Personalized refresher training based on a model of competency acquisition and decay

W. Lewis Johnson[1], *Alicia Sagae*[1]

[1]Alelo Inc.
Los Angeles, CA, USA
{ljohnson, asagae}@alelo.com

ABSTRACT

This paper will describe current work at Alelo to track learner interactions with learning systems and analyze the data to derive profiles of learner competencies. This includes both linguistic knowledge (e.g., the ability to recognize and recall words and phrases) and linguistic skills (e.g., the ability to perform communicative functions involving speaking and listening). The learner modeling framework is based on a theoretical model of the learning and forgetting of complex contextualized cognitive skills. The model predicts that learners first acquire declarative knowledge and then automatize it, resulting in procedural skill. If acquired knowledge and skills are not used for extended periods, they undergo decay, depending upon the level of mastery achieved, the decay period, and the history of skill reinforcement. The model also takes into account the variety of contexts in which skills are used, in order to predict the robustness of learning, i.e., the ability to transfer learning to new situations.

Keywords: Learner Modeling, Training & Simulation, Language Acquisition, Skill Attrition

1 INTRODUCTION

To optimize the learning process it is important to take into account not just what each learner does and does not know, but also what they have learned in the past and may be in the process of forgetting. This is especially important for learning world languages and cultures. Most people who study a language in class and do not live in a community where the language is spoken is likely to go through extended periods without practicing the language. When an opportunity arises for using the language, they then face the challenge of quickly recovering those skills.

This paper will describe current work at Alelo to track learner interactions with learning systems and analyze the data to derive profiles of learner competencies. This includes both linguistic knowledge (e.g., the ability to recognize and recall words and phrases) and linguistic skills (e.g., the ability to perform communicative functions involving speaking and listening). The learner modeling framework is based on a theoretical model of the learning and forgetting of complex contextualized cognitive skills. The model predicts that learners first acquire declarative knowledge and then automatize it, resulting in procedural skill. If acquired knowledge and skills are not used for extended periods, they undergo decay, depending upon the level of mastery achieved, the decay period, and the history of skill reinforcement. The model also takes into account the variety of contexts in which skills are used, in order to predict the robustness of learning, i.e., the ability to transfer learning to new situations.

These competency profiles in turn are used to derive personalized learning trajectories for learner, a set of learning activities that focus on recovering and reinforcing the learner's language competencies.

In addition to facilitating refresher training, this approach can be used to improve self-study learning, by helping to ensure that learners retain their skills as they progress through the curriculum. It can also help improve the efficiency of classroom and group instruction, by helping learners with varied backgrounds and skill to prepare for the learning activity.

Cross-cultural competency is a critical need for military personnel. For example, the US Defense Regional and Cultural Capabilities Assessment Working Group has identified the ability to integrate cultural knowledge and skills into mission execution as a critical cross-cultural competency for general purpose forces (McDonald, et al., 2008). Training of these skills, knowledge, and abilities is resource-intensive for both trainees and organizations. Simulation-based training promises anytime, anywhere access that can allow instructional material designed by a single trainer to be delivered in an effective, interactive way to thousands of trainees at lower cost and higher convenience (Fletcher, 1990). However, when instructors and domain experts encode this material, the current tools offered to them typically produce script-like, monolithic data structures that are culture-specific, non-reusable, and difficult to update or apply to new cultures and missions.

In the CultureCom project, we address these problems by developing a new system for creating training simulations in cross-cultural competency; this system produces flexible, model-driven simulations that apply culture-general and culture-specific rules in combination. As a result, we achieve the novel capability to swap cultural models, in the form of rule sets, in and out of a social simulation to reveal pedagogically relevant differences at the level of behavior (utterances, gestures) and intention (communicative function). For comparative analysis, we encoded a pair of simulations using the CultureCom architecture and using a state-of-the-art architecture based on finite-state-machines. We show that the new model-driven architecture requires comparable authoring time for an initial simulation, but allows more objects to be reused, reducing authoring time and total number of objects crated for each subsequent simulation. We also show that the risk of inconsistent simulation behavior is reduced using the new architecture.

2 A MODEL OF SKILL ACQUISITION AND DECAY

In 1885, Hermann Ebbinghaus discovered the exponential nature of forgetting. The following formula can roughly describe the forgetting:

$$R = e^{-\frac{t}{S}}$$

where *R* is memory retention, *S* is the relative strength of memory, and *t* is time. Ebbinghaus also discovered that repeated training increases the relative strength of memory, and so after each subsequent training episode the profiles of memory decay, i.e., the forgetting curves, are progressively shallower.

