable learners to review learning materials and refresh their language skills. One platform that has drawn significant interest is the iPod. We have just completed development of an iPod version including lesson pages, dialog videos, and glossaries of words and phrases. There is even need for printed versions of dialog materials, which learners can use in dialog role-playing exercises.

A lightweight learning management system, named Kahu, is specifically designed to support the unique characteristics of Alelo learning laboratories. Users sometimes run TLCTS systems on local networks that may or may not be on the Internet; they sometimes run it on laptop computers which they may temporarily disconnect from the local network and move to other training locations. Kahu is therefore designed to support discontinuous use, unlike most learning management systems for the Web.

As learners use an Alelo environment, the system automatically records the their actions and the system's responses, as well as the learners' speech recordings. When possible we retrieve these logs and recordings, and store them in a data warehouse, Hoahu. This warehouse is an important vehicle for our adoption-based research, as it is enables us to examine how learners use these environments in natural settings. The speech recordings are annotated and used to incrementally improve the performance of the speech recognizer.

THE ROLE OF ARTIFICIAL INTELLIGENCE IN ALELO COURSES

Automated speech recognition (ASR) is a pervasive element of Alelo courses. Learners progress from simple Skill Builder lesson and exercise pages in which learners repeat and recall specific phrases, to more complex dialog exercises and game episodes where learners must engage in conversation with non-player characters. Speech recognition enables the system to give learners feedback on the quality of their speech, and makes spoken dialog possible.

The speech recognition capability in TLCTS is designed specifically to support tutorial feedback and simulated dialog, and to reflect the characteristics of learner language. This starts with the design of the

underlying acoustic models used in speech recognition. We train acoustic models on a combination of native speech and learner speech, ensuring that recognition rates are acceptably high on intelligible but mispronounced learner speech. This training proceeds iteratively. The first version of the speech recognizer is trained on a combination of native speech, typically extracted from published corpora, and samples of learner speech that we collect ourselves. Later, as TLCTS systems are introduced into use, they save recordings of each learner's speech. We collect these recordings and use them to retrain the acoustic models as needed.

Speech processing makes use of language models that are also customized to reflect the properties of learner language. First, they employ vocabulary that is limited to what is covered in the course, and in fact is limited to the sets of learning modules that the learner is currently working on. The language models are grammar-based, and are constructed from the phrases that authors enter in the Kona book. This reflects the fact that novice language learners typically make heavy use of memorized phrases in their speech (ACTFL, 1983). The curricula typically include a variety of alternative phrases for expressing particular communicative acts, often with differing degrees of politeness and formality. Thus for example a communicative act to ask someone's name might be rendered in Iraqi Arabic as /inta sh-ismak?/ (What is your [m] name?), /inti sh-ismich?/ (What is your [f] name?), /sh-ismak, ibni?/ (What's your name, son?), or even /sh-ism HaZirtak, ya seyyid?/ (What is Your Presence's name, oh sir?). Incorporating these phrases into the language model gives the speech recognizer enough flexibility to recognize language in a range of different conversational contexts.

Because the language model captures a variety of ways of expressing communicative acts, it helps detect pragmatic errors that learners commonly make. For example, in the Skill Builder lessons on greetings and introductions, learners practice addressing a variety of characters of different ages and social status, and receive tutorial feedback if their choice of wording is too blunt, impolite, or otherwise inappropriate for the conversational context. Additionally, we extend the language models for certain types of exercises to capture common errors in word choice, syntax, and morphology, enabling the system to provide targeted tutorial feedback. When authors create conversational exercises for the Skill Builder, they attempt to anticipate the most common learner mistakes, and author them into the exercise specifications. Then later these error models are revised and extended based on data collected from TLCTS users. The Skill Builder also includes focused speech pronunciation exercises in which learners practice discriminating commonly confused sounds in the target language.

Note that in the Mission Game many of these error types are not included, and learners must speak reasonably correct utterances in order to be recognized. This distinction reflects pedagogical and pragmatic concerns, as well as technical ones—it is uncommon in real life for people to correct foreigners' speech in polite conversation, and so likewise there is limited value in creating non-player characters for game scenarios that can identify a wide range of learner language errors.

