Conversational Agents in Language and Culture Training

Alicia Sagae, W. Lewis Johnson, Andre Valente
Alelo, USA

ABSTRACT

This chapter describes the design, implementation, and use of an agent architecture that has been deployed in Alelo, Inc.'s language and culture training systems, which offer practical training for foreign language skills and intercultural competence. These agents support real-time conversation in the language of interest (Dari, Pashto, Arabic, French, and others), using automatic speech recognition and immersive simulation technologies. In earlier work we developed a number of agent-based language and culture trainers, based on the Tactical Language and Culture Training System platform (Johnson & Valente, 2008). Our experience has revealed a number of desiderata for authorable, believable agents, which we have applied to the design of our newest agent architecture. In this chapter we describe the Virtual Role-Players (VRP), an agent architecture that relies on ontological models of world knowledge, language, culture, and agent state in order to achieve believable dialogue with learners. Authoring and user experiences are described, along with future directions for this work.

INTRODUCTION

Conversational agents have particular appeal in the educational domain of training learners in communicative competency. Experts in the language education community, such as the American Council on the Teaching of Foreign Languages (ACTFL), describe communicative proficiency as a goal that includes both *declarative knowledge* and *procedural skill* (Lampe, 2007). In a computer-based learning environment, conversational agents allow students to engage these procedural skills by speaking in real time and observing believable responses.

Our concern is the development of conversational agents for practical training systems for foreign language skills and intercultural competence. Such agents must meet a range of challenging design constraints. They need to produce behavior that has an appropriate level of realism and cultural accuracy, so that they provide a suitable model for training. At the same time, they need to be easily authorable and configurable, by people who may not be specialists in agent modeling frameworks and formalisms. Ideally it should be possible for trainers and educators to create their own training scenarios, populate them with conversational agents, and have the agents behave in culturally and situationally appropriate ways.

In our earlier work we developed a number of agent-based language and culture trainers, such as the Tactical Iraqi trainer for Iraqi Arabic (Johnson, in press). These have been used widely by military service members and other individuals preparing for overseas work. The trainers include a large number of practice dialogues and scenarios, each of which includes one or more conversational agents. All told, a typical course includes fifty or more agents, each designed to

Sagae, A., Johnson, W. L., & Valente, A. (to appear). 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*. Madrid: IGI Global.

converse with learners in a particular scenario or situation. The behavior of each agent is specified in a finite state machine framework.

Although this has proven to be an effective approach for creating conversational agents in this domain, it suffers from some key limitations. One is that each agent has to be authored for the specific scenario context in which it is intended to be used. This multiplies authoring effort and limits the number of agents that can be incorporated into a given scenario. It also prevents trainers from creating new training scenarios or adapting scenarios to their own needs. Another limitation is that it is up to agent authors to make sure that the agent's behavior is culturally appropriate, and it is difficult to validate whether they have done so. To overcome these limitations we have recognized the need to develop an agent authoring framework that supports the creation of agents that can be employed flexibly in a range of training scenarios, and which incorporate explicit, validated models of culturally appropriate behavior.

In response to these and other desiderata, this chapter introduces Alelo's new Virtual Role-Player framework, a conversational agent architecture that has been employed in a number of language and culture training systems. The Virtual Role-Player (VRP) architecture is a flexible platform for combining models of conversation, facilitating model reuse and behavior validation. Models of politeness, culture, and strategy (e.g. lying for a social purpose) have all been implemented in this framework in an effort to drive more believable agent behavior. The result is a set of artificially intelligent, conversation-ready agents who can be attached to visual avatars in a variety of serious game environments. These agents are called Virtual Role-Players (VRPs). Individual VRPs have been instantiated as characters in Alelo's Operational Language and Culture Training System series (OLCTS), and in mission rehearsal applications built on the 3-D platforms Virtual Battlespace 2 (VBS2) and RealWorld.

In this chapter we connect our experience with conversational agent systems to a selection of related examples from the literature, followed by a detailed description of our newest agent architecture. In the remainder of this section, we explain the context of this work with examples from Alelo's OLCTS language and culture training suite. This leads to a set of desiderata for a new agent architecture, and a discussion of related work using these desiderata as a guide. Following related work, we give a technical overview of the VRP architecture, along with additional detail on the models used for dialogue and culture in VRPs. Finally, we describe user experiences of authoring and interacting with VRPs.

BACKGROUND

Language and Culture Training with Alelo Conversational Agents

Many users of Alelo courses are members of military services who require training in intercultural skills prior to deployment overseas on missions such as reconstruction and humanitarian assistance. Figure 1 shows an episode from one such course, the US version of the Operational Dari language and culture trainer. In this example the learner, playing the role of the military character on the left, is meeting with a local leader (known as a malek) and other elders in an Afghan village. To play the role, the player must speak into the microphone in Dari, the local language of the area, as well as choose culturally appropriate gestures. The dialogue history is shown in the top center of the figure. The player needs to develop rapport with the malek in order to gain his cooperation. For this reason, the dialogue history up to this point in the episode consists of greetings and inquiries about the malek's family, and there has been no discussion as yet of the business purpose of the meeting, namely the reconstruction of a local school.

Figure 1. Operational Dari language and culture training system. (© 2010, Alelo)

The Operational Dari training suite includes interactive lessons that help learners acquire the communication skills that they need in situations such as the one depicted in Figure 1, and a variety of immersive episodes in which learners can practice their skills. These include single-player learning environments, as shown in Figure 1, as well as multi-player environments, in which trainees can practice their communication skills as part of an overall simulated military mission. Figure 2 shows a multi-player episode, or *scene*, running in Virtual Battlespace 2 (VBS2). The player's intent in this conversation is similar to that in the previous example: to build rapport with a member of the village to gain their cooperation.