Classical verbal learning research serves as the motivation for many current language learning products, such as CL150 Rapid Rote. These products focus on vocabulary memorization, and seek to optimize the presentation of vocabulary prompts to learners so that they quickly develop the ability to recall vocabulary items. However, language skill involves more than recall of memorized vocabulary items; it involves the fluent production of appropriate language in context. To properly predict the the acquisition and loss of complex contextualized langauge skills, a richer theoretical model is required.

3 INITIAL EXPERIMENTS

We conducted an analysis of data collected from people who trained with Tactical Iraqi for extended periods. This included Marines who had trained with Tactical Iraqi at military training sites, as well as Alelo personnel who had learned languages using Tactical Language courses. We were interested in data from learners who had trained for extended periods of time, and preferably had gaps in their training history, which presumably would lead to attrition in language performance.

3.1 Study of Archival Marine Learner Data

This study sought to operationalize factors that may contribute to language acquisition and attrition and empirically test the predictive relationships between these factors and speech production during TLTS training. Three potential indicators of speech production skill regularly tracked with TLTS users were identified. Firstly, correct versus incorrect recognition of an utterance by the automated speech recognizer reflects learners' general ability to accurately produce speech in the target language. Empirical analysis was conducted using this skill indicator in the current study. Secondly, the duration of time to produce correct utterances reflects learners' ability to process and produce speech quickly and fluidly. Lastly, appropriate attempts to transfer utterances to conversations in the Mission Game reflect learners' ability to recall and generalize utterances learned during Skill Builder lessons to a simulated transfer environment.

We contacted three Marine training sites with significant amounts of training activity: Marine Corps Air Ground Combat Center (MCAGCC) 29 Palms, Camp Lejeune and Camp Pendleton. Of these three, we selected data from 29 Palms because they had both logged user profiles and the actual speech recordings of Marines speaking into Tactical Iraqi. The other sites had configured Tactical Iraqi to turn off saving of speech recordings. We retrieved a total of 294 separate trainee profiles from 29 Palms, and out of these 34 had trained for at least a month, and had skipped at least a week at some point during their training. These trainees appeared to be good candidates for studying the effects of language attrition. Many of these trainees had conducted extensive training with Tactical Iraqi, at least 25 Skill Builder lessons. These learners were native English speakers. Training dates for data used in this study ranged between October 2007 and September 2008. The final data set consisted of 9,615 unique speech attempts made during the course of training. They had trained for over periods of one to four months, often with gaps of a month or more during which language attrition could occur. We used these 20 training profiles as the basis for analysis. These profiles were assigned anonymous codes to protect the identity of the Marines involved.

Several potential predictors of utterance success in the practice setting of the TLTS Skill Builder were identified and operationalized for this empirical study. The number of previous speech attempts both in general and for specific target utterances was calculated. The number of previous successful utterances with a given target utterance was also calculated. 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 a given target utterance was computed. This variable reflects a retention interval, representing the amount of time that has passed since the learner has performed the skill of successfully producing the target utterance. 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 given target utterance was calculated. To the extent that the other utterances were semantically similar to the target utterance, this variable may reflect interference.

While the variables operationalized in the current study do not provide exhaustive coverage of the potential factors influencing skill retention, they do reflect the best and most appropriate information we could collect from the data currently available using TLTS logs.

3.2 Data Analysis and Results

Detailed data regarding learner behaviors while training with Tactical Iraqi were collected and tracked by the system. Automated logs tracked the content of specific training pages visited, as well as learner speech interaction with the system. Data for individual speech attempts were extracted from these system logs and employed in the current analysis. Table 1 details the variables utilized in this study.

Table 1. Description of Study Variables Obtained from Tactical Iraqi Logs

Variable	Description
Learner	Identifies learner
Page Type	Identifies type of training page (e.g., pronunciation training, memorization recall, etc.)
Target Utterance	Target utterance of speech attempt
Recognized Utterance	Recognized utterance of speech attempt
Correct	Coded '1' if recognized utterance matches target utterance (otherwise coded '0')
Previous Total Attempts	Number of previous speech attempts made by learner across all target utterances
Previous Target Attempts	Number of previous speech attempts made by learner for the current target utterance
Cumulative Target Correct	Number of previous correct attempts on current target utterance
Time Since Correct Attempt on Target	Number of hours that passed since the last correct attempt on the current target utterance
Intervening Targets Since Last Attempt	Number of other target utterances attempted since the last attempt on the current target utterance

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 (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. As is customary in multilevel modeling, a series of models were 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 presented in Table 1. 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.