The speech recognizer incorporates a "garbage model" that catches utterances that fail to adequately match any expected utterance. The configuration parameters of the garbage model can be adjusted at run time, making it possible to vary the degree of tolerance of speech disfluencies and language errors. Initial lessons typically have a higher tolerance for error, to avoid frustrating beginning learners. In advanced lessons and game levels, learners are required to achieve a relatively high level of language accuracy, often higher than what is required of second language learners in real-world interactions with native speakers. By requiring a certain degree of language accuracy from more advanced learners, we aim to help learners to overcome the language acquisition plateau that many learners encounter, where certain types of inaccuracies and disfluencies persist and are difficult to eradicate. The language acquisition

plateau is often attributed to overemphasis on communication in language teaching, at the expense of accuracy (Richards, 2008).

Alelo courses incorporate dialog models, to support robust spoken dialog with non-player characters. Conversational interaction proceeds through a three-stage process. First, the system interprets the learner's spoken utterance and selected non-verbal gesture and generates a description of the intended speech act. The non-player character then chooses an action (speech act and/or nonverbal action) to perform in response, and may change its attitude toward the player accordingly. This proceeds as follows in the example in Figure 1, where the player greets the man in the café with a polite greeting /as-salaamu 'aleykum/ and the palm-over-heart gesture indicating sincerity and respect. The system interprets the learner's action as a polite instance of the Greet communicative act, and so the man in the café counters with a polite respond-greet act, realized as the polite utterance /wa 'aleykum is-salaam/ and a palm-over-heart gesture. In addition the man's level of rapport with the learner increases, visualized in the game by a green plus floating up from the character's head (a repurposing of the icon used in Unreal Tournament to indicate when the player's health has increased).

The emphasis in TLCTS on spoken dialog differentiates it significantly from other spoken language CALL (computer aided language learning) systems. Other systems such as Auralog's TELL ME MORE and the NativeAccent pronunciation tutor (Eskenazi et al., 2007) place heavy emphasis on pronunciation error detection, while TLCTS gives priority to conversational practice, and treats pronunciation as one of many enabling skills for conversational proficiency. Earlier versions of TLCTS provided constant feedback on pronunciation errors, and and this was found to cause learners to focus too much on pronunciation and not enough on other skills such as vocabulary recall and conversational skills (Johnson & Beal, 2005); consequently we now restrict pronunciation feedback to focused pronunciation exercises. Most CALL systems that offer conversational practice, including Rosetta Stone and Virtual Conversations (Harless et al., 1999), simulate dialog in a highly structured manner, in which the system displays three or four possible conversational moves at each turn and the learner is expected to read one of the moves off of the screen. TLCTS in contrast allows a wide range of conversational moves (possible communicative acts and possible realizations of those acts), and encourages learners to figure out what to say on their own, without reading the answer off the screen. Seneff et al.'s (2007) language learning systems come closest to supporting complex conversational speech, but their approach employs stylized language tasks (e.g., spoken translation), is not tailored to learner language, and has only limited ability to recognize common language errors.

Underlying many of TLCTS's capabilities is a rich ontology of language and cultural skills. Skills center on the ability to utilize knowledge of language and culture in face-to-face communication. For example, the Tactical Iraqi lesson on meeting strangers includes the following skills, among others: use formal greetings, respond to a greeting, introduce yourself by name, understand gender roles and constraints in the area, understand the importance of eye contact, use the appropriate gesture to accompany greetings, and welcome guests. Lesson pages, exercise pages, and quiz pages are annotated according to the skills that they address. Authors use these skills to check the quality of learning materials, to ensure that exercise and quiz materials give learners adequate opportunity to practice critical skills. They also provide the foundation for a dynamic learner model. Each time the learner successfully completes a quiz or exercise, or performs an action in a dialog that requires application of a particular skill, the learner model updates its estimate of the learner's mastery of that skill, in a manner akin to knowledge tracing (Corbett & Anderson, 1995).

One use of the learner model is to provide learners with advice as to what lessons and game episodes they should undertake next (Wu, 2008). As explained above, we wish to give learners freedom to decide

what games and other learning activities to undertake when, so that they have a sense of being in control of the learning activity. At the same time, there is a risk that if they undertake activities that are too challenging for them, they might become frustrated or discouraged. A Tutor Advice Model checks the learner model before the learner attempts a game episode or Skill Builder quiz, alerts the learner if he or she appears to lack certain key prerequisite skills, and offers suggestions of Skill Builder lessons that the learner could practice to develop those skills.