Figure 2. Scene from an Operational Dari scene in Virtual Battlespace 2. (© 2010, Alelo)

As these examples illustrate, Operational Dari and other OLCTS courses enable learners to converse with a number of different agents. The agents tend to have many similarities in their behavior, as well as differences specific to that conversational situation. For example, the Afghan agents in Figure 1 and Figure 2 can both converse about their respective families, but the details of their families may differ. More importantly, there may be differences in the two conversations stemming from the fact that one is a formal gathering and the other is a chance encounter on the street. This poses challenges from an agent modeling standpoint. In our older Tactical Language courses, separate agent models had to be constructed for each agent in each episode, without any opportunity for reuse. This resulted in a substantial amount of authoring effort to develop each agent and then validate its behavior to make sure that it is culturally and situationally appropriate. The Virtual Role Player architecture helps to reduce this effort. The agents in Figure 1and Figure 2 can both be developed from a common agent model for Afghan elders, and customized as needed for each specific conversational situation.

Desiderata for Communicative Agents

With experience from these systems in mind, we can describe some of the desiderata for a new communicative agent architecture that is appropriate for building large-scale language and culture training courses.

Desideratum 1: Reusable characters. Full-scale training systems, comprising multiple training scenarios and situations, require a method for reusing characters across situations. To see why, consider the alternative, which would be to author each agent in a context-specific way, with a set of verbal and non-verbal responses that apply only in the specific scenario where the agent is intended to appear. The effect of having a character appear in multiple scenarios can be achieved in this approach only by creating multiple agents with some overlapping behavior (both respond with “I am Razan” when asked for their name) but which are non-interchangeable due to scenario-specific details (when asked “how is your family,” one responds “they are well” while the other responds “my son has been injured”).

Such an approach multiplies authoring effort and limits the number of agents that can be incorporated into a given scenario. It also prevents trainers from creating new training scenarios or adapting scenarios to their own needs. Instead, we desire communicative agents to be authored in a reusable fashion, with general behavior rules that can apply in multiple scenarios. In the example given above, the agent Razan should have a programmatic way to determine and then report on the health of his family, rather than relying on the author to enter context-specific answers by hand. This principle applies to all layers of the agent architecture; when possible, we prefer to build reusable components that can be composed to achieve a desired agent behavior.

Desideratum 2: Explicit models of appropriate behavior. The connection between authoring processes and agent behavior is another important consideration. In many cases, the designer of agent behavior is a subject matter expert (SME) who has detailed knowledge of what constitutes “believable” language, gesture, and movement for the character who will be played by the communicative agent. How is this knowledge captured in the authoring process, and how is it translated into the programmed instructions for the agent? In many systems, including the Alelo’s earlier finite state machine (FSM) approach, the knowledge is captured informally, with each author working from his own mental model and attempting to author dialogues that conform to it. In the new architecture, we desire a way to use explicit models of culturally appropriate behavior that can be examined and validated independently, then dropped into the agent architecture, ensuring greater consistency. To further improve authorability, the architecture would support libraries of predefined models, under the hypothesis that modifying or adapting models is easier than creating them from scratch.

Desideratum 3: Awareness. In similar fashion to explicit models of behavior, we propose that communicative agents in this setting should have access to explicit factual knowledge about themselves and their social and physical environment. To apply social norms correctly, an agent must be aware not of the concept of social distance, but also of the agent’s specific position in society, in order to reason about the relative social distance in its communications with the learner.

The agent must also have an internal set of beliefs and knowledge about the world, in order to behave appropriately. This supports flexible behavior of the type described in our earlier example: ideally, we want an agent to respond to the question “how is your family” based in part on whether or not the agent actually knows this information, or even if he has a family. World knowledge should include the physical and cultural environments, and the current state of the conversation. For example, the agent may condition certain responses on whether or not the learner has established enough rapport to be trusted. Some decisions may even depend on synthesis of facts from the physical and cultural environment. For example, to decide on an appropriate gesture (say, shake hands), agents may need to know not only about the social distance with an interlocutor but also the physical distance.

Desideratum 4: Performance. Performance factors are critical for agents that operate in immersive environments, particularly when there are large numbers of such agents operating at the same time. 3-D graphics rendering and automated speech rendering both require a considerable amount of computational processing, and compete with agent reasoning for computing resources. Agent models that take a long time to make a decision, or that compete strongly for computing resources, can adversely impact the believability and usability of the simulation.

RELATED WORK

The systems described in this chapter draw on concepts from the literature in social science and computational modeling, in addition to conversational agent systems. We address each of these areas in the section below.

Social Science and the Science of Learning

The language and culture training curricula at Alelo are founded in research in the social sciences, including socio-cultural and linguistic anthropology, as well as the science of learning. This research plays an important role in choosing the desiderata described above, since the end goal for conversational agents in Alelo courses is to provide an effective method for exercising and building learner knowledge, skills, and attitudes (KSAs) in accordance with these curricula.