Across the 9,615 speech attempts, there was significant variability in the probability of speech success at level 2 (between target utterances and within speakers). As indicated by grandmean log-odds intercept (□00) of the unconditional model (□00 = .845), the grandmean average probability of success across targets was around .70 (i.e., the average speaker got 7 in 10 speech attempts correct), but the 95% confidence interval ranged from average probabilities as low as .27 to as high as .93. The level 1 explanatory variables accounted for approximately 60% of the variance in likelihood of success between target utterances (as indicated by the proportional reduction of the random effect for probability level, □00, from the unconditional to the conditional model), while the level 2 variables didn't account for any additional variance across target utterances. Thus, the explanatory variables modeled accounted for a sizable amount of the observed variability in success between target utterances.

Unsuccessful speech attempts were generally negatively related to the likelihood of future success for a given target utterance, but this relationship lessened as the number of unsuccessful attempts increased. For example, holding all other variables in the model constant (at their grandmeans), the first unsuccessful attempt at an utterance (after 1 prior correct attempt) was associated with a .029 decrease in the probability of the success on the next attempt. The next unsuccessful attempt was associated with a slightly smaller decrease (.024) in the probability of success on the next attempt.

In contrast, successful speech attempts were positively related to the probability of future success. This relationship lessened as the number of successful attempts increased. Additionally, a successful attempt was less indicative of future success than an unsuccessful attempt was of future failure. That is, holding all other variables in the model constant (at their grandmeans), the second successful attempt at an utterance was associated with a .015 increase in the probability of the success on the next attempt. The next successful attempt was associated with a slightly smaller increase (.011) in the probability of success on the next attempt.

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 1. This relationship was moderated by how frequently the target utterance appeared in the training content, such that more frequently appearing utterances showed a slightly larger decrease in likelihood of success as function of time since the last correct attempt relative to less frequently appearing utterances.

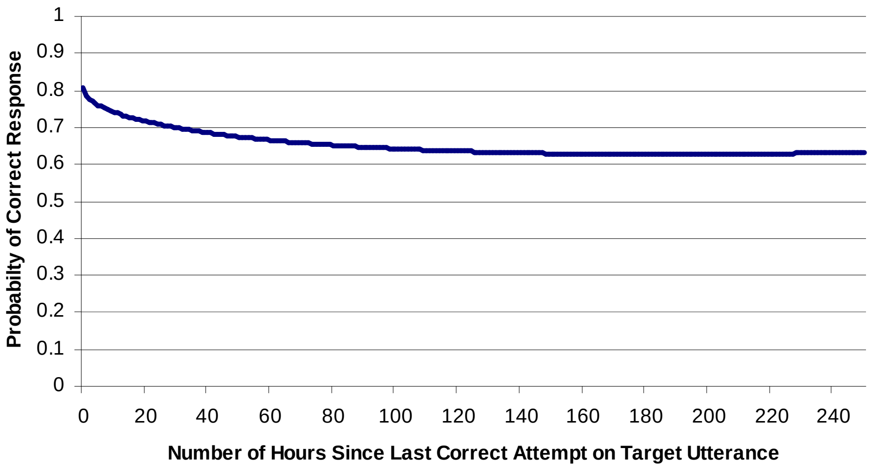

Figure 1. Time elapsed since last correct attempt and probability of a correct response

The length of a target utterance (in letters/characters) showed a curvilinear impact on probability of success. As the target length increased the relationship changed from positive to negative. The unique relationship between utterance length and probability of success is presented in Figure 2. This relationship was moderated by how frequently the target appeared in the course content, such that target length has a slightly larger impact on the likelihood of success as the frequency of occurrence in training increases. This finding is consistent with our predication that the learners would make heavy use of memorized phrases, which tend to be relatively short, but more than single words. When learners are called upon to produce memorized phrases, their performance rate was relatively high, even higher than it as for individual words. When learners were required to produce longer phrases that were not memorized, their performance decreased rapidly. The finding may also be influenced by the technical characteristics of the speech recognition algorithm, which tends to perform with higher reliability on longer phrases than on individual words.

Lastly, the proportion of non-English phonemes in a target utterance was negatively related to the probability of success (see Figure 5).

