To appear in: Durlach & Lesgold (Eds.), Adaptive Technologies for Training and Education. New York: Cambridge University Press.

The Art and Science of Developing Intercultural Competence

W. Lewis Johnson and LeeEllen Friedland
Alelo Inc.
12910 Culver Bl., Suite J
Los Angeles, CA 90066 USA

Aaron Watson and Eric Surface
SWA Consulting Inc.
801 Jones Franklin Rd
Suite 270
Raleigh, NC 27606-3381

Introduction

As the world becomes more interconnected, it is increasingly common for people to interact and work with others who come from different cultural backgrounds and who speak different native languages. These intercultural interactions require *intercultural competence*: the ability to communicate successfully with people of other cultures (Byram, 1997). Intercultural competence is, therefore, an increasingly important skill for professionals in a range of fields, including management, education, health, and emergency response (Earley, 1987; Schneider et al., 2009; Kosoko-Lasaki et al., 2008; USDHHS, 2008). Intercultural competence is also receiving increasing attention within the military services, as service members take part in overseas deployments and engage in missions that involve frequent contact with people of other cultures (McDonald et al., 2008).

Providing such adult professionals with adequate intercultural competence training can be a difficult challenge. Many believe that cultural proficiency requires extensive education and years of study (NFLC, 2007). Working professionals who are not specialists in linguistics, foreign languages, or cultural analysis typically do not have the time to study foreign languages and cultures extensively. They may have studied a foreign language in school, but by the time they are called upon to use their foreign language skills on the job, those skills have often gone dormant or been forgotten.

This chapter describes a computer-based approach to intercultural competency training that helps learners quickly develop and retain the intercultural skills they need to be effective in their work. This approach is realized in the Alelo family of language and culture training products. It focuses on the particular responsibilities that each learner is preparing for, the situations the learner is likely to encounter in carrying out those responsibilities, and the particular cross-cultural skills that apply to those situations. Learners get the opportunity to practice their skills in computer simulations of realistic work situations and get direct feedback about their levels of

effectiveness in those situations. These simulations incorporate artificially intelligent animated characters that represent members of the target culture, and that can engage in spoken conversations with learners in the foreign language. We then provide learners with sustainment training tools that help them to retain and refresh their communication skills, even after long gaps in training and periods of disuse.

The design of the curricula, and the methods used to implement it, are particularly intended to support and motivate lifelong learners. Unlike students in typical classroom settings, we do not assume that learners are a captive audience. Many are studying in their free time, and have many other demands on that time. If they find the training boring, or feel they are making insufficient progress, they may give up and discontinue learning. The training employs a serious game approach (Barrett & Johnson, 2010), utilizing techniques drawn from interactive games to engage learners and maintain their interest and motivation.

This work is informed by, and contributes to, the art and science of cultural competency training, at multiple levels. The training courses advance the state of the art in language and culture training by providing simulations of cross-cultural interactions at an unprecedented level of fidelity. The serious game approach incorporates the work of graphic artists, animators, and other artistic professionals. We conduct scientific analyses of how learners interact with our courses in the field, which helps us determine whether our training methods are effective and prioritize efforts to improve them.

This approach has been realized in computer-based training courses that have been used by tens of thousands of learners around the world to learn to communication with people of other cultures, mostly in foreign languages. Military service members in particular have benefited greatly from this training, which helps them establish good rapport with local people, increase operational effectiveness, and ultimately avoid conflict and combat casualties.

The chapter is organized as follows. The next section presents some examples from Alelo courses that illustrate the training methods and technologies employed. After this comes an overview of the instructional design methods that are employed, and a description of some of the particular adaptive technologies that are used to realize them, particularly those for simulated dialog and skill sustainment training. It describes how data collected from field use of Alelo courses are analyzed to measure technology performance and learning effectiveness, and to inform further development effort.

Examples from Alelo Courses

The following examples from Alelo's Operational Dari and Operational Pashto courses and the earlier Tactical Dari course will serve to illustrate the key points in

this chapter. These courses are designed to help military service members prepare for deployments to Afghanistan, by acquiring the communication skills that they will need in their encounters with the Afghan people. As shown in Figure 1, trainees can practice their skills in encounters with Afghan non-player characters. In this example the learner plays an American small unit leader (left) who is meeting with the malek (tribal leader) and other elders in the village to discuss collaborating on a reconstruction project. The learner can select gestures for his character to perform, and speaks in the foreign language on behalf of his character, in this case, in Dari. The system recognizes and interprets what the learner says, and the artificially intelligent non-player characters in the scene respond accordingly. The transcript of the conversation is shown on the top of the screen. In this example, the learner has greeted and introduced himself, and the malek has responded to the greeting and introduced himself. Learners are evaluated in terms of their ability to complete the objectives of the mission, and scored on their use of communication skills. The communication score (top right corner of the figure) depends upon the number of conversational turns they are able to complete, and the amount of scaffolding (i.e., hints and assistance; see Seeley Brown et al., 1989) that they must rely upon in order to complete them. In this example, the learner is relying on transcriptions and translations of the malek's phrases to get through the exercise, and so receives a lower score than would be the case if these scaffolds were turned off.

Figure 1. Meeting with the malek in Operational Dari.

The objective of scenarios such as this is not simply to practice the foreign language, but to develop intercultural competence. To succeed, the learner needs to act and

behave in a manner that is culturally appropriate in this situation. This includes such things as making polite and appropriate inquiries about the malek's family, showing reciprocity and mutual respect, and establishing a good relationship and rapport with the malek. The malek character responds positively to each instance of culturally appropriate behavior, and his attitude toward the learner gradually improves. The malek will agree to collaborate on the project with the learner only if the learner has established a sufficient level of rapport. Such scenarios give learners opportunities to practice their skills and help them develop familiarity with these sorts of cross-cultural situations, so that they will be confident and more at ease when they deal with similar situations in real life.

Each course contains a number of such simulated episodes, in what is known as the Mission Game component of the course. Some episodes provide useful practice for anyone working in the foreign country, while others, such as the episode shown in Figure 1, are more aimed at people with particular mission responsibilities. Before starting the course each trainee completes a questionnaire in which they indicate their particular mission responsibilities and learning goals. The system then automatically generates a tailored course of instruction, of lesson modules and game episodes. These enable the trainee to quickly and efficiently acquire the particular communication skills that he or she requires.

The course incorporates a combination of lesson materials and learning activities that help learners acquire the knowledge, skills, and attitudes necessary to master these scenarios. These materials can be delivered in a coordinated fashion on a combination of desktop and mobile devices. Figure 2 is an example of an exercise delivered on a mobile device, in this case, an iPod Touch. Here the learner practices one of the basic greetings in Dari, "sa-laam aa-ley-kum." The learner can hear a native Dari speaker say the phrase, speak the phrase into the microphone, and compare their own speech against that of the native speaker. The desktop trainer has a corresponding exercise, that integrates automated speech recognition, to verify that the learner's speech is intelligible and is at an acceptable level of pronunciation accuracy for a beginning language learner.

Figure 2. A language instruction exercise on an iPod Touch

Once the learner has acquired some basic phrases, he can proceed to simple conversational exercises. Figure 3 shows one such exercise, called an Active Dialog, from the desktop training system. Here the learner practices selected communicative skills—in this case basic greetings and introductions—with his Afghan counterpart. The setting for the dialog is a virtual classroom, suggesting that this is a practice activity rather than a higher-stakes simulated mission. Otherwise, the learner's objectives in the activity are similar to those in the Mission Game episodes. In this case, the learner needs to greet his counterpart and exchange names with him, and in the process he or she receives a communication score based upon the number of conversational turns completed and the amount of scaffolding the learner relies upon to complete them.

Figure 3. An Active Dialog from the Operational Dari course

The system tracks the learner's activities and progress on both the desktop trainer and on the mobile trainer. This enables it to dynamically construct refresher training lessons to recover knowledge, skills, and attitudes that may be vulnerable to forgetting and decay (Figure 4). Refresher material is selected depending upon how recently it has been trained, with spaced practice intervals that progressively increase over time.

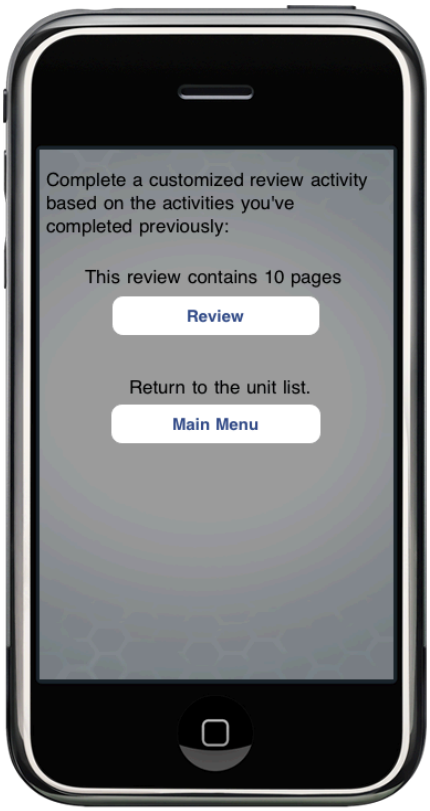

Figure 4. Dynamic refresher training lessons

To facilitate transfer of communication skills to mission operations in Alelo's military courses, a Virtual Role Player (VRP) capability has been developed that integrates conversational virtual humans into virtual training and mission rehearsal environments. VRP technology integrates with multiple commonly used mission rehearsal systems, including Virtual Battlespace 2 (VBS2) and RealWorld. VRPs extend the capabilities of these training systems so that trainees can perform the same military actions as before (e.g., maneuver vehicles, fire weapons), but can also engage in conversations with the characters in the scenario, just as in the Mission Game shown in Figure 1. Multiple trainees can participate in the virtual environment at any one time, making it possible for teams to carry out simulated missions together.