Alelo courses are designed to teach intercultural competence: the ability to communicate successfully with people of other cultures (Byram, 1997). Because many of our products are deployed to service members in the US military who are preparing to be deployed overseas, we also consider research on intercultural competence from military institutions. The recommendations of the Defense Regional and Cultural Capabilities Assessment Working Group (RACCA WG) are a seminal publication describing cross-cultural competence in this community (McDonald, McGuire, Johnson, Selmeski, & Abbe, 2008). They include a list of 40 learning statements, covering knowledge, skills, and personal characteristics, which are designed to foster cross-cultural competence in military and civilian personnel. An example from each category is shown below:

- Knowledge: Knowing cultural concepts and processes
- Skills: Integrating culture into planning and execution for mission success
- Personal Characteristics: Demonstrating a willingness to engage

Within the domain of intercultural competency, we adopt a task-based instructional strategy that is well aligned with other task-based approaches to language learning, such as Ellis (2003). Our approach is described in the Situated Culture Methodology, or SCM (Valente, et al., 2009). The SCM focuses on communication necessary to perform real-world 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, Tessmer, & Hannum, 1999). The learning activities in Alelo courses, including conversations with agents, are all task-oriented in this sense.

These learning objectives and strategies from the literature on cross-cultural competence underscore the importance of our agent desiderata. Explicit models of appropriate behavior allow us to build learning objectives such as those described by the RACCA WG into the conversational agent software, by associating the objectives with learner behaviors (i.e. specific turns in a dialogue) and tracking when those behaviors occur. When these behaviors are complex, the agent must rely on more than the most recent learner input to identify them. For example, demonstrating willingness to engage may be a behavior that is associated with an entire dialogue. The desideratum of awareness captures this need for a breadth and depth of knowledge on the part of the agent. Reasonable performance ensures that the training experience closely resembles the real-world setting where the learner will have to apply his skills, and re-usable characters allow us to generate a variety of experiences that are tailored to individual learners' needs. In a task-based curriculum, for example, we may want the learner to collaborate with the same character on a variety of different missions.

Computational Models of Dialogue

Social science literature also influences the individual models of dialogue that are implemented in VRPs. In many conversational agent systems, models of human behavior drive the agents' reaction to user input. These models can capture features of personality that affect the agent in all interactions, or emotional features that arise in reaction to a particular chain of events (Marsella & Gratch, 2009). They often capture details of the particular conversational context, as in goal-driven models of negotiation (D. Traum, Marsella, Gratch, Lee, & Hartholt, 2008) or task-based models of conversation with explicit conditions for success (Johnson & Valente, 2009).

Because VRPs engage in interpersonal dialogue with one player or a small group, social science models from the field of microsociology are especially relevant. Microsociology focuses on the dynamics of everyday social life, and on the structure and processes of how society works in real-time, on the ground. This tradition can be traced to Max Weber (1994), one of the founders of sociology, and Alfred Schutz (1970), who grounded the approach in the philosophy of Husserl. It has continued to develop in the disciplines of social psychology (Blumer, 1986; Goffman, 1990; Mead, 1967), ethnomethodology (Garfinkel, 1967), sociology (Habermas, 1984), and anthropological "practice theory".

These theories have influenced the formal microsocial models of Hobbs & Gordon (Hobbs & Gordon, 2010), which are used directly in the VRP architecture to model socio-cultural expectations for interpersonal dialogue. Among the other models of social behavior that are used in VRPs are politeness (Wang, et al., 2008), and strategy (e.g. lying for a social purpose) (Sagae, Wetzel, Valente, & Johnson, 2009).

Conversational Agents

VRPs apply these background theories in the context of a conversational agent system; specifically, a pedagogical system where the conversational behavior of the agent is both a teaching tool and a teaching target.

A wide variety of related systems exist in the literature today. Many of these focus on generation capabilities – the ability of the agent to express its affective states in ways that are appropriate for a given conversation. Greta is one example of a conversational agent aimed at realistic behavior generation (Rosis, Pelachaud, Poggi, Carofiglio, & Carolis, 2003). Like a VRP, Greta is designed to use explicit models to drive believable expressions of the affective state of the agent. However Greta's models focus on the agent's self-knowledge and her expressions are performed on a micro-scale, using features like gaze, eyebrow, and mouth position. Other agent systems in this class include Laura, the health counseling agent (Tmothy Bickmore & Pfeifer, 2008), the Sensitive Artificial Learners (McRorie, Sneddon, Sevin, Bevacqua, & Pelachaud, 2009), and experiments in agent emotion by Marsella and Gratch (2009) and Mello and Gratch (2009). These agents model their affective states at a fine level of granularity, and they express these states as a way of building rapport with the user.

VRPs contrast with such systems in several ways. First, VRPs are typically used to teach the *user* how to build rapport with the *agent*. This distinction is important, since VRPs can be used in training scenarios where rapport-building is intentionally challenging or delicate (as in simulated negotiations). Second, the models of conversation used in VRPs are more coarse-grained, in order to focus on dialogue-level effects such as dialogue outcome (e.g. did the user manage to perform required social conventions? Did the user negotiate a commitment for future action from the NPC?). This granularity also allows us to focus on broad socio-cultural influences, rather than personal influences, on agent behavior. Most interactions with VRPs are intended to equip the user with communication skills, applicable in any conversation with a similar cultural and linguistic context, not to build an ongoing relationship between the user and a particular VRP character. Finally, VRP models are intentionally designed with authorability in mind. Because VRPs are deployed in a large number of systems, new models must be derived from existing ones quickly, by non-programmer authors. In summary, the agent systems described above are designed to have very emotionally expressive behavior in a small range of circumstances, while VRPs are designed to have flexible behavior, which is culturally and linguistically appropriate in a wide range of circumstances (different languages, different cultures, different tasks).