INITIAL TRANSITION INTO SERIOUS USE

As described in an earlier paper (Johnson & Beal, 2005), TLCTS was developed though an iterative evaluation process, consisting of formative evaluations, limited ablation studies in controlled contexts, and exploratory classroom use. These early studies were followed by a process of transitioning into widespread use, evaluating results, and drawing lessons to be learned, which informed further system development. This process continues to this day. The following discussion summarizes lessons learned from this process, focusing on Tactical Iraqi. Tactical Iraqi is one of the most widely used TLCTS courses, and one with which there has been extensive adoption-related experience.

A complete prototype of Tactical Iraqi was completed in June 2005. By that time it had already attracted significant interest from members of the US military. Nevertheless, it was evident that 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. Ordinarily military services do not acquire new technologies unless they have independently identified a requirement for them, and even then they are unlikely to adopt a research prototype, from whatever source. Additionally, the US military already had access to a range of language learning materials, including an Army-wide license for Rosetta Stone. Tactical Iraqi required up-to-date videogame computers to run, which most military units do not have for training purposes. The US military imposes severe security restrictions on software that runs on its networks, so 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 US Marine Corps. A series of two pilot two-week training courses were administered at a Marine Corps training facility at Camp Pendleton, California, on new laptop computers provided expressly for this purpose. Each training course showed promising results, and also identified problems that were corrected 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 Unreal Tournament could not be run separately. The second pilot evaluation started showing significant promise, while still pointing to areas where further improvements were 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 were 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 and 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 in Arabic. 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. Once this customization technique was introduced into the authoring process, it began to be used in other ways as well. For example, there are now variants of Tactical Dari, Tactical Iraqi and Tactical Pashto for use by Australian Defence Force, the British Army, and foreign military forces generally. Each version has slight differences in dress, vocabulary, etc. The customization technique is also used to author content for the iPod. Although the material on the iPod is closely based on the material in the PC game, some adjustments needed to be made to the page texts to make them suitable for voiceover narration.

Although it was convenient to develop the initial prototype of the Mission Game as a game 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 Tactical Iraqi. 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 Unreal Tournament game environment.

Software testing is particularly critical for an AIED learning environment such as TLCTS. 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. This has since been integrated into the dialog authoring tool, so that authors can test and validate dialogs themselves. 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 informative in

learning environments such as TLCTS. Speech recognition performance needs to vary based on the recognition context 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 higher, 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.

The learning environment itself is just one component of the set of training materials that needed to be developed. User manuals and "train-the-trainer" courses needed to be developed. Although educators are familiar with the importance of staff development, such notions are not common in the videogame community. As a consequence many prospective users assumed that a software CD was 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 suboptimal 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 Alelo staff (mainly this author) and early adopters of the learning environment. Early adopters would sometimes develop their own 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, and often get stuck in one part of the course. Also, the programs of instruction encourage learners to repeat lessons on successive days, until they are able to demonstrate full mastery of the target skills. Without this, some learners would plough ahead through the curriculum without mastering the material, and have difficulties when they reach the more advanced material.

We have since further expanded Tactical Iraqi to include multiple alternative programs of instruction, depending upon the learner's rank and learning objectives. On one hand, some trainees have a very limited amount of time to train, and want a program of instruction that can provide them with useful skills quickly, perhaps in just a few hours. To accommodate this, separate programs of instruction are provided for basic infantrymen, squad leaders, and higher level leaders. Each has different duties to perform in country, and so the training that they receive focuses on different topics. On the other hand, some learners wish to acquire an in-depth knowledge of Arabic, so that they can pass a proficiency test at an intermediate level and receive a pay bonus. For these types of learners, a rich, in-depth curriculum is provided that helps learners develop a broader range of communication skills and apply them in a variety of contexts.

RESULTS OF INDEPENDENT EVALUATIONS

Although the evaluations described in the preceding section showed promise, and led to further small-scale use of Tactical Iraqi, additional evaluations were required in order to establish that the course was

suitable for widespread use. The following discussion focuses on three evaluations, conducted with three different subject populations. SWA Consulting either led or assisted each of these studies (Surface & Dierdorff, 2007).

These evaluations had the form and character of summative evaluations, but the term "summative" is used advisedly here. These were not evaluations of the final version of Tactical Iraqi, but rather were *in vivo* evaluations of the system in its state at the time (versions 3.0 and 3.1), and were sources of experimental results that led to further improvements both to the software architecture and to the curriculum. Each was conducted under conditions dictated by the sponsors of the respective studies, as well as the training schedule of the participating units. This makes results somewhat difficult to compare. On the other hand, these *in vivo* evaluations give a more realistic picture of effectiveness in practice, which is difficult to derive from laboratory studies.