Figure 5. VRP mission rehearsal scenario

Course Development Methodology and Design Principles

One of the keys to the effectiveness of Alelo courses for lifelong learners is their *focus*: they are designed to concentrate on the skills that learners need to carry out their jobs or missions. In this respect they are similar to courses that teach foreign languages for special purposes (LSP) (Fiorito, 2005) in that they are driven by a needs analysis that determines what skills are most needed by the learners. However, in contrast with LSP courses, they take into consideration the *breadth* of intercultural knowledge, skills, and attitudes (KSAs) required for such purposes (McDonald et al. 2008), not just the language skills. They are designed to promote *transfer* of skills from the learning environment to the real world, e.g., through simulated scenarios that allow learners to practice their skills in realistic contexts. Other important considerations for the design of these courses are *accuracy* and *consistency* of the learning material. The courses must reflect current linguistic usage and cultural practices relevant to the intended purposes of the communication skills. Consistency is an issue because the courses incorporate multiple learning activities on multiple learning platforms, which need to treat the intercultural KSAs in a coordinated fashion. It is a particular problem for colloquial

and less commonly taught languages, for which there are often no agreed-upon standards of form or orthography.

To ensure that each course meets these objectives, we have developed and implemented a courseware development methodology, called the Situated Culture Methodology, that creates a holistic approach to using the situational context of the learner's job or mission to generate learning objectives, uses those learning objectives to identify key cross-cultural and culture-specific factors and competencies, then applies them in a tailored curriculum focused on the intercultural competence needed to successfully perform the relevant job or mission.

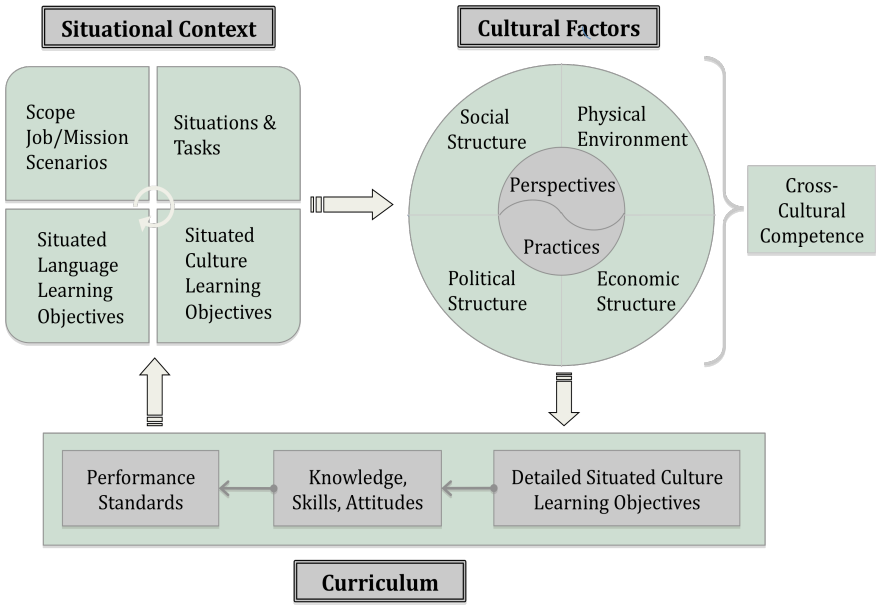

Figure 6. Situated Culture Methodology.

Principles of instructional design guide the creation of innovative curricular components, as illustrated in the examples found in the section above. There is also a scientific basis of theory and method from sociocultural and linguistic anthropology (Hymes 1974, 1987; Hall 1966; Birdwhistell, 1970) that shapes the underlying research and data development required to ensure the fidelity of course content in relation to learning objectives.

Ethnographic interviews are conducted with subject matter experts (SMEs) in order to collect data about jobs and missions (Bernard, 1998). This typically includes information about work processes and tasks, as well as the types of contexts in which the work takes place, interactions that commonly occur as part of performing the work tasks, how social and cultural dynamics affect the way one might perform the work in different settings, and criteria for successful performance of tasks. Of prime interest are first-person narratives about how intercultural communication played a role in real-life mission encounters; videos of these narratives are often

included in the course to illustrate examples of how principles can apply to specific scenarios.

If language instruction is part of a given course, native speakers of the target language are interviewed in depth, often repeatedly, to collect realistic examples of spoken dialog. This is sometimes supplemented by engaging the native speakers in role-playing in order to prompt them to provide variations and alternative approaches to specific scenario dialogs. These linguistic data are then transcribed and used in creating mission-based lessons and exercises. Usually one or more language SMEs are also used to review course materials and sometimes advise on specific issues pertaining to speech acts or interactions in the curriculum.

Because Alelo takes a situated approach to both language and culture, all curriculum materials are presented in an integrated way that includes information and guidance regarding contextual factors to consider, culturally appropriate practices and behaviors, and the potential consequences of inappropriate behavior. This instruction and the simulated scenarios include a naturalistic combination of factors that are at play during real-world interactions, including verbal and nonverbal behavior, and the use of social space (e.g., distance between people interacting face-to-face).

Job and mission tasks—based on the ethnographic data collected, and the role that learners will play in the simulated environment (e.g., senior military officer or junior enlisted service member)—are also analyzed in relation to social and cultural factors relevant to the geographic region in which the course is based, as well as sociopolitical, socioeconomic, sociodemographic, and sociocultural characteristics of the place and its people. This analysis is then used to shape the design of lessons, exercises, and interactive scenarios in the course to ensure that appropriate examples, values, and performance criteria are used. Information about the social and cultural factors of the specific region or cultural group is also gathered primarily through ethnographic interviews with natives of the region or cultural group, supplemented by academic research and analysis

In Figure 6, the circle showing cultural factors is labeled with five categories based on a recent framework for operational culture developed by the US Marine Corps (Salmoni & Holmes Eber, 2008). It should be noted that, while these categories are useful, the Situated Culture Methodology is not dependent on any one categorization or paradigm of social or cultural factors. The methodology can be used with any conceptualization of social and cultural factors based in social theory. Any paradigm or categorization that is chosen would be represented in the circle shape within the figure and relevant values (e.g., labels, categories, subheadings) would then be used to structure the presentation of information in the course.

Ten key design principles in Alelo courses

The Situated Culture Methodology provides a proven means of collecting and developing ethnographic linguistic and sociocultural data to use as primary course materials that prepare learners for the roles they will play in real intercultural mission-based situations. The Alelo development methodology for course design includes the following ten key principles:

1. **Task-based learning.** The SCM approach is well aligned with task-based approaches to language learning (Ellis, 2003), in that it focuses on communication necessary to perform real-world tasks, and in that it has learners practice their communication skills in the context of realistic tasks. It assumes that intercultural competence curricula, like other types of curricula, should be based upon a task analysis of the skills to be taught, and the work situations in which those skills are to be applied (Jonassen et al., 1999). The learning activities in Alelo courses are all task-oriented in this sense. The task-based approach applies to cultural skills training, in that the courses focus on situated cultural action.
2. **Language in culture.** Alelo products are developed with the understanding that language and culture are inextricably linked and cannot be separated. Communicative skills are taught in the context of face-to-face communication in social situations, in which learners are alerted to the ways in which power relations and social distance influence the use of language in those situations (Brown & Levinson, 1987). The courses cover nonverbal aspects of communication, such as gestures, as well as the verbal aspects. Courses survey the range of cultural factors that are likely to be relevant in the real-world situations that learners are likely to encounter in other countries.
3. **Communicative approach.** Alelo curricula emphasize interaction as the ultimate learning goal, as well as the means to achieve this goal. Curricula are organized around communicative functions in context, and seek to promote communicative competencies in those contexts (Canale & Swain, 1980). They incorporate many activities that involve interaction, particularly with animated characters. Curricula teach individual communicative functions and then progress to tasks and dialogs that involve combinations of these functions.
4. **Tailored curricula.** Alelo courses are tailored to the individual, and this customization takes place at multiple levels. For example, we have developed customized versions of our courses for military and nonmilitary use, and further customize by military organization (e.g., US vs. NATO allies) and branch of military organization (e.g., Army vs. Navy). Each may have differences in terms of situations and missions, and hence differences in curricula. Learning systems further customize the curriculum according to the needs of the individual learner, depending on their military rank, their training objective (proficiency vs. survival), desired training missions, and time available for training.
5. **Localization.** Alelo courses are localized according to the linguistic and other cultural practices of the intended learning community. This helps to ensure

that learners regard the courses as something written with them in mind. This includes images, audio recordings, cultural content, and translation into the native language of the learners, but extends beyond that to the norms of usage in particular communities. This comes up particularly in courses for military services, each of which tends to develop slightly different practices in the use of titles, forms of address, and specialized jargon. It also relates to choice of artwork in the animated scenarios, e.g., uniforms and physical appearance of the animated characters.