Other agent systems that have influenced our work on VRPs by addressing socio-cultural factors in task-based dialogue include cultural agents by Roque and Traum (2007), and conversational modeling and turn-taking by Vilhjálmsson and Thórisson (2008). These systems both make use of rich object representation of dialogue context, in support of the desideratum of awareness.

Finally, VRPs draw on the literature in pedagogical agents such as Autotutor (Graesser, Chipman, Haynes, & Olney, 2005), Why-2 Atlas and Atlas-Andes (Rose, et al., 2001), among many others. Some of these pedagogical agent systems communicate with the learner in natural language, as VRPs do. However few of these systems are intended to teach, as well as use, natural language communication. As a result, it is often possible to assess the linguistic and cultural performance of the agents separately from their pedagogical strategies. In contrast, the communicative performance of a VRP serves not only to communicate to the learner, but to serve as an example of the skill being taught.

THE VRP ARCHITECTURE

The following is a discussion of the VRP architecture, and how it is used to create communicative agents (or VRPs). We focus on the use of VRPs that engage in multimodal communication with trainees in a foreign language, using a combination of speech and gesture, as shown in the examples in the introduction.

This is the most challenging application of the VRP architecture, since it integrates many different aspects of behavior interpretation, intent planning, and behavior generation. VRP characters can also be realized in training modalities where the interaction between the trainees and the communicative agents is more limited, e.g., trainees select communicative actions from a menu of options, or communicate with the character by typing rather than speaking. The particular choice of interaction modality depends upon the training objectives of the course and the expected skill set of the trainees undertaking the activity. The VRP architecture is designed to be flexible enough to support a range of different interaction modalities.

In multimodal dialogue scenarios, user actions are recognized from menu-driven gestures along with speech input that passes through an automatic speech recognition (ASR) module, and VRP responses are performed by 3-D avatars with speech and gesture generation capabilities. The result is a character who moves and acts physically in the virtual environment, but who also listens and responds in real-time conversations with human users. At the heart of this system is the social simulation pipeline, the software component that gives each agent its ability to understand and respond to learner inputs. In this section we discuss the social simulation pipeline used by VRPs.

Overview

The architecture adopts a variant of the SAIBA paradigm (H. Vilhjalmsson & Marsella, 2005), which separates intent planning (the choice of what to communicate) from production of believable physical behavior (how to realize the communication). The current version extends the architecture described by Samtani, Valente & Johnson (2008).

A visual representation is given in Figure 3. The player usually engages the system by speaking into a microphone. Starting from the lower-left corner of the area labeled Social Simulation Module, player speech input is captured and passed to an automatic speech recognizer (ASR). The ASR component produces a plain-text string, stored in an object called an *utterance*, which is used by the behavior interpretation module to generate a meaning representation. The player may also perform gestures, using a menu-driven interface. This allows the learner to accompany his spoken turns with movements, such as “hand-over-heart,” a gesture that accompanies greetings in some Muslim cultures. Gestures pass through the same behavior interpretation module that interprets speech, resulting in a meaning representation that captures spoken and gestured intent.