The first evaluation was conducted with the 3rd Battalion, 7th Marines (3/7), in the fall of 2006. The 3/7 assigned two marines in each squad of approximately twelve marines to receive Tactical Iraqi training. The vast majority were junior enlisted personnel, of a rank of corporal or lance corporal. Each participating marine underwent two hours of self-paced computer-based training twice a week over a period of three months, for a total of 40 hours of training. A total of 89 3/7 Marines participated in the study.

Another independent evaluation was performed by the Defense Language Institute Foreign Language Center (DLIFLC) in Monterey, California in the spring of 2007. The DLIFLC recruited 8 military personnel who were between assignments to participate in the study. The subjects spent five days of intensive training, eight hours of computer-based training per day, following Alelo's recommended program of computer-based instruction. At the end of the program, the subjects underwent a standard oral proficiency interview (OPI). 6 of the 8 subjects demonstrated a 0+ level of proficiency on the Interagency Language Roundtable (ILR) scale, meaning that they demonstrated some ability to use rehearsed Arabic phrases in common situations. An ILR score of 0+ is equivalent to Novice or Novice High on the ACTFL scale. It should be noted that Tactical Iraqi is designed to impart particular mission-oriented and task-oriented communication skills, whereas an OPI is designed to measure general proficiency. Thus although the OPI is a standard measure of spoken communication skill, it may not give an accurate picture of how well a learner might perform on the tasks and missions that Tactical Iraqi is designed to teach. In the DLIFLC's case the OPI interviewers were instructed to ask the learners questions that had little to do with typical military tasks, such as identifying the furniture and other objects in the room where the interview took place.

The third evaluation group consisted of trainees in the military advisor training program at Ft. Riley, Kansas. One class of 268 participants took part in the program. They completed twelve hours of classroom instruction, together with 28 hours of training with Tactical Iraqi, in intensive day-long self-paced sessions. After the overall training program was complete, the trainees took part in a final mission rehearsal exercise (MRX), in which the trainees engaged in live missions in a mockup Iraqi town, populated by Iraqi role players. Roughly one third of the trainees felt that they were effective in their Arabic speaking and listening skills during the MRX exercise.

All three groups showed significant increases ($p < .01$) in knowledge of Iraqi Arabic language and culture, as measured by an independently constructed post-test focusing on the vocabulary and skills taught in Tactical Iraqi. The DLIFLC group had the largest increase in knowledge. This group spent more time training with Tactical Iraqi than the groups participating in the blended learning condition. The DLIFLC also followed our recommended program of instruction most closely, and were regularly supervised. The other units received varying degrees of supervision. Although game-based learning offers

the promise of motivating learners to gain knowledge on their own, in these studies at least the greatest learning gains were achieved by learners who used Tactical Iraqi under supervised conditions.

Training with Tactical Iraqi also affected the trainees' confidence in their ability to understand and speak Arabic. Subjects in the Ft. Riley group and the DLIFLC group both reported significant increases in speaking self-efficacy and listening self-efficacy ($p < .01$). Comparable data from the Marine group are not available. Subjects in all groups reported slightly increased motivation to transfer (i.e., motivation to apply their Arabic language and culture skills) and increased motivation to participate in future training opportunities.

On the other hand, the Ft. Riley learners reported a slight decrease in learning self-efficacy overall. This is typical of many training programs for difficult subjects, where learners may start the training program with unrealistic expectations of the how easy the subject matter will be to learn.

Subjects were asked to evaluate the utility of each component of Tactical Iraqi: the Skill Builder, the Arcade Game, and the Mission Game. All groups of subjects rated all components of the Tactical Iraqi slightly positively. On average, the subjects rated the Skill Builder most highly in terms of utility, although there was also the highest variance in responses concerning the Skill Builder.

We suspect that there were may have been multiple reasons why the Skill Builder was judged more highly. At the time, the content in the Tactical Iraqi Mission Game was focused a civil reconstruction scenario, while the Skill Builder covered a wider range of missions such as house-to-house searches. More recent versions of Tactical Iraqi contain a much wider of range of content pertaining to a wider range of missions, both in the Skill Builder and in the Mission Game.