6. **Immersive learning.** Immersive environments play an important role in providing the context for communication and facilitate skill transfer to real-world situations, in accordance with theories of situated cognition (Lave, 1988) and constructivist learning (Vygotsky, 1986). Immersive environments also provide the context for game-based learning where learners develop the contextualized skills and confidence they need to succeed.
7. **Scaffolding.** Alelo employs an elaborate scaffolding methodology within our courses (page-level, lesson-level, and course-level) to progressively develop the skills of the learner and reduce their reliance on hints, assistance, and other forms of support. Courses employ a scaffolding-and-fading approach, in which scaffolding is gradually removed until learners are able to complete the task unassisted (Seeley Brown et al., 1989). Scaffolding takes the form of translations and transcriptions, hints, and other signals that the learner's action was consistent or inconsistent with cultural practice. It is also integrated into the feedback from pedagogical agents (virtual coaches and virtual conversational partners). Virtual coaches help learners prepare for practice scenarios and give learners feedback on their performance. The reactions of the conversational partners are highlighted to help learners recognize when they have acted in a culturally inappropriate way. Learners first develop their skills in practice scenarios where scaffolding is added, and then test their skills in assessment scenarios where the scaffolding is removed.
8. **Feedback.** The frequency, timing, and type of feedback given to a learner is critical during language instruction. In our courses, feedback takes the form of evaluation and scoring of the learner's performance, reactions of the animated characters to culturally correct or incorrect behavior, or feedback from virtual coaches. Because many activities are highly interactive and game-based, particular attention is paid to ensure that feedback is provided in a way that does not disrupt the flow of the learning activity. This helps to get learners into a flow state of engaged activity (Csikszentmihalyi, 1990) and helps promote conversational fluency. We therefore employ a two-loop model in providing feedback. Feedback in the outer loop, between learning activities, encourages reflection on the preceding learning activity and drawing lessons learned. Feedback in the inner loop, in the midst of learning activity, is action-oriented, rapid, and designed to maintain the flow of the activity. Other researchers such as VanLehn (2006) employ a two-loop model

in providing feedback, but do not place as much emphasis on maintaining flow in the inner loop.

9. **Learner motivation.** Alelo courses are designed throughout to promote and maintain learner motivation. Motivation is a common problem in second language courses, which many learners find tedious and boring. It is a particular issue in courses for lifelong learners, who are frequently attempting to learn a language in their spare time. Such learners are likely to lose interest and discontinue learning if the learning activities are not sufficiently engaging and motivating. Research has shown some foreign language learners who lack confidence in their language skills are more likely to disengage from the training environment compared to those with greater confidence (Bienkowski, Watson, & Surface, 2010). Following the work of researchers on motivation in learning (e.g., Lepper, 1988), learning activities are designed to optimize learner confidence and sense of control, and optimize the level of challenge. This is achieved in part through the use of feedback that offers encouragement and adheres to norms of politeness (Wang et al., 2008). The principles and design techniques of game-based learning (Gee, 2003; Prensky, 2001) are employed throughout to promote sustained engagement.
10. **Skill sustainment.** Courses are designed not just to develop communication skills, but to sustain them as well. This is a particular concern for courses that attempt to impart skills in a compressed period of time. The contextualization of learning helps promote sustainment, since it supports the formation of associations in memory that both promote retention and facilitate recall (Neisser, 1984). In accordance with research on spaced practice and memory (Baddely & Longmand, 1978; Bahrick, 1979; Melton, 1970), we have developed methods for dynamically selecting refresher learning materials based upon the learner's past training history, to help the learner to retain their communication skills better and quickly recover them when they start to lose them. We provide sustainment training on mobile devices, so that learners can refresh their skills anywhere, anytime.

Technology Architecture Overview

The following is an overview of the technology architecture used to author and deliver these courses. The remainder of the chapter will focus on selected adaptive technologies within this architecture. More detail about the other technologies described here may be found in (Johnson & Valente, 2009; Johnson, 2010).

Authors develop courses using a set of collaborative authoring tools named Kona, and accessed using a Web-based portal named Hilo.[1] Kona maintains a repository of

[1] A number of the component names in this architecture, as well as the company name Alelo, have Hawaiian origins. Hawaii provides an example of the type of

course content, including specifications of learning activities and a collection of art assets (images, sound recordings, animations, etc.) used in these learning activities. A multidisciplinary team of authors utilize these tools, including linguists, anthropologists, media specialists, animators, software engineers, and experts in the language and culture subject matter of the particular course.

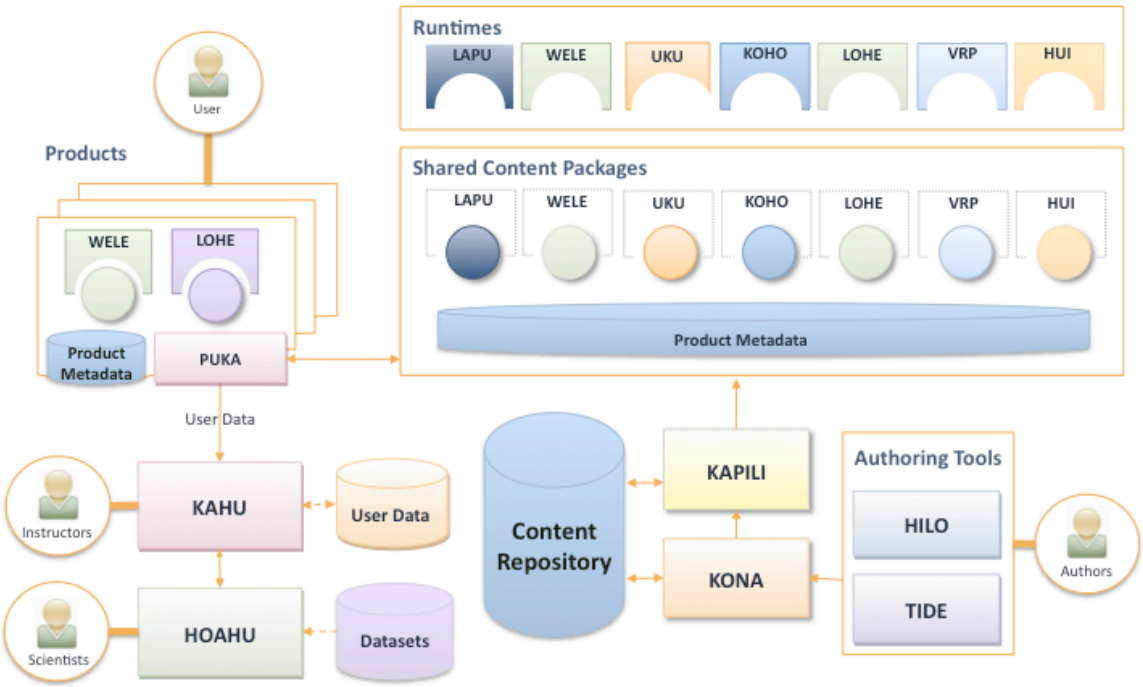

Figure 6. Alelo technology architecture

Course content specifications are represented in XML, and specify the learning content in a platform-independent manner. This makes it possible to deliver learning materials on both desktop computers and mobile devices, and create integrated suites of learning materials across multiple platforms. Each content element is tagged with metadata to indicate whether it applies to all delivery platforms or is intended for a specific platform such as a mobile device. In a typical course most content applies to all platforms, but it is sometimes necessary to tag selected elements for a particular device to reflect the specifics of how the course is delivered on that device. For example, our iPhone application does not yet include a Mission Game, so references to the Mission Game in the course materials must be marked as desktop-only.

---

harmonious interchange between cultures that Alelo as a company seeks to promote.

The content specifications can support multiple source languages, which helps us to achieve the Localization goal mentioned in the previous section. For example, Alelo has developed a multilingual Web site for Voice of America to teach American English language and culture, named goEnglish.me. This course has so far been developed for Chinese and Farsi speakers, and localizations for speakers of other languages are planned for the near future.

The content specifications also include metadata to specify the intended audience of a particular learning activity. For example, the Operational Dari lesson on introducing your team has different versions of content for US Marines, other US service members, and other NATO allies operating in Afghanistan, each using different names for military ranks. US Marine learners learn how to say Marine-specific military ranks such as gunnery sergeant or lance corporal in Dari, while the other services do not. This enables us to employ a mass customization approach to content development, and plays an essential role in achieving the Tailored Curricula goal mentioned in the previous section.

To create the individual learning products, a build system named Kapili converts content into shared content packages that can be loaded onto particular delivery platforms. Each delivery platform has a runtime player that can play the content on that particular device. For example, the scenarios in Figure 1 and the dialog in Figure 3 were delivered using the Lapu desktop runtime platform, built on top of the Unreal Engine from Epic Games. The language instruction activity in Figure 2 and the refresher training activity in Figure 4 were delivered using the Uku mobile runtime platform on the Apple iOS operating system for iPhones, iPod Touches, and iPads.

Each learning product typically uses some combination of the available runtime platforms, depending upon the needs of the particular learning organization and the computing resources available to them. A front-end portal for the learning package, named Puka, enables learners to select the particular learning modules they require, based upon their particular learning goals and needs. Puka can then download the content packages for those particular learning modules, providing each learner with a customized course of instruction.