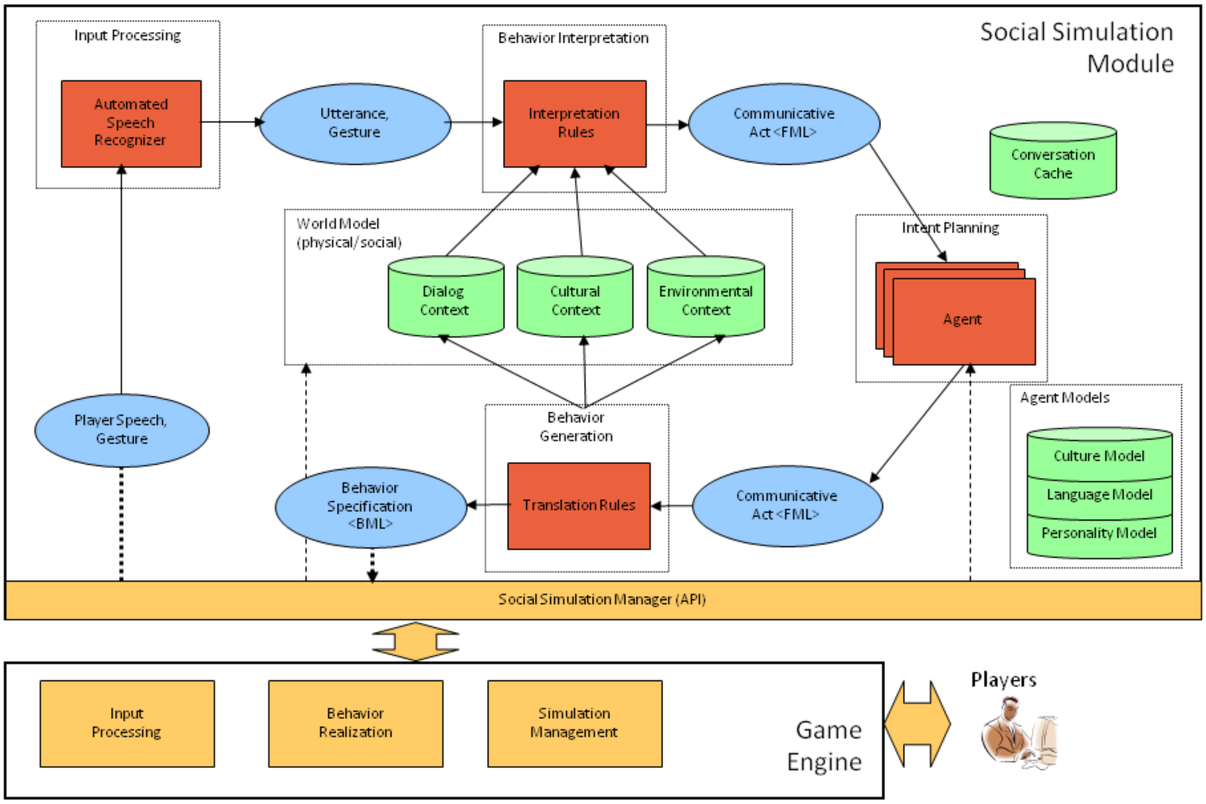

Figure 3. The VRP Social Simulation Pipeline. (© 2010, Alelo)

After the interpretation stage, an intent planning module maps the learner's intended input to an agent-specific response, and the rest of the pipeline is devoted to realizing this response in the form of speech and gestures. Speech is generally performed via pre-recorded voice files, but the mechanism is generic and can integrate speech synthesis if required.

To pass data between these stages, we use a control mechanism implemented in the expert systems programming language CLIPS[i]. The data structures, including mapping rules that determine what the output of each stage should be, are stored as instances in an ontology, edited and stored using Protégé[ii]. At runtime, the CLIPS module loads a set of rules as data, and uses them to analyze the learner input, much like a syntactic parser that loads a grammar in order to produce a parse.

This design choice supports our desiderata of re-usability, explicit models, and awareness, at the potential expense of performance (this tradeoff can be managed with some optimizations, described later). Rules and similar data structures represent explicit, composable models of agent behavior that can be reused in multiple dialogues, and by multiple agents. For example, we can consider a set of mappings from strings to meaning representations to be a model of the language that a conversational agent understands. This language understanding model is authored by subject matter experts with general knowledge of intercultural communication and expertise in the current language (e.g. Dari, Pashto, French, Arabic, and others). The same model can be loaded by a variety of agents, each of which can now "understand" the language at hand. The behavior interpretation and behavior generation modules shown in Figure 3 apply this model at run-time to determine the meaning representation for an input utterance, or to choose an utterance that expresses an intended output.

The agent itself is also an object in the ontology, which can store personal facts in addition to pointers to language and behavior models. As a result, when the dialogue manager sends learner input to an agent, that agent is able to supply his own world knowledge, composed of all of the models and facts that have been attached to him, as additional input to the intent planning module that computes the agent's reply. This contributes to the agent's awareness of the situation, resulting in consistent replies for a given agent across all learner interactions.

Behavior Interpretation

As described above, spoken learner behavior is converted into a string by an automatic speech recognition (ASR) module and stored in an object called an utterance. The meaning representation is structured as an object called a communicative act, or "act". Act structures are derived from the work of Austin (1975), using refinements published by Traum and Hinkleman (1992). Each act has a core function, i.e., the illocutionary function of the utterance (to greet, inform, request, etc.), as well as a semantic content field that holds a structured representation of the thematic role relations (Jackendoff, 1990) expressed in the utterance.

To determine which act corresponds to the current learner utterance, we rely on search through a collection of bi-directional mapping objects. At interpretation time, we search using the utterance as a key, retrieving all of the objects that map this utterance to an act. Next, a context-specific selection of a single return act is made, which is passed as the result of behavior interpretation into the intent planning stage. When no mapping is found for the current learner utterance, the behavior interpretation module returns an act called "garbage", meaning that the agent failed to understand the learner's input. The garbage act can also be used to explicitly capture predictable learner errors, ensuring that they lead to a non-understanding response from the VRP. This type of in-game feedback for the learner can be important for maintaining realism. An extended discussion of this topic is given in (Sagae, Johnson, & Bodnar, 2010).

By performing this mapping to act-level abstractions, we gain efficiency in later stages. For example, "Hi" "Hello" "Howdy" may all be interpreted as casual greetings. By mapping all of these behaviors to the same act, we enable one intent planning rule to generate a reply for all of them. In addition, we achieve a division of tasks that corresponds to a division of labor among our authoring team. Authors with language expertise can work on authoring behavior mappings, associating a variety of strings in the language being taught (Pashto, Dari, French, Arabic, etc.) with their interpretations at the act level. Intent planning rules can be authored in parallel by authors with expertise at the cultural level, who know what the agent should plan to do in a given situation, or at the task level, who may have specialized knowledge of how conversations go in a technical domain.

Intent Planning

In the intent planning module, VRP behavior is generated by a series of behavior-mapping rules that reflect the subject-matter expertise of the rule authors. A simple example of such a rule, expressed in natural language, is shown below:

IF *the learner asks about the health of your family,*
THEN *reply that your family is well*
AND *increase trust for the learner* **(1)**

Rules like (1) apply some types of knowledge implicitly in the following sense: when creating the rule, an author may be guided by an assumption like the axiom shown in (2).

In cultural setting X, *family inquiries help establish rapport* **(2)**

Without a software object where such axioms can be represented explicitly, the author is forced to keep a list of these guiding axioms in mind and to check the rules he writes for consistency with them. Since multiple authors may be involved in authoring rules, they would need to coordinate to make sure they all have a consistent view of these axioms. In contrast, the VRP architecture uses explicit models for axioms of type (2), stored as meta-rules in that can be loaded as appropriate for a given agent.