Also, many of the subjects in the Ft. Riley and marine groups received very little in the way of orientation to the software. In our recommended program of instruction we encourage trainees to alternate between the Skill Builder and the game components, both as a way to get a wider variety of practice opportunities and as a change of pace. Our analysis of the log data from Ft. Riley in particular suggest that many trainees did not do this, and in fact spent very little time at all in the Mission Game. The DLIFLC group was encouraged to make greater use of the Mission Game and Arcade Game, and their ratings of those components were relatively higher, although still lower than their ratings of the Skill Builder.

Some trainees reported problems with usability and technical glitches, particularly with the speech recognizer. This was to be expected, as this version of Tactical Iraqi was still undergoing transition from research prototype to product-quality learning system. Speech recognition for learner language remains a challenging problem. Further improvements has since been made to system reliability generally and speech recognition specifically.

In late 2007 the 3/7 Marines returned from their tours of duty in Iraq. The tour of duty of the 3/7 was particularly successful in that it did not suffer a single combat-related injury or fatality during their deployment, which was in Anbar province. This same battalion had been involved in heavy fighting in Ramadi in its previous deployment.

To understand the role of Tactical Iraqi in the 3/7's success, the Marine Corps Center for Lessons Learned conducted a study of the effect of Tactical Iraqi training on the performance of the 3/7 (MCCLL, 2008). Returning Marines in these units were asked to complete a questionnaire regarding their use of Tactical Iraqi, their experience using Iraqi Arabic while deployed, their assessment of their own Arabic communication skills, and their recommendations regarding the use of Tactical Iraqi for training. Meanwhile, the officers in charge of the 3/7 were interviewed regarding their satisfaction with the training that Tactical Iraqi provided, and its effect on operations during deployment.

In the assessment of the officers of the 3/7, knowledge of Arabic contributed to the unit's operational capability in theater, and this in turn contributed to the unit's effectiveness. The battalion had a limited

number of translators assigned to it, and the Tactical Iraqi training enabled marines to interact with Iraqis when translators were unavailable. In the words of one company commander: "I can tell you that it did greatly enhance our operational capability... It just increased our tempo, it increased our understanding, it increased most importantly our relationship with them. They understood that we came in with a basic set of knowledge and that we were also willing to learn and be able to communicate and those communications skills came through."

Surveys of the marines were consistent with the officers' assessment. 76% reported had opportunity to speak Arabic on a daily basis, and only 15% of those surveyed reported that they were rarely or never effective in listening and understanding Iraqi Arabic. 50% of the marines who trained with Tactical Iraqi reported that the skills that they acquired affected the success of their mission. They described a number of types of situations in which Tactical Iraqi training came into play: patrols, small talk over meals with the Iraqis, working with Iraqi army and police, directing traffic, and asking questions of visitors at checkpoints. In the words of one marine: "[Tactical Iraqi] impacted my success as well as the 3/7. Iraqis saw that we were different from other Marines. I still have friends from Iraq calling me on the cell phone to keep in touch."

The success of the 3/7 is no doubt the result of a number of factors, not least of which was the strong commitment and support of the unit's commander. Nevertheless, the experience of the 3/7 strongly suggests that Tactical Iraqi achieved level 4 outcomes on the Kirkpatrick scale (Kirkpatrick, 1994): it yielded results in terms of performance on the job. The Marine Corps Center for Lessons Learned study was reported widely within the Marine Corps, and since then increasing numbers of marine units have employed TLCTS in their preparations for deployment to Iraq and Afghanistan.

In view of this adoption experience and evaluation results, let us know return to the research questions raised in the introduction.

- **Is the learning environment effective?** The experience of the 3/7 Marines and other units deploying overseas indicates that it is effective. However, it is important to define effectiveness consistent with the learning goals of the people adopting the environment. The 3/7 assessed effectiveness in terms of how the training helped them perform their mission, and from this perspective Tactical Iraqi performed well. The DLIFLC assessed effectiveness in part using a general oral proficiency interview, and by this measure Tactical Iraqi did less well. However the DLIFLC did not consider whether the skills tested in the OPI were relevant to communication needs in country.
- **Does it enhance motivation, and encourage learners to continue learning?** There is some evidence in support of this. The subjects participating in the evaluation studies reported increased motivation to participate in future training opportunities. Some of the 3/7 learners continued training with Tactical Iraqi after they had deployed to Iraq. This is noteworthy both because it was outside of scheduled training, and because the marines had a range of learning options available to them, including immersion in the local Iraqi Arabic linguistic community. Still, we need to collect more data of learners engaged in voluntary training with TLCTS, as opposed to mandated programs of instruction.
- **Does the game-based learning approach contribute to the overall learning effect?** So far, evidence of this incomplete. We know that most learners divide their time between the Skill Builder and game components of TLCTS. We also know that the game activities offer learners a welcome change of pace, when they are engaged in intensive training. At the same time, the learners in the earlier studies rated the Skill Builder higher than the game activities. We need to evaluate the newer

versions of TLCTS courses, which incorporate improved, mission-relevant game activities, to see whether learner attitudes change.