Finally, an important element of the technology architecture is the support for data collection and analysis. We have developed a common framework for logging learner activities, influenced by Mostow and Beck's (2009) model of educational data logging. Each runtime player includes a logging mechanism consistent with this framework. These data can be uploaded to a learning management system so that instructors can track learners' progress. We also can import the logging data into a data warehouse named Hoahu for detailed analysis (Johnson et al., 2010). This enables us to investigate how the learners are using the systems, study trajectories of learning over time, and derive measures of learning gains and system performance. This makes possible an iterative cycle of course development, evaluation, and improvement, where lessons learned from deployed versions of

learning systems help to inform the development of successive versions. Data collected from training sites is incorporated into the training corpora for some of the underlying technologies such as automated speech recognition, resulting in progressive improvement of recognition performance over time.

Adaptive Curriculum Tailoring

As indicated above, most Alelo curricula are designed to support a range of different learning requirements, with the understanding that not all learners have the same learning needs. Some of this variation in learning requirements is captured by classifying learners into different groups, e.g., different military services, and tagging learning content according to the group that it is intended to support. This results in a family of learning products for a given topic, each tailored for a particular group of learners. Adaptive curriculum tailoring methods are then employed to further tailor the curriculum according to the learning goals of smaller groups of learners as well as individual learners.

The particular set of learning modules in the curriculum for a given course is determined by the course of instruction (COI). Courses of instruction can be selected and changed at run time, according to choices made by the instructor and/or the learner. Using the learning management system (Kahu in Figure 6), an instructor can select a course of instruction for his or her class or small unit, and download it to the learning portal (Puka) on each learner's training device. Then when the learner uses the training device, he or she will then see a tailored course curriculum, separated into required or recommended learning modules (included in the COI) and optional learning modules (not included in the COI). The COI is shared across the available training devices. So for example if the course includes both a desktop runtime player and a mobile runtime player, learners will see the same COI on both runtime players.

COIs can be further tailored according to the needs and preferences of each individual learner, and in the case of self-paced learning this is the primary tailoring mechanism. At the beginning of the course learners can complete a questionnaire, in which they indicate their particular role in their organization, the range of jobs or missions that they are training for, the particular region they are interested in, and in some courses, the level of communicative proficiency they are aiming for. Puka then selects a recommended course of instruction for the learner, based upon those choices. If during the course learners decide to change their learning objectives, they can go back and change their questionnaire answers and obtain a revised COI. The latter feature was introduced based upon feedback from Marine trainees who had trained with the Tactical Iraqi language and culture trainer. Some of these trainees were originally assigned to complete a compressed predeployment training course, but as their language skills increased their training ambitions also increased, and they started getting interested in preparing for a language proficiency test so that they could obtain college credit and bonus pay. The latest versions of the Iraqi

Arabic and Sub-Saharan French courses now allow the trainee to expand their learning objectives at any time, and switch from a predeployment training COI to a language proficiency training COI.

The following examples, taken from the Virtual Cultural Awareness Trainer for the Horn of Africa (VCAT-HOA), further illustrate how courses of instruction are tailored based on learner questionnaire choices. VCAT is intended for use by military service members and other government personnel on assignment to the Horn of Africa region.

- Learners can choose from among thirteen countries of interest within the Horn of Africa region. The COI will then include a combination of modules pertaining to the Horn of Africa region as a whole and modules specific to the country of choice.
- Learners can select a level of responsibility: Junior, Mid-Level and Senior. Junior enlisted personnel fall into the junior category, non-commissioned officers fall into the mid-level category, and officers fall into the senior category. The COI will then include learning modules relating to the particular types of situations that learners at that level are likely to encounter, e.g., senior personnel complete modules covering the cultural issues relating to meetings with host country leaders.
- Learners can select a type of mission to train for: civil affairs, security operations, humanitarian assistance operations, and noncombatant evacuation operations. Based on this choice, training modules relating to those particular types of missions are included in the COI.

Based on these choices, each learner can receive one of 156 possible tailored courses of instruction.

Conversational Virtual Humans as Adaptive Learning Technologies

Adaptive technologies are also integrated into the individual learning activities within a course of instruction. These include the speech and multimodal communication technologies that enable us to create conversational virtual humans that engage in face-to-face dialog with learners, as illustrated in Figures 1 and 3. These play a critical role in providing learners with opportunities to practice their culturally appropriate communication skills. In order for their behavior to be culturally appropriate, it is important that they adapt to the learner. The virtual humans are designed to model the way people adapt their interaction with other people based upon the changing power and social distance relationships between them. If learners demonstrate knowledge of culturally appropriate behavior and develop rapport with the animated characters, the characters will reciprocate and become more friendly and cooperative. This makes it possible to model intercultural interaction scenarios such as the meeting with the malek in Operational Dari, where appropriate cultural action is essential for a successful mission outcome. In addition, virtual humans play an essential role in providing feedback to the learner. The

manner in which virtual humans react to what learners say and do provides learners with cues as to whether or not they are behaving in a socially appropriate way, as in real life. These reactions can be highly salient learning experiences, particularly if the virtual human reacts emotionally to what the learner has said.

The virtual humans can engage in simulated multimodal conversation with language learners. The virtual humans need to be able to react to nonverbal actions such as gestures that the learners might perform, actions that they might perform on objects in the virtual world, as well as spoken utterances that they might say. In most Alelo products the user selects gestures and actions from on-screen menus, which then causes the learner's avatar in the scenario to perform the action. Spoken utterances are input via microphone, except in products that lack a spoken input channel, in which case learners choose utterances from menus of options. The conversational virtual human then interprets the inputs in a manner consistent with the culture, and then generates a response.

Unlike many spoken language processing systems that are designed for use by native speakers, Alelo spoken dialog systems are specifically designed to process and understand *learner language*, i.e., language forms produced by language learners (Ellis & Barkhuizer, 2005). All components of the spoken dialog system architecture, from automated speech recognition through virtual human response, are designed with the characteristics of learner language in mind, particularly for novice-to-intermediate language learners, as defined on the ACTFL spoken proficiency scale (ACTFL, 1983). This is necessary to ensure that the dialog system is robust enough to interpret the learner's speech in spite of mispronunciations and other errors. Another objective is to detect and categorize errors when they occur, so that the system can assess learner performance and provide feedback.

Interpreting learner language can be challenging for people (ACTFL, 1983), and even more so for computers. However by taking into account the particular characteristics of the speech of the target learners, it is possible to maximize understanding rates. The acoustic models used in the automated speech recognizer are trained on samples of learner speech as well as native speech, to ensure that the recognizer is tolerant of learner accents. Analysis of interaction logs and speech recordings collected from the courses make it possible to understand in detail the properties of the learners' language as they progress through the course. The following are some key properties that the framework takes advantage of:

- Limited vocabulary, primarily drawn from the courses themselves and therefore somewhat predictable;
- Frequent use of memorized phrases, drawn from the curriculum, which makes the language even more predictable, and
- Limited use of complex, multi-phrase utterances, so that the meaning of most utterances can be characterized as simple speech acts.

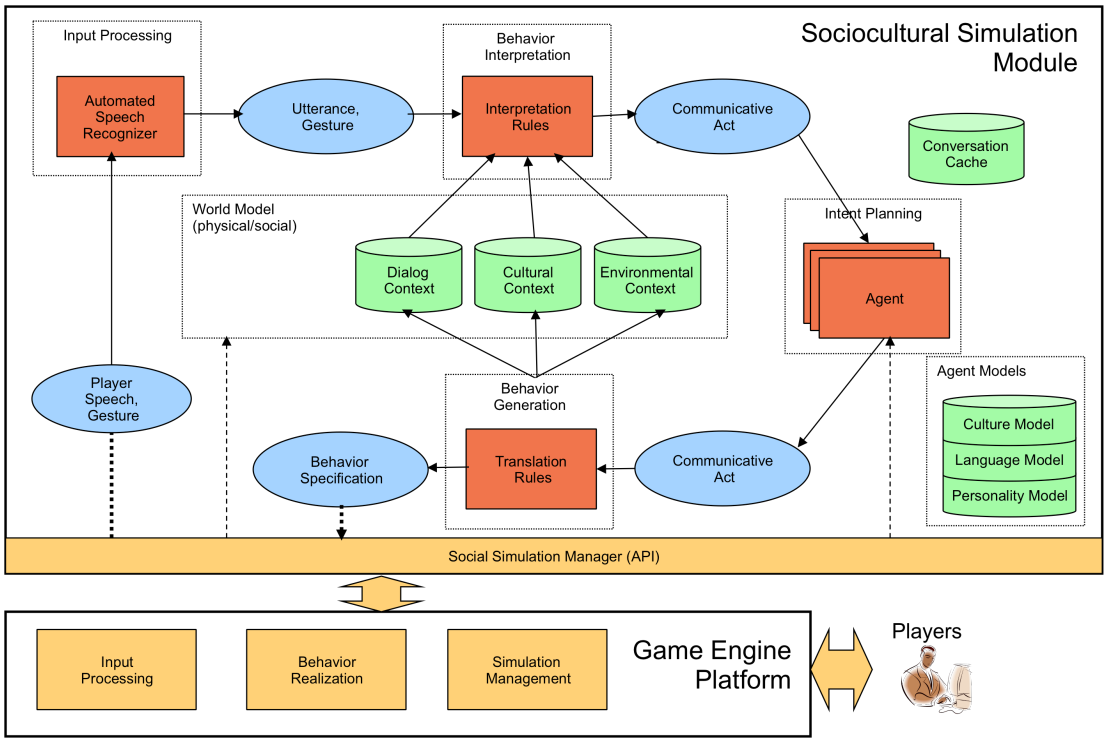

Figure 7. Conversational virtual human architecture

Figure 7 gives an overview of the conversational virtual human architecture. The architecture is built on top of game engine platforms that manage the virtual worlds in which the simulations take place, including the physical environment, objects, and animated characters. It is independent of any particular game engine, but relies upon the game engine interface to provide the conversational interface that receives the learners' inputs (speech, gestures and other action choices, and locomotion) outputs the virtual humans' spoken responses, and animates the movements of the animated characters and objects in the scene.