Meta-rules operate by using not only the current learner input, but a model of the dialogue history, to modify the agent's internal state. For example, the rule shown in (2) may search the dialogue history for instances of family inquiries, and increment trust whenever one is found. Without meta-rule (2), the author is responsible for adding clauses like "AND increase trust for the learner" to every family inquiry he authors. Rule (1) is only one example – there may be inquiries about the names or ages of family members, their employment status, or where they attend school. In VRPs, the author creates each of these rules independently, while the system ensures that higher-order consequences are triggered in a consistent manner based on an axiom that is only authored once, and applied everywhere.

Behavior Generation

Behavior generation in VRPs is the inverse of behavior interpretation, and it reuses the same models of language and agent knowledge. This choice supports our desideratum for reusable agents and components. However, these shared models can grow quite large, requiring search optimizations to ensure reasonable performance. The collection of mapping objects between utterances and acts is one example. In the current VRP system we are investigating indexing options that will allow us to quickly retrieve the set of mappings that apply to a given utterance (during interpretation), or to a given act (during generation), rather than searching the entire model during every learner turn at run-time.

MODELS OF DIALOGUE AND CULTURE

The architecture described above is intended to support flexible models of dialogue. In this section we describe what these models look like and how they affect VRP behavior in important ways.

Conversational agent systems in the literature employ a variety of models to generate believable behavior during interactions with users. Examples include models of compliance (Roque & Traum, 2007), stress (Hofstede & Hofstede, 2005), and cooperation (Allwood, 2001), among others. These models allow features from the social, cultural, pragmatic, and linguistic contexts of a dialogue to influence its outcome.

One example of a pragmatic-level model that has been implemented in the VRP framework is the response strategy model (Sagae, et al., 2009). In this work, a formal model is used to capture the strategic intention of a conversational agent with respect to truth values and conversational cooperation. Response strategies are implemented as an ontological layer in the VRP architecture, resulting in agents that can employ strategies like *Inform, Lie,* or *Redirect*, among others.

In addition to pragmatic models like response strategies, we have used the VRP architecture to explore socio-cultural influences on dialogue behavior. Experiments to date have included models at the macro- and micro-sociological level. In the SocialSim-MR application, macro-sociological models generated by FactionSim (Silverman, Johns, Weaver, O'Brien, & Silverman, 2002) were instantiated on a VRP platform. FactionSim is a tool that mimics actual ethnopolitical conflicts around the world. It draws on best-of-breed models from the literature in culture, cognition, personality, emotion, stress, sociology, and developmental economics, integrating them into a socio-cognitive agent framework. By connecting this framework to the real-time conversation capabilities of VRPs, the SocialSim-MR project allows real-time interpersonal interactions to be influenced by a group-level social simulation that runs in the background, based on validated social models.

In the CultureCom project, VRP behavior is driven by a microsocial model of commonsense culture, based on work by Hobbs & Gordon (2010). This model defines axioms of behavior that can be translated directly into VRP-framework rules at a meta-behavior level, similar to the example shown in (2). Both cross-cultural and culture-specific axioms have been defined that allow VRP behavior to depend on a mixture of cultural influences.

All of these models contribute to the desideratum of explicit models of appropriate behavior. One advantage of using such models is that we can validate the assumptions that drive agent behavior in isolation, before applying them in a training context. In a prototype evaluation of the CultureCom project, we collected examples of culturally appropriate dialogues that exemplify use of promises and commitments in Pashto-speaking Afghanistan. These dialogues were used to validate a model of Afghan culture, encapsulated in an ontology with rules of the type shown in (2). With the model in place, the behavior of the agents matched the behavior given in the dialogue with very high accuracy (>90% for word-level precision and recall). When the model was replaced with an American culture model (retaining the Pashto language model), the actual agent behavior diverged significantly from the behavior predicted in the Afghan dialogues. This experiment was conducted with a small data set and serves as a proof-of-concept for the ability to swap in and swap out cultural models in the VRP architecture. This is a great advantage of making models of culture explicit in the agent architecture: we now have the opportunity to express and compare the behavior of agents under different modeling assumptions.

INTEGRATION WITH IMMERSIVE GAME ENVIRONMENTS

As described in the introduction to the VRP architecture, the full social simulation pipeline, with ASR input and spoken and gestured output, is one modality of interaction between users and VRPs (text-based typing interaction is an alternative, as is menu-driven interaction). A defining feature of this modality is that it requires integration with an immersive 3-D environment. This integration depends on third-party game engines. Game engines commonly incorporate features for creating virtual worlds and populating those worlds with objects and animated characters, or avatars. Language understanding capability, including ASR, is provided by the VRP social simulation module, as shown in Figure 3.

Alelo's Operational Language and Culture Training Systems (OLCTS), shown in Figure 1 and Figure 2, integrate VRPs with a variety of game engines, some of which are single-player, and some of which are multi-player. Figure 1 shows a single-player training scenario where VRPs are integrated with the Unreal 2.5 game engine. VRP technology is packaged in a software wrapper, which can be called by the Unreal engine when the learner enters a conversation with an agent (embodied by an avatar in the simulated world). Figure 2 shows a multi-player scenario where VRP software has been deployed as a plug-in that integrates with Bohemia Interactive's VBS2. VRPs have also been integrated with Total Immersion's RealWorld.