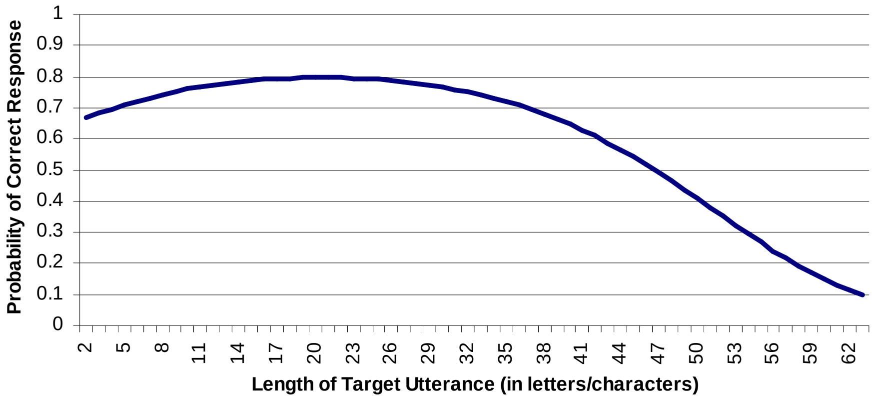

Figure 2. Length of target utterance and probability of a correct response

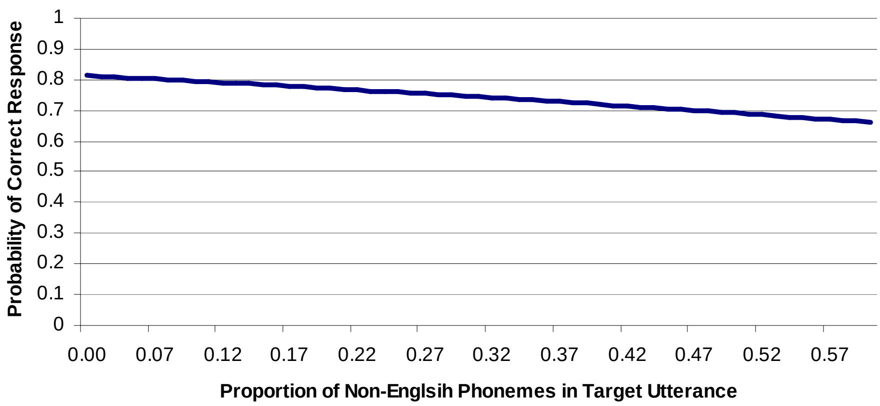

Figure 3. Proportion of non-english phonemes in target utterance

3.3 Preliminary Conclusions

These results support several general conclusions regarding the training behaviors and performance observed in the available data. Within individual learners, there were substantial differences in the rate of success between different target utterances. These differences were largely attributable to the variables examined in this study. By accounting for these variables (i.e., holding them constant), the differences in the success rate across target utterances was reduced by 60%.

Also, the learners' speech attempts were recognized as correct approximately 70% of the time, on average. This suggests producing correct speech attempts was not overly difficult for these learners. The results suggest that producing a correct or an incorrect speech attempt is predictive of future success (or failure) for a given target utterance, even for those learners who have demonstrated the ability to correctly produce the utterance (which was the entirety of our sample). Additionally, speakers who fail to produce a correct utterance are more likely to fail the next attempt than speakers who produce a correct utterance are to be correct again on the next attempt.

As shown in Figure 3, learners were less likely to produce a correct speech attempt as the interval of time since their last correct attempt increased. That is, with a longer retention interval came greater probability of attrition. Interestingly, this relationship was equally strong for more complex (i.e., longer utterances, with a greater proportion of non-English phonemes) and less complex utterances. The complexity of target utterances did impact the likelihood of correct utterances. The impact of target utterances length was complex (see Figure 2), while a higher proportion of non-English phonemes was always associated with a lower likelihood of success (see Figure 3).

These results illustrate the potential utility of tracking learner behaviors and response patterns, in that future likelihood of success could be monitored and could even be used to dynamically adjust the order or manner in which content is presented to guard against skill decay.

4 A HIERARCHICAL MODEL OF LANGUAGE COMPETENCIES

Based on these preliminary experiments, we are currently developing a more sophisticated model that can be used to estimate learner proficiency not only with particular spoken utterances, but across a network of related knowledge, skills, abilities, and attitudes (KSAAs) that comprise linguistic competencies. In addition, we plan to extend the application of this model to not only highlight lessons that the learner should revisit, but to dynamically select and reorder content that is personalized for the current learner and his or her learning needs, so that learning stays within the learner's zone of proximal development (Vygotsky, 1978).