The automated speech recognizer takes the learner's speech and attempts to recognize it as an utterance in the foreign language. It utilizes the Julius open-source speech recognition decoder (Lee & Kawahara, 2009), applied to recognition models developed specifically for learner language. It utilizes a grammar-based recognition approach, in which the recognition grammars are generated automatically from the database of foreign language utterances within Kona that is constructed by the course authors as they develop each course. Depending upon the type of learning material, the grammar may recognize ill-formed utterances in addition to well-formed utterances, enabling the system to recognize and respond to learner mistakes. In addition to the recognition model there is a "garbage model" that will match with low probability against any input utterance. This helps to ensure that the speech recognizer will report an utterance as recognized if it is recognized with sufficiently high probability. This helps to reject unintelligible learner speech, in which cases the virtual human generates responses such as "Excuse me?" or "I didn't understand you" in the foreign language. It also helps to avoid situations where the

system misinterprets the learner as saying something that he or she didn't say, which learners find frustrating and lowers their confidence in the reliability of the recognition system.

Next, the virtual human takes the learner's behavior, i.e., the learner's utterance, gestures, and other actions, and attempts to interpret the communicative intent of that behavior. Communicative intents are presented as communicative acts, e.g., like speech acts (Searle, 1975), but not limited to speech. Following the work of Traum and Hinkelman (1992), each includes a *core function* (to greet, ask, inform, request, etc.), and *grounding* function, i.e., the act's role in coordinating the conversation (e.g., to initiate a conversation, continue, acknowledge the other speaker, etc.). Interpretation depends upon the particular culture being modeled, as well as the dialog context. If the learner says something that does not make sense in the current dialog context the virtual human may respond as if he or she does not understand the learner, just as people do in similar circumstances.

Once the system interprets the learner's behavior, the virtual humans in the conversation or within earshot in the vicinity choose a communicative act to perform in response, in the intent planning phase. The communicative act is then realized as a coordinated set of behaviors on the part of the virtual human, which may comprise a spoken utterance, one or more gesture animations, lip animation, and other animated actions. These depend upon the particular culture, the dialog state, states of objects in the virtual world, as well as the virtual human's current attitudes toward the learner. This results in a high level of adaptivity and flexibility in the virtual human behavior.

Depending upon the type of learning activity, portions of the overall architecture may be used. In the simplest case, called *language-instruction pages*, just the automated speech recognition component is used. Learners are prompted to say the phrase, and receive feedback as to whether the learner's utterance matches the phrase. This type of exercise is illustrated in Figure 2. (Note that automated speech recognition is currently available only in the desktop runtime platforms, not the iPhone, so on the iPhone the platform plays back the learner's speech and the native speaker's speech, and lets the learner compare the two.)

In another common type of activity, called *mini-dialogs*, learners practice individual conversational turns, and get feedback on the accuracy and appropriateness of their response, frequently provided by a Virtual Coach. Learners build up their skills by practicing conversational turns in mini-dialogs, and then combine them in more extended dialogs. For the mini-dialogs, only the automated speech recognition and behavior interpretation phases of the pipeline are required. Because the context of the exercise is more constrained in an open-ended dialog, it is possible to expand the range of ill-formed or inappropriate utterances that the system is prepared to recognize, and give learners feedback regarding those errors when they occur (Kumar et al., 2009). Thus the bulk of the coached practice on language forms and language use occurs in the context of these mini-dialogs.

Evaluating virtual human performance

Getting the behavior right for the virtual humans requires collaboration among social scientists, instructional designers, animators, engineers, in fact, pretty much the entire multidisciplinary team. The front-end analysis of the culture results in libraries of culturally appropriate behaviors, which in turn are created and sequenced by trained animators. The resulting virtual human behavior must be appropriate from a social science standpoint (e.g., are the malek character's gestures appropriate for an Afghan of that social standing in that situation?), from an artistic standpoint (e.g., are they fluid and natural?), as well as from a pedagogical standpoint (e.g., will a novice-level trainee be able to recognize and interpret the cultural cues in the malek's behavior?). Scaffolding is sometimes added to the virtual human's behavior, in the form of visual and auditory signals and callout commentaries, to help novice learners with limited intercultural competence to interpret the situation and the behaviors of the virtual humans correctly.

To make accurate assessments of the performance of the spoken dialog system, it is necessary to collect and analyze data from actual dialogs between learners and virtual humans, in the context of field use of the systems. Field data is essential for an accurate assessment because speech recognition rates depend upon the accuracy of the learners' speech, and learners continually adapt their speech as a result of their interactions with the system.

As an illustration, the following are some evaluation results from field data collected in late 2009 (Johnson et al., 2010; Sagae et al., in press). Forty-five active-duty Navy subjects were recruited as volunteers to undergo self-study training with Tactical Iraqi and Tactical French. Subjects were asked to train for extended periods of time over the course of several months. The purpose of this was to obtain longitudinal data on learner and system performance, to see how the performance of the dialog system changes as learners develop and improve their language skills. As it turned out, many of the subjects were unable to take significant amounts of time out of their busy training schedule for this volunteer study. A total of 13 subjects trained for four or more hours, 8 out of 25 Arabic trainees and 5 out of 20 French trainees. Analysis focused on these subjects.

One useful overall measure of dialog system performance is the *meaningful response rate*, i.e., the frequency of virtual human responses that indicated that the virtual human understood what the learner said, and did not produce responses like "Sorry, I didn't understand you." The virtual human may fail to have a meaningful response for a number of reasons: the microphone may have failed to pick up the learner's voice, the automated speech recognizer may have been unable to recognize the learner's utterance, or the virtual human may have been unable to interpret the meaning of the utterance in the context of the dialog.

If the virtual human did not understand what the learner said, the learner would need to repeat or rephrase what he or she said, i.e., perform additional *utterance*

attempts, until the virtual human understood and was able to respond and complete the dialog turn. So another useful measure of dialog system performance is the number of utterance attempts per dialog turn, or the number of utterance attempts per meaningful response. If the learner must repeat themselves multiple times in order to be understood, a *dialog breakdown* is said to have occurred. Dialog breakdowns inevitably occur in real life when language learners attempt to communicate with native speakers. However, severe breakdowns with multiple speech attempts can cause learner frustration, and should be avoided if possible, e.g., by hints that help learners to rephrase what they are trying to say in order to be understood.

In this particular study, the meaningful response rate for the French learners was 59%, and the meaningful response rate for the Arabic learners was 58%. The number of utterance attempts per dialog turn was 1.64 for French, and 1.73 for Arabic. Of the total number of dialog terms, 7.1% of the French dialog turns were severe breakdowns involving four or more dialog attempts, while 6.5% of the Arabic dialog turns were severe breakdowns.

These statistics indicate that learners had to repeat themselves from time to time, but usually were able to make themselves understood with relatively few attempts. These frequencies are likely not greatly different from what beginning learners would experience in real life when engaging in conversations with native speakers. However a small fraction of the dialog turns were severe, and brought down the overall meaningful response rate. Further study of these breakdown occurrences revealed that even when the learners received hints, in the form of possible foreign language utterance texts for the learner to say, they continued to experience difficulty being understood, especially if the utterances were in French. The learners apparently were confused by the written French hints, which is not surprising since French spelling is somewhat idiosyncratic and includes letters that are pronounced differently from how they are pronounced in English. As an outcome of this analysis, the hint system in the new Operational courses has been improved to provide more spoken hints, to help learners improve their pronunciation and overcome serious breakdowns.

Although these measures provide useful overall indicators of dialog performance, they do not distinguish failures due to system performance problems from failures resulting from poor learner speech. Therefore two annotators were recruited to review learner recordings out of the sample set, and judge the accuracy of the system's interpretations of the French learners' speech. Two sets of dialogs were selected for comparison in this study: one set of beginner-level dialogs and one set of intermediate-level dialogs. Annotators were asked first to indicate whether they agreed with the system as to whether the learner's speech was intelligible or not (i.e., recognizable utterance vs. garbage), and whether they agreed with the system's interpretation of the utterance as a communicative act. A total of 345 dialog turns were annotated. Overall inter-annotator agreement for this task, as measured by Cohen's Kappa, was 0.8, which is good.

Overall, there were very few instances where the annotators judged that the system had misunderstood the learners by inappropriately assigning a communicative act to the learner's speech. Less than 1% of the dialog turns were considered unintelligible by the annotators but interpreted by the system. For 3.5% of the dialog turns (8 for annotator 1, 16 for annotator 2), the system assigned a communicative act interpretation and the annotators assigned a different interpretation. In the remaining cases where the system assigned an interpretation to the dialog turn (167 or 48.4% for annotator 1, 160 or 46.4% for annotator 2), the annotators agreed with the system's interpretations.