The game engine platform provides an interface that allows the learners to choose which agent characters to talk to, and a set of graphical controls for managing the dialogue. These include buttons for starting and stopping speech input (top right corner in Figure 1) and controls for choosing nonverbal gestures for the learner's character to perform. The interface provides other information such as a summary of the objectives to complete in the scene, hints about what to say and what to do, and a transcript of the dialogue.

The game engine platform is also responsible for performing the behavior of each character. This includes speech, animated gestures, and body movement. The VRP package includes a set of pre-recorded lines for the characters to say and gesture animations for the characters to perform, which are loaded into the game engine. At runtime, the VRP social simulation module selects appropriate speech and gestures for the agents, and directs the game engine to perform them using these animations and sounds.

VRP logic is engine-independent, meaning that the same VRPs can be reused for characters in multiple game engines. In mission rehearsal engines, the trainer responsible for configuring a particular mission rehearsal scenario can assign VRPs to specific avatars. Once the assignment is made, that character will be able to hold dialogues using the VRP personality, knowledge, and decision making. In addition to the game-engine plug-in, the VRP system is deployed with a library of individual VRP characters including children and adults, men and women, leaders and followers, etc. Each character uses appropriate language and culture depending on its gender, age, etc. The desideratum of reusable agents makes this use case possible.

Virtual role-players can also be used in augmented reality or mixed reality environments in much the same way they are used with desktop game environments. Alelo is currently working on integrating VRP as part of the Infantry Immersive Trainer (IIT), a mixed reality training facility prototype for small units in Camp Pendleton, California (Lethin & Muller, 2010). The goal is to create and control a virtual Afghan role-player, who is projected on a wall inside one of the rooms in the IIT. The role-player can then hold conversations with service members who come to that room within the scope of a specific mission (e.g., census taking). A prototype of this integrated system is currently in testing, while we investigate solutions for several technical challenges. First, we are adapting the speech recognition technology used in VRP to work adequately in a noisy environment, and with audio input devices much less precise than the USB headsets normally used with desktop game engines. Second, we are working to remove the push-to-talk requirement, and move to a continuous-recognition method for input.

EXPERIENCES AND RESULTS

In this section we describe the experience for the authors and users of the VRP platform. At the time of writing, VRPs have been authored in three languages and have been integrated into training and mission rehearsal products. Our experience in delivering these products revealed interesting design issues with the VRP platform, which are described along with their solutions in this section.

The Authoring Experience

As our desiderata from previous sections show, authorability contributes directly to the overall quality and efficiency of communicative agents. Some issues we have discussed in the context of VRP authoring include the following:

Granularity of authoring rules: Non-novice learners sometimes express multiple intentions in the same turn, for example “Excuse me. I’m John. What’s your name?” The communicative act data structure used in VRPs captures one core function at a time, which supports reuse (the same object represents “I’m John” no matter what precedes or follows it) but constrains the authors’ ability to create an intent planning rule for utterances like this one, since those rules are currently structured to take one act as input before producing one or more acts as output. This constraint can be appropriate in the case of beginning language learners, since their turns are short and rarely break the single-intent assumption. It also simplifies the ASR search space, since the single-intent assumption implies that every learner turn is essentially a choice of one item from a known (although large) vocabulary of utterances. However we are currently exploring ways to break this assumption in the agent architecture, to meet the needs of intermediate learners.

Summarizing agent world knowledge: In addition to behavior rules for an agent, authors create a plain-text description that is displayed to learners who “click to know more” about a character in the game environment. This description draws on data associated with the agent in the ontology, the same data used by the agent at runtime to support his world- and self-awareness. However the world state may be large – how do authors select the most relevant pieces of knowledge as a “summary” or description of the agent? Can we support them in this process by traversing the ontology and making semi- or fully-automatic recommendations? This could contribute to more efficient and possibly more consistent authoring of agent summaries.

Multi-author access to shared libraries: VRP authoring results in libraries of reusable agents and their components. As in any software system, access and version control on the agent libraries is critical. However access to these libraries by multiple authors is also important for efficiency. Our current authoring tools rely on subversion locking mechanisms on the back end, allowing multiple authors to read an agent library, but limit write-access to one author at a time.

Access to new model features: The VRP architecture is intentionally flexible. When a new model is introduced, authoring tools must be updated. For example, when features of politeness are added to the model of language interpretation, authoring tools must be extended to allow authors to annotate the value of politeness for language samples that already exist. If a model-based solution is proposed, but authors do not have access to the new model features, the solution fails.

Supporting novice and advanced authors: As with many complex tasks, agent authoring requires tool-based support that simplifies commonly-performed actions while still allowing advanced users to access low-level details as needed. For example, assembling a new rule based on existing acts ("when the player greets, I greet in return") should be a simple, wizard-driven authoring task. In contrast, creation of new low-level components like acts ("what are the semantic and thematic roles for a formal greeting?") must be available for advanced users.

The User Experience

User feedback for the systems described here has been positive, for Alelo systems using finite-state technology as well as VRPs. The effectiveness of this approach has been documented in various published studies (e.g., Surface, et. al (2007)). The most dramatic evidence of effectiveness comes from a study conducted by the US Marine Corps Center for Lessons Learned (MCCLL) (MCCLL, 2008), which studied the experience of the 3rd Battalion, 7th Marines (3/7 Marines) in Anbar Province, Iraq in 2007.

Prior to deployment to Iraq, the battalion assigned two members of each squad of approximately thirteen marines to spend forty hours in self-study training with Alelo's Tactical Iraqi course. It should be noted that (1) forty hours is not a long time to spend learning a foreign language, (2) Arabic is a difficult language for most English speakers, and (3) self-study computer-based language learning tools often fail to produce significant learning gains. However the 3/7's experience drew attention both prior to deployment and after deployment.