The centerpiece of our approach is a simple but powerful framework for modeling the range of world language KSAAs. It encompasses not just vocabulary mastery (as is commonly the case for competing products), but also language use in context, and culture. It therefore gets to the heart of learning goals we want to accomplish in accordance with the ACTFL 5C standards. Our modeling approach is simple enough that language teachers will be able to understand and use it with relatively little training. Yet it is rich and powerful enough that we can develop computational tools that can work with it automatically in the context of learning. It relies on the connections among KSAAs and between KSAAs and other types of data.

KSAAs will be structured and organized as follows.

- They are grouped into 3 categories: forms, functions, and practices.
- Forms include various subcategories of linguistic forms and patterns, such as sentence structures, vocabulary, and phrases;
- Functions cover the use of language to express and communicate meaning, most notably communicative functions;
- Practices cover the cultural norms and practices governing communication and language use;
- They can have attributes that characterize and circumscribe their use. One pervasive set of attributes comprises speaking, listening, reading, and writing;
- They are organized into subcategories (taxonomies) and part-whole hierarchies (meronomies). Conversely, they can be grouped into collections. The result is a rich lattice of interconnected KSAAs that allows our tools to reason about KSAAs and their mastery.

Figure 4. Virtual Virginia/Alelo Chinese prototype

This framework allows us to take advantage of Alelo's unique platform for delivering language and culture training, shown in Figure 4. In this figure, the student's character (left) is arranging to meet with a friend named Zhang Li (right), an animated character. The learner speaks into the microphone in Chinese, and Zhang Li responds in Chinese. Zhang Li has just asked the learner: "Wǒmen lǐbàijǐ jiànmiàn?" (What day should we get together?) The learner needs to respond by suggesting a day to meet. A help menu on the left lists a number of possible ways of responding, with different days of the week and choices of phrasing. Once the learner is able to converse without reliance on prompts and hints, the dialog becomes a realistic simulation of real-world task-oriented dialog.

With the model described in Section 3, we are only able to capture the identity of the utterances the learner practices in this conversation. In contrast, our new structure for modeling KSAAs allows us to capture more of the unique advantages provided by this training system, taking them into account in the model of learner proficiency. Using Figure 4 as an example, Zhang Li (the virtual character) says "Wǒmen lǐbàijǐ jiànmiàn" (What day should we get together?). This phrase is a linguistic form made up of component parts (words and subphrases). It utilizes grammatical forms (e.g., the use of the "jǐ" particle to form a question). It has a modality attribute (listening, since Zhang Li said it to the learner). It is an expression of a communicative function (requesting a meeting date). It conforms to a cultural practice (absence of face-threat mitigating strategies, as is customary for conversation among friends). So there is a rich network of KSAAs associated with just this one example.

5 FUTURE WORK

ACKNOWLEDGMENTS

The authors would like to acknowledge the Office of Naval Research for their support of this work.

REFERENCES

Fletcher, D. J. (1990). Computer-Based Instruction: Costs and Effectiveness. In A. Sage (Ed.), Concise Encyclopedia of Information Processing in Systems and Organizations. Elmsford, NY: Pergamon Press.

Johnson, W. L., & Friedland, L. (2010). Integrating Cross-Cultural Decision Making Skills into Military Training. In D. Schmorrow & D. Nicholson (Eds.), Advances in Cross-Cultural Decision Making. London: Taylor & Francis.

Johnson, W.L. & Valente, A. (2009), "Tactical Language and Culture Training Systems: Using AI to teach foreign languages and culture." AI Magazine, 30 (2), 72-83.

McDonald, D. P., McGuire, G., Johnston, J., Selmeski, B., & Abbe (2008), "Developing and Managing Cross-Cultural Competence Within the Department Of Defense: Recommendations For Learning and Assessment." US Defense Regional and Cultural Capabilities Assessment Working Group.

Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods, 2nd Ed. London: Sage Publications.Hermann

Sagae, A., Johnson, W. L., & Valente, A. (2011). Conversational Agents in Language and Culture Training. In D. Perez-Marin & I. Pascual-Nieto (Eds.), Conversational Agents and Natural Language Interaction: Techniques and Effective Practices (pp. 358-377). Madrid: IGI Global.

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes (Ed. by M. Cole, V. John-Steiner, S. Scribner, & E. Souberman). Cambridge, MA: Harvard University Press.