The annotators were able to assign interpretations to a number of utterances that the system rejected as garbage (94 or 27.2% for annotator 1, 134 or 38.8% for annotator 2). To understand the causes for these disagreements, the annotators were asked to revisit these cases and judge whether the utterance had a pronunciation error, a grammatical error, or was free of errors. Most of these cases (58 out of 94 or 62% for annotator 1, 85 out of 134 or 63% for annotator 2) were judged as having pronunciation errors, and very few (0 for annotator 1, 2 for annotator 2) were judged as having grammatical errors. Remaining cases (36 or 10.4% of the total for annotator 1, 47 or 13.6% for annotator 2) were judged as well-formed, and therefore the system should have interpreted them, assuming they were appropriate to the context of the dialog.

These results indicate that the Alelo dialog systems investigated in this study are quite successful at avoiding misunderstandings, where the system thinks the learner said something that he or she did not mean to say. This is a significant advance over the speech recognizer used in the first version of Tactical Iraqi, which lacked a garbage model and therefore tended to misunderstand learner speech frequently, causing considerable learner frustration.

The results also suggest that dialog performance could be further improved by increasing the system's tolerance for mispronunciations—or by providing improved training so that learners improve their pronunciation. For the reasons noted above, the current focus of development at Alelo is on improving pronunciation hints. Improving tolerance for mispronunciations in automated speech recognition may be useful as well; however, it likely is not necessary to achieve the same level of tolerance for error as the annotators in this study. For one thing, it may be useful for pedagogical purposes to require learners to strive for greater pronunciation accuracy, to avoid the plateau and fossilization effects that many language learners experience (Han & Odlin, 2006). The results also indicate that the current system rejects some well-formed utterances, and so further effort may be needed to reduce this rejection rate—as long as it does not increase misunderstanding rates as a side effect.

Adaptive Technologies for Sustainment Training

The next set of technologies that we will discuss relate to the problem of adapting curricula to meet the needs of skill sustainment and refresher training. As we attempt to meet the needs of lifelong learners to develop communication skills quickly, there is risk that they might also lose them quickly, if they have not trained them sufficiently. We are employing adaptive training in two ways to help address the skill sustainment problem.

- The sequence of training activities presented to the learner is adapted dynamically until the learner has mastered the skills being trained, so that the learned skills are resistant to forgetting and skill decay.
- After long gaps in training, learners receive compressed and tailored curricula that help them quickly regain the skills that they may have lost.

Adaptive technologies have the potential for revolutionizing refresher training for language and other communicative skills. Language learners typically experience skill decay after periods of disuse of their language skills, but most language curricula are aimed at initial acquisition of language skills, not refresher training. It is difficult to design a good refresher language training course that meets the needs of all learners, since the effectiveness of the course depends on the knowledge, skills, and attitudes that each individual learner brings to the course—what they have previously learned as well as what skills have since decayed. However, through adaptive technologies it is possible to customize curricula based upon each learner's skill profile and training history, so that learners can recover their skills as quickly and efficiently as possible.

This is a good example of where scientific analysis of field use of Alelo's learning environments has led to their further improvements. By analyzing the pattern of skill acquisition and decay in learners who use our courses for extended periods, with gaps in their training, it was possible to design improved adaptive training methods that resist and counteract skill decay, which are now being integrated into the new versions of Alelo's training systems.

Background: Research in skill acquisition and decay

The research literature demonstrates skill decay over time. Research has also identified learning conditions and principles that can reduce decay (i.e., increase retention) across a variety of skills. For instance, spaced practice can help lessen the effects of skill decay (e.g., Baddely & Longmand, 1978; Bahrick, 1979; Melton, 1970). While massed practice may enhance performance in the training context, research suggests such a strategy often does not translate to enhanced retention after periods of time without using the skill (e.g., Bahrick, 1979; Estes, 1955). Spaced practice, however, tends to result in slower, incremental improvements in performance during training (relative to massed), but higher levels of retention after some time interval (e.g., Baddely & Longmand, 1978; Bahrick, 1979; Melton, 1970). This is the basis for many flashcard-based language learning algorithms that promote rote memorization of vocabulary items.

But, learning communicative language skill is more complex and involved than rote memorization of declarative knowledge. From a general skill acquisition perspective (e.g., Anderson, 1982), learners must acquire knowledge, proceduralize that knowledge creating initial skill (Neves & Anderson, 1981), compile the skill, automatize the skill, and apply the skill. Furthermore, communicative competence involves a combination of a number of component skills, including pronunciation skills, skill in producing well-formed utterances, listening comprehension skills, ability to recognize cultural cues, etc. Skill decay can therefore potentially manifest itself in multiple ways. Learners may retain the ability to recognize vocabulary items, but lose fluency. They may retain the ability to pronounce foreign language phrases, but lose the ability to use those skills effectively in conversation.

Research points to various strategies that can be employed to alleviate decay in complex cognitive skills, and which may apply to communicative skills. Variable practice (i.e., practicing a set of skills across a range of potential situations) has been found to be beneficial across many skill domains (see Ghodsian et al., 1997), such as extending knowledge and skills acquired to novel situations (see Van Rossum, 1990 for review) and a variety of motor learning tasks (Catalano & Kleiner, 1984; Kerr & Booth, 1978). Furthermore, retention can be enhanced by constructing practice environment(s) to match real-world contexts in which learners will use their skills. Such training approaches will call for trainees to process information and produce skills under the same conditions they will encounter after leaving training (Morris, Bransford, & Franks, 1977). Encoding specificity theory (Cormier, 1987), which states information is better recalled under conditions similar to those in which it was originally processed and encoded, suggests learning and practicing skills in an environment similar to the transfer context will enhance retention.

Overlearning, or the act of continuing to train and practice a skill after it has been learned, has been proposed to lead to skill automaticity (i.e., requiring fewer attentional demands to retrieve and perform) (Shiffrin & Schneider, 1977; Schneider, 1985). Research evidence suggests overlearning has a moderate effect on retention of both cognitive and physical tasks (see Driskell, Willis, & Cooper, 1992). However, the evidence also suggests this effect weakens as the retention interval lengthens (Driskell, Willis, & Cooper, 1992).

Approaches such as variable practice and overlearning can be thought of as adaptive in nature, in that they require training content to adjust (or adapt) according to learner behaviors. For example, one cannot begin to *overlearn* a skill until he or she has first sufficiently demonstrated the skill has been learned. The major challenge in implementing such adaptive approaches is customizing them to fit the unique (yet volatile) skill profiles of individual learners. These skill profiles can change both during periods of training and during periods of disuse of the learned skills.

Empirical study of skill decay in Tactical Iraqi

To the extent possible, Alelo curricula are designed with these findings in mind, to maximize retention. Learners overlearn key phrases such as greetings, to maximize retention, and practice vocabulary in a range of phrases and contexts, to provide variable practice. However, the goal of providing effective training in a compressed amount of time makes it necessary to find the right mix of learning activities and to make tradeoffs between these various activities. The research literature does not provide clear indications of how to make these tradeoffs. For example, how much training time should be devoted to overlearning of key phrases, and how much should be devoted to variable practice in a range of phrases?

To address these questions, Alelo collaborated with SWA Consulting Inc., to analyze archival data from trainees who worked for extended periods with Alelo's Tactical Iraqi language and culture trainer. The subject trainees were US Marines who had trained at the Marine Corps Air Ground Combat Center (MCAGCC) at Twentynine Palms, California.

We retrieved a total of 294 separate trainee profiles from Twentynine Palms. Out of these, 34 had trained for at least a month and had skipped at least a week at some point during their training. Many of these trainees had conducted extensive training with Tactical Iraqi, at least 25 Skill Builder lessons. These learners were native English speakers who trained with Tactical Iraqi between October 2007 and September 2008. The final data set consisted of 9,615 unique speech attempts made during the course of training.

This study focused on patterns of successful speech attempts (or utterances) produced by learners. We identified skill decay as instances in which learners failed to reproduce an utterance they previously had produced correctly. Several potential predictors of utterance success in the practice setting of the Tactical Language Training System (TLTS) Skill Builder were identified and operationalized for this empirical study. The follow variables were calculated from log data for this study:

- The number of previous speech attempts, both in general and for specific target utterances, was calculated.
- The number of previous successful utterances was also calculated for each target utterance. This variable could reflect the degree of overlearning undertaken by the learner, with a larger number of successful attempts indicating greater overlearning.
- The time elapsed (in real time) since a learner's last successful utterance for each target utterance was computed. This variable reflects a retention interval–the amount of time that passed since the learner last produced the target utterance successfully.
- Training pages that both provide example speech clips and require learners to repeat those same utterances were differentiated from pages that require learners to produce speech from memory. The number of other target utterances attempted between two attempts of a specific target utterance

was calculated. To the extent that the other utterances were semantically similar to the target utterance, this variable may reflect interference.