In final exercises prior to deployment, Iraqi speaking role players commented that the 3/7 demonstrated Arabic language ability far beyond the skills of typical marine units preparing for deployment. During their entire tour of duty in Iraq, the 3/7 Marines did not experience a single combat casualty. The MCCLL interviewed officers in charge of the unit, and conducted surveys of the individual trainees, finding numerous instances where language and cultural skills contributed directly to mission effectiveness. Most importantly, members of the unit demonstrated an appreciation and willingness to learn about Iraqi culture, which caused the Iraqis to be more positive and cooperative, and set in motion a virtuous cycle: cooperation lead to operational effectiveness, which lead to mutual trust and further cooperation, and so on.

FUTURE RESEARCH DIRECTIONS

Some of the research questions we would like to pursue in the future with VRPs include better modeling of cultural cues, investigation of modality trade-offs in conversational agent systems, varying the levels of difficulty in agent interactions, and improving authorability.

As we observed in the related work section, VRPs can be contrasted with conversational agents that are focused on fine-grained expression of the agent's affective state. However, we recognize that cues such as gaze, posture, and facial expression can provide important information to a language learner regarding how well his attempts at communication are being received by the virtual character. For example, to show the learner that his use of formal language has a positive effect on his rapport with a local leader (played by a VRP), a current VRP system would display a series of glowing green plus symbols rising from the VRPs avatar. Similarly, a learner choice that has a negative effect results in red minus symbols. These cartoon-like effects are a placeholder for higher-quality behavior generation modules that might express reduced rapport with the learner, or more detailed affective states, by a change in the avatar's frame or stance, along the lines of (Timothy Bickmore, 2008).

A VRP is a platform-independent representation of the logic that defines an artificially-intelligent character. We can use this logic in a variety of engagements with the learner, ranging from chat-style interaction where the learner types his input and reads textual responses, to short dialogues featuring speech recognition and generation but only a fixed visual image of the VRP character, to long immersive engagements where spoken input and output are accompanied by rich body and facial movements on a 3-D avatar. Each of these modalities may be appropriate for different learning conditions. For example, text-based interaction could be helpful to learners with limited access to desktop computing hardware, while 3-D interaction may be desirable for multi-learner engagements. Because VRP technology can deploy the same agent on a variety of platforms, we now have the opportunity to perform experiments that compare these modalities.

With explicit models of dialogue behavior, as we have in VRPs, we gain the ability to tune a given agent to be more or less tolerant of violations to the assumptions made in the model. For example, we may define an agent that uses some of the meta-cultural rules of the type shown in (2), but which ignores others for the sake of being lenient on the learner. A rule that requires formal greetings when dealing with social superiors, for example, may be turned off for beginning learners and on for advanced learners. To fully address this topic, we need to understand how learner skills may vary, either from learner to learner or over time, in a single person. With a flexible architecture like VRPs, we would like to start using this understanding to match the difficulty of the conversational engagement to the needs of a particular learner at a particular point in time.

CONCLUSION

Alelo has a depth of experience in building conversational agents for education and training. Based on this experience we have described a set of desiderata for such agents, and used it to frame a new agent architecture based on highly composable models of dialogue, language, and culture. This architecture, called Virtual Role-Players, is compatible with multiple 3-D game engines, providing access to the agents in a variety of training contexts. Under this architecture we are able to set the groundwork for future research related to cultural cues, multi-media agent experiments, and variable levels of difficulty in conversational interactions. VRPs represent a unique application of conversational agent technology, not only to teach, but to demonstrate communicative skill for learners who hope to acquire it.

ACKNOWLEDGMENTS

The authors would like to thank the members of the Alelo team who contributed to this work. This work was sponsored by PM TRASYS, Voice of America, the Office of Naval Research, and DARPA. Opinions expressed here are those of the authors and not of the sponsors or the US Government.

REFERENCES

Allwood, J. (2001). Cooperation and Flexibility in Multimodal Communication. In H. Bunt & R.-J. Beun (Eds.), *Cooperative Multimodal Communication*. Berlin/Heidelbert: Springer Verlag.

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

Bickmore, T. (2008). *Framing and Interpersonal Stance in Relational Agents.* Paper presented at the The Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Workshop on Functional Markup Language, Estoril, Portugal.

Bickmore, T., & Pfeifer, L. (2008). *Relational Agents for Antipsychotic Mdication Adherence*. Paper presented at the CHI 2008 Workshop on Technology in Mental Health.

Blumer, H. (1986). *Symbolic Interactionism: Perspective and Method* Berkeley: University of California Press.

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

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

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Goffman, E. (1990). *The Presentation of Self in Everyday Life*. New York, NY: Doubleday.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education, 48*, 612-618.

Habermas, J. (1984). *The Theory of Communicative Action*. Boston: Beacon Press.

Hobbs, J. R., & Gordon, A. (2010). *Goals in a Formal Theory of Commonsense Psychology.* Paper presented at the the 6th International Conference on Formal Ontology in Information Systems.

Hofstede, G., & Hofstede, G. J. (2005). *Cultures and organizations: Software of the mind*. New York: McGraw-Hill.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, Mass MIT Press.

Johnson, W. L. (in press). Serious Use of a Serious Game for Learning Foreign Language. *Journal of Artificial Intelligence in