STATUS AND FUTURE WORK

The Alelo learning environment architecture, which is used to develop TLCTS courses, is one of the most widely used AIED platforms in the world today. Tens of thousands of learners have already used TLCTS courses. One training site at Ft. Riley, Kansas is alone responsible for training approximately 10,000 trainees per year. Alelo maintains a Web site for download access to software, with over 10,000 registered users. The total number of users of TLCTS systems is difficult to determine precisely, but certainly numbers in the tens of thousands.

A number of US military posts and training have set up computer labs for training, in the United States, Europe, Japan, and in the Middle East. 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 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.

We have since released a new version of Tactical Iraqi, 5.0, with a greatly expanded range of Mission Game scenes and Skill Builder lessons, dealing with a much wider range of missions. New lessons are included focusing on common problems in Arabic pronunciation, to help learners get a better grounding in the sound system of Arabic. Tactical Iraqi 5.0 now supports two types of curricula: predeployment training and proficiency training. Trainees who choose proficiency training receive more in-depth training, and practice their Arabic skills in a wider variety of contexts, so that they can prepare to take and pass a spoken proficiency test at an intermediate level. We plan a series of evaluations to determine whether learners are in fact able to achieve this level of proficiency.

Supporting higher levels of language proficiency has necessitated extensions and improvements to the natural language processing and authoring technologies used in TLCTS. As learners gain in proficiency they are more likely to use their vocabulary to construct novel phrases, instead of relying on memorized phrases. As a result the recognition rates achieved by TLCTS's phrase-based speech recognizer starts to drop on higher proficiency learner language. We are currently extending our authoring tools and language processing capabilities so support a wider range of variability in learner language. Our new dialog authoring tools enable to authors to construct patterns of possible utterances from example utterances, thus expanding the range of possible inputs. We are developing a new agent modeling tool, named LightPsych, which enables authors to specify a wider range of possible character responses to learner inputs.

We have made continuous efforts to improve all components of Tactical Iraqi as well as the underlying Alelo architecture. We strive to increase the interactivity of the Skill Builder lessons, and give learners greater opportunity to practice their communication skills in realistic situations. This is resulting in increasing use of interactive conversational characters in the Skill Builder. We have also added scoring mechanisms to the interactive dialogs in the Skill Builder, to make them more game-like. We are currently exploring further ways of employing game design techniques to motivate learners, including on-line collaboration and competition between learners.

Meanwhile, Alelo technology is being applied to new types of language and culture courses. We have developed a Web-based Virtual Cultural Awareness Trainer (VCAT) that focuses primarily on

cultural competence, in contrast to the Tactical Language courses that focus primarily on linguistic competence. We are also developing new educational courses based on the TLCTS model. Alelo has also developed Web-based learning environment for English as a second language, which is integrated into Voice of America's news portal. These efforts have enabled tens of thousands of learners worldwide to start taking advantage of this new approach to language learning, and will no doubt introduce a new set of adoption-based research questions.

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 anonymous reviewers made many constructive suggestions for improving the paper. 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 TRASYS. The opinions expressed in the paper are those of the authors' and do not reflect those of the sponsors or the US Government.

GLOSSARY OF HAWAIIAN TERMS

Kona	big island	Honua	world
Hilo	a Polynesian navigator	Keaka	theater
Hua (hua'olelo)	word	Wele (puna welewele)	(spider) web
Waihona	library	Uku	flea
Lapu	ghost	Paheona	art
Huli	search	Hoahu	warehouse
Kahu	administrator		

REFERENCES

American Council for the Teaching of Foreign Languages (ACTFL) (1983). *ACTFL Proficiency Guidelines*. Hastings-on-Hudson, NY: ACTFL Materials Center.

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

Corbett, A., & Anderson, J.R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253-278.

Csikszentmihalhi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper Perennial.

DeSmedt, W.J. (1995). 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* (pp. 153-174). Mahwah, NJ: Lawrence Erlbaum.