To examine the role of training context characteristics and learner behaviors on language acquisition and attrition, hierarchical linear modeling (HLM) for a binary outcome was employed (see Raudenbush & Byrk, 2002). The models estimated are similar to common logistic regression models, with the exception that model intercepts were allowed to vary across groups, or clusters, of related observations. Analyses were carried out using the GLIMMIX procedure in SAS for generalized linear mixed models. The three-level multilevel analysis considered speech attempts nested within target utterances nested within speakers, with the probability of success (vs. failure) of each speech attempt constituting the predicted outcome of the model. Only speech attempts for which learners had at least one successful previous attempt for that specific target utterance (i.e., had demonstrated the ability to produce the utterance correctly) were included in the analysis. As is customary in multilevel modeling, a series of models was estimated to determine a) the observed unconditional variability (i.e., not controlling for any explanatory variables) in the probability of successful speech attempts across target utterances (after controlling for between-speaker differences), and b) the predictive roles of various explanatory variables entered into the model.

Across the 9,615 speech attempts, there was significant variability in the probability of speech success between target utterances and within speakers. The following are some of the factors that were related to successful or unsuccessful attempts at speech production. The amount of time since one's last correct attempt (on a given target) was negatively related to the probability of success on any given attempt. The impact of time lessened as the time interval increased, with the largest decrease occurring during the first 50 hours, followed by a more gradual decrease up until 150 hours (at which point time since the last correct attempt was unrelated to future success). This relationship is presented in Figure 8, which also depicts the role of prior successes with respect to the target utterance. This plot shows that skill decay due to elapsed time was more pronounced in learners who had fewer prior correct utterances than those who had more correct utterances. That is, advanced learners' skills were somewhat less susceptible to decay compared to those of novice learners.

The complexity (or difficulty) of the target phrase was also related to skill decay. The probability of a correct response varied by the degree of difficulty of the phrase, both in terms of utterance length and the frequency of difficult phonemes. As utterance length increased, the probability of a correct response reached a maximum at a length of 21 characters (approximately the length of common phrases such as "as-salaamu 'aleykum" and "wa 'aleykum as-salaam"), and then decreased significantly thereafter (see Figure 9). This may reflect the tendency of novice learners to memorize and repeat short phrases. Also, the proportion of non-English phonemes in a target utterance was negatively related to the probability of success.

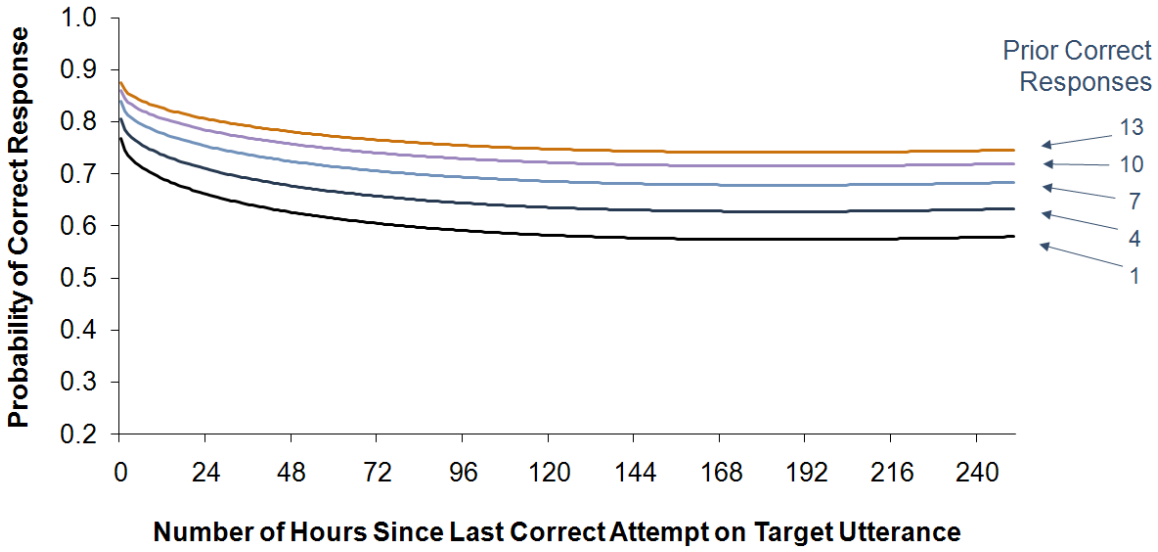

Figure 8. Time Elapsed Since Last Correct Attempt and Probability of a Correct Response[2]

These findings show promise for incorporating an automated process for tracking which language skills are most susceptible to decay for individual learners at a given time during training.

Adaptive presentation of learning activities for sustainment

Based upon the lessons learned from this and similar analyses, we are developing new methods for dynamically customizing each learner's course of instruction and sequence of learning activities, based upon the learner's trajectory of learning and predicted forgetting. The first version, illustrated in Figure 4, integrated into the Alelo's mobile training platform, prompts learners to review language instruction pages that the learner has previously studied, to ensure that the learner has properly mastered these phrases. It keeps track of each language instruction page that the trainee has visited, and the amount of time that has elapsed since it was visited. If a page has not been visited for sufficient amount of time, it is included in a set of pages for the trainee to review. Each time the learner reviews a page, the time interval until the next review increases, resulting in spaced practice.

The next version will track each learner's progress of language acquisition and decay, and dynamically generate a tailored curriculum that helps them quickly achieve or regain mastery of the language skills that they need for a particular purpose, such as completing a particular mission. Material will be organized by topic, as in the current curricula, but the material included under each topic will change dynamically based upon what the learner has previously mastered or is in danger of forgetting. If the system judges that the learner has fully mastered a particular topic, it will encourage the learner to skip it. Later, if the learner has stopped training on a particular topic for a period of time, the system will provide

[2] Reported probabilities calculated holding all other model variables constant at their grandmean.

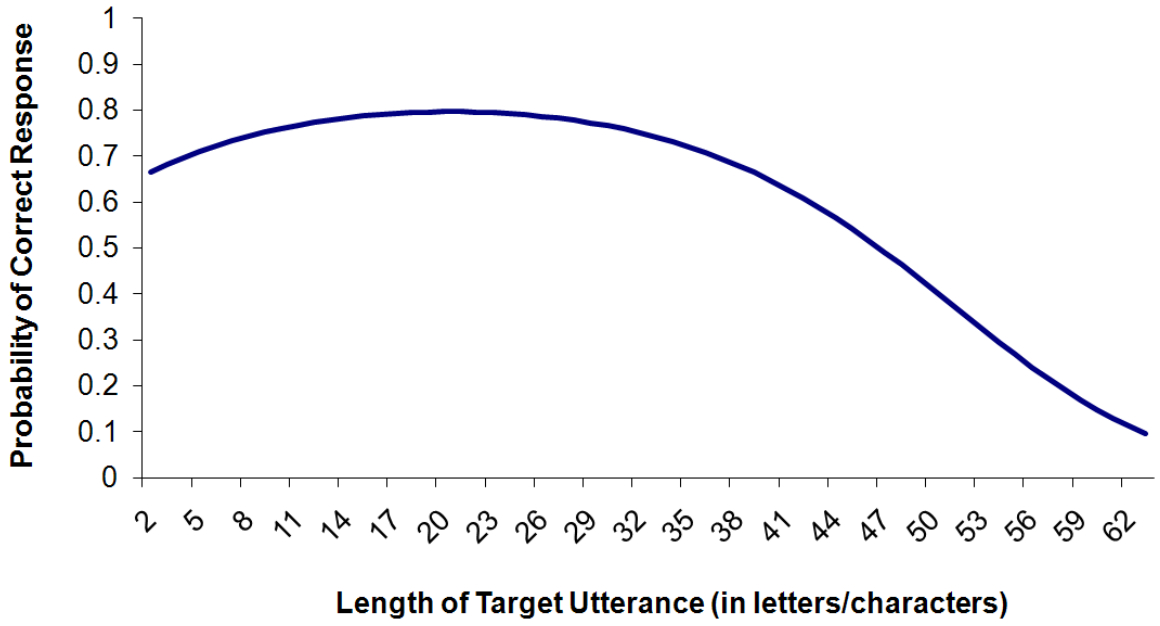

Figure 9. Length of Target Utterance and Probability of a Correct Response

the trainee with a focused set of exercises that will help them quickly recover their skills.

The key to the approach is a detailed learner model that tracks the learner's degree of mastery of each phrase and topic, based upon the theoretical language acquisition model and empirical studies of skill decay such as the one described in the previous section. For each phrase and vocabulary item, the learner model will track the type of skill the learner has demonstrated (recognition and recall, pronunciation, use in conversation) and the level of mastery demonstrated (declarative level vs. automatized, overlearned level). This requires tracking both the number of correct and incorrect attempts, as in the study in the previous section, as well as the amount of time the learner requires to complete a particular exercise. Each exercise and activity will be tagged with metadata indicating what language skills it teaches and/or tests. If an exercise involves language skills that the learner has already mastered, the system will let the learner skip the exercise and focus on the exercises that cover skills where the learner is weaker.

If the learner has stopped training for an extended period of time, the mastery estimates in the learner model may no longer be valid due to skill decay. In such cases the learner will be encouraged to complete a summary quiz or practice scenario covering a particular topic, and use the data to recalibrate the learner model. This will enable the system to recommend refresher and remedial training to trainees who have failed to master certain skills or are in the process of losing those skills. We also plan to incorporate a mechanism for predicting skill decay, precalibrated from analyses of archival learner data, so that the system is better

able to make recommendations of refresher training activities without having to first retest the learner's skills.

We plan to continue to collect data from field use of this sustainment training model, and use it to iteratively improve the personalized refresher training mechanism. This will involve progressively improving the ability of the learner model to account for and predict skill acquisition and decay, both statistically for the population of learners as a whole, as well as for each individual learner. In the longer term, we anticipate that individual differences in learner skill acquisition and decay patterns will emerge, which may be indicative of differences in learning strategies, metacognitive skill, and/or aptitude.

Summary

This chapter summarized the approach to developing intercultural competence underlying Alelo courses, and described some of the adaptive technologies used to implement this approach. The approach is particularly suited to the needs of lifelong learners, who may have specific needs for intercultural competence in their job or work, and who have relatively little free time to study and train. Course designs adhere to ten key principles, grounded in research in learning science, psychology, anthropology, and applied linguistics: task-based learning, language in culture, communicative approach, customization, localization, immersive learning, scaffolding, feedback, learner motivation, and skill sustainment. Courses are developed in accordance with the Situated Culture Methodology, which is based upon theories and methods from sociocultural and linguistic anthropology. An integrated technology architecture, incorporating adaptive technologies, supports authoring and delivery of the courses using computer game engines, Web browsers, and mobile devices. Data from field use of the courses is used to measure system performance and effectiveness, and identify areas for further improvement.

Experience indicates that when used properly, these courses can be very effective in promoting intercultural competence. For example, the 3rd Battalion, 7th Marines used Tactical Iraqi to prepare for its deployment to Iraq in 2006 and 2007. Each squad of thirteen Marines included at least two trainees who had trained for 40 hours with the course prior to deployment. This battalion did not suffer a single combat casualty during this deployment, and was the first Marine battalion to do so in the Iraq war. The Marine Corps Center for Lessons Learned interviewed the officers and surveyed the enlisted personnel afterwards (MCCLL, 2008). The study uncovered evidence that the level of intercultural competence of these Marines upon arrival in Iraq was much higher than that of previous units, and that this helped establish a positive relationship between the Marines and the local people and contributed significantly to mission effectiveness.

Future plans call for further expanding the utilization of the products described here, particularly the new mobile solutions. We plan to further develop the

personalized refresher training capabilities. Meanwhile we aim to expand the approach to new products aimed at additional learner communities, languages, and cultures.

Acknowledgments

The author wishes to express his thanks to the various members of the Alelo team who contributed to this work. Mike Emonts edited and commented on the manuscript. This work was sponsored by USMC PMTRASYS, Voice of America, Office of Naval Research, and DARPA. Opinions expressed here are those of the authors and not of the sponsors or the US Government.

References

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

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369-406.

Austin, J.L. (1975). *How to Do Things with Words*. Harvard University Press, Cambridge, MA (1975)

Baddely, A.D. & Longman, D.J.A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21,* 627-635.

Bahrick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*, 296-308.

Barrett, K.A. & Johnson, W.L. (2010). Developing serious games for learning language-in-culture. In *Interdisciplinary Models and Tools for Serious Games: Emerging Concepts and Future Directions*, Richard Van Eck, Editor. Hershey, PA: IGI Global.

Bernard, H. Russell, ed. *Handbook of Methods in Cultural Anthropology*. Lanham, MD: AltaMira Press.

Bienkowski, S. C., Watson, A. M., & Surface, E. A. (2010, August). Performance-avoid goal orientation and task engagement: Moderating effect of self-efficacy. Paper presented at the American Psychological Association Convention, San Diego, CA.

Birdwhistell, R. L. *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia: University of Pennsylvania Press.

Brown, P. & Levinson, S.C. (1987). *Politeness: Some universals in language usage*. New York: Cambridge University Press.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence.* Clevedon: Multilingual Matters.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics***,** *1*, 1-47.

Catalano, J.F. & Kleiner, B.M. (1984). Distant transfer in coincident timing as a function of variability of practice. *Perceptual & Motor Skills, 58,* 851-856.

Cormier, S.M. (1987). The structural processes underlying transfer of training. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp 152-182). San Diego, CA: Academic Press.

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

Driskell, J.E., Willis, R.P. & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615-692.

Earley, P.C. (1987). Intercultural training for managers: A comparison of documentary and interpersonal methods. *The Academy of Management Journal, 30*(4), pp. 685-698.

Ellis, R. (2003). *Task-based language learning and teaching*: Oxford University Press, USA.

Ellis, R. & Barkhuizen, G. (2005). Analyzing Learner Language. Oxford: Oxford University Press.

Estes, W.K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369-377.

Fiorito, L. (2005). Teaching English for Special Purposes (ESP). http://www.usingenglish.com/articles/teaching-english-for-specific-purposes-esp.html

Gee, J. P. (2003). What video games have to teach us about learning and literacy. Computers in Entertainment (CIE), *1*(1), 20-20.

Ghodsian, D., Bjork, R.A., & Benjamin, A.S. (1997). Evaluating training during training: Obstacles and opportunities. In M.A. Quinones & A. Ehrenstein (Eds.) *Training for a rapidly changing workplace,* (pp 63-88). Washington, DC: American Psychological Association.

Hall, E. T. (1966). *The Hidden Dimension.* Garden City, N.Y.: Doubleday.

Han, Z. & Odlin, T. (2006). Studies of fossilization in second language acquisition. Clevedon: Multilingual Matters.

Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Phildelphia, PA: University of Pennsylvania Press.

Hymes, Dell. (1987). Communicative Competence. In Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier, eds., *Sociolinguistics: An International Handbook of the Science of Language and Society*, pp. 219-229. Berlin: Walter de Gruyter.

Johnson, W.L. (2010). Using immersive simulations to develop intercultural competence. In Culture and Computing. Berlin: Springer-Verlag.

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

Johnson, W.L, Ashish, N., Bodnar, S., & Sagae, A. (2010). Expecting the unexpected: Warehousing and analyzing data from ITS field use. In V. Aleven, J. Kay, & J. Mostow (Eds.): *ITS 2010, Part II*, LNCS 6095, 352-354. Berlin: Springer-Verlag.

Jonassen, D.H., Tessmer, M., & Hannum, W.H. (1999). Task analysis methods for instructional design. Mahwah, NJ: Lawrence Erlbaum.

Kerr, R. & Booth, B. (1978). Specific and varied practice of motor skill. Perceptual & Motor Skills, 46, 395-401.

Kosoko-Lasaki, S., Cook, C.T., & O'Brien, R.L. (2008). Cultural Proficiency in *Addressing Health Disparities*. Boston: Jones & Bartlett.

Kumar, R., Sagae, A., & Johnson, W.L. (2009). Evaluating an authoring tool for mini-dialogs. In Proc. of AIED 2009. Berlin: Springer-Verlag.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Lave, J. (1988). *Cognition in practice*: New York: Cambridge University Press.

Lee, A. & Kawahara, T. (2009). Recent Development of Open-Source Speech Recognition Engine Julius. In Proc. of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference.

Lepper, M. R. (1988). Motivational considerations in the study of instruction.*Cognition and instruction*, 289-309.

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

McDonald, D.P., McGuire, G., Johnson, J., Selmeski, B., & Abbe, A. (2008). Developing and managing cross-cultural competence within the Department of Defense: Recommendations for learning and assessment. Technical report, RACCA WG.

Melton, A.W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596-606.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Mostow, J. & Beck, J. (2009). What, How, and Why should Tutors Log? In: Proceedings of EDM 2009, pp. 269-278.

National Foreign Language Center (NFLC) (2007). Cultural Proficiency Guidelines (3.2). http://www.nflc.org/culture-prof-guide_3-2.pdf.

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General*, 113, 32-35.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57-84). Hillsdale, NJ: Erlbaum.

Prensky, M. (2001). Digital Game-Based Learning. Columbus, OH: McGraw-Hill.

Sagae, A., Johnson, W.L., & Bodnar, S. (in press). Validation of a dialog system for language learners. Proceedings of SigDIAL, in press.

Salmoni, B. & Holmes-Eber, P. (2008). *Operational Culture for the Warfighter*. Quantico, VA: Marine Corps University Press.

Schneider, P., & Sadowski, D. (2009). The effects of intercultural collaboration strategies on successful PhD education. In S. Fussell, P. Hinds & T. Ishida (Eds.), *Proc. of IWIC 2009.* New York: ACM Press.

Schnieder, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors, 27,* 285-300.

Seeley Brown, J., Collins, A., & Deguid, P. (1989). Situated cognition and the culture of learning. Educational Research **18** (1), 32-42.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Traum, D., & Hinkelman, E. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence* 8, pp. 575–599.

US Dept. of Health and Human Services (USDHHS) (2008). Cultural competency curriculum for disaster preparedness and crisis response. https://cccdpcr.thinkculturalhealth.org/

VanLehn, K. (2006). The behavior of tutoring systems. *Int. J. of Art. Int. in Education* 16 (3), 227-265.

Van Rossum, J.H. (1990). Schmidt's schema theory: The empirical base of the variability of practice hypothesis: A critical analysis. *Human Movement Science, 9,* 387-435.

Vygotsky, L. S. (1986). *Thought and Language*. Cambridge, MA: MIT Press.

Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., & Collins, H. (2008). The Politeness Effect: Pedagogical Agents and Learning Outcomes. *International Journal of Human Computer Studies* 66(2), 98-112.