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

Eskenazi, M., Kennedy, A., Ketchum, C., Olszewski, R., & Pelton, G. (2007). The NativeAccent™ Pronunciation Tutor: Measuring success in the real world. In M. Eskenazi (Ed.) *Proceedings of SlaTE Workshop on Speech and Language Technology for Education*. ISCA Tutorial and Research Workshop, Farmington, PA USA.

Gamper, G. & Knapp, J. (2001). A review of CALL systems in foreign language instruction. In J.D. Moore et al. (Eds.) *Artificial Intelligence in Education* (pp. 377-388). Amsterdam: IOS Press.

Gee, J.P. (2003). *What Video Games Have to Teach Us about Learning and Literacy*. New York: Palgrave Macmillan.

H. Hamberger, H. (1995). 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* (pp. 183-199). Mahwah, NH: Lawrence Erlbaum.

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

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

Johnson, W.L. & C. Beal, C. (2005). Iterative evaluation of a large-scale, intelligent game for language learning, In C.-K. Looi et al. (Eds.), *Artificial Intelligence in Education* (pp. 290-297). Amsterdam: IOS Press.

Johnson, W.L., Beal, C., Fowles-Winkler, A., Lauper, U., Marsella, S., Narayanan, S., Papachristou, D. & Vilhjalmsson, H. (2004a). Tactical Language Training System: An Interim Report. In J. Lester et al. (Eds.), ITS 2004, LNCS 3220 (pp. 336-354). Berlin: Springer-Verlag.

Johnson, W.L., Marsella, S. & Vilhjálmsson, H. (2004b). The DARWARS Tactical Language Training System. In Proceedings of I/ITSEC 2004. Arlington, VA: NTSA.

Johnson, W.L. & Valente, A. (2008). Collaborative authoring of serious games for language and culture. In *Proceedings of SimTecT 2008*. Lindfield, NSW: SIAA.

Johnson, W.L. & Valente, A. (2009). Tactical Language and Culture Training Systems: Using AI to teach foreign languages and cultures. *AI Magazine* 30(2), 72-84.

Johnson, W.L., Vilhjalmsson, H., & Marsella, S. (2005). Serious games for language learning: How much game, how much AI? In C.-K. Looi et al. (Eds.), *Artificial Intelligence in Education* (pp. 306-313). Amsterdam: IOS Press.

Johnson, W.L. & Wu, S. (2008). Assessing aptitude for learning with a serious game for foreign language and culture. In B.P. Woolf et al. (Eds.), *Intelligent Tutoring Systems 9th Intl. Conf, ITS 2008*, LNCS 5091 (pp. 520-529). Berlin: Springer-Verlag.

Kirkpatrick, D.L. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J., 1993. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as Cognitive Tools* (pp. 75-105). Hillsdale, NJ: Erlbaum.

MCCLL (Marine Corps Center for Lessons Learned) (2008). Tactical Iraqi Language and Culture Training System. MCCLL Newsletter 4 (8), August 2008.

Meron, J., Valente, A., & Johnson, W.L. (2007). Improving the authoring of foreign language interactive lessons in the Tactical Language Training System. In M. Eskenazi (Ed.) *Proceedings of SlaTE Workshop on Speech and Language Technology for Education*. ISCA Tutorial and Research Workshop, Farmington, PA USA.

Richards, J.C. (2008). *Moving Beyond the Plateau: From Intermediate to Advanced Levels in Language Learning*. Cambridge: Cambridge University Press.

Squire, K. & Jenkins, H. (2003). Harnessing the power of games in education. *Insight*, http://www.iaete.org/insight.

Surface, E. & Dierdorff E. (2007). *Special Operations Language Training Software Measurement of Effectiveness Study: Tactical Iraqi Study Final Report.* Special Operations Forces Language Office, Tampa, FL, USA.

Vilhjalmsson, H. & Samtani, P. (2005). MissionEngine : Multi-system integration using Python in the Tactical Language Project. In *Proceedings of PyCon 2005.* Python Software Fondation.

Wang, C., & Seneff, S. (2007). A spoken translation game for second language learning. In R. Luckin et al. (Eds.), Artificial Intelligence in Education (pp. 315-322). Amsterdam: IOS Press.

Wu, S. (2008). *Reducing unproductive learning activities in serious games for second language acquisition*. Ph.D. Thesis, Univ. of Southern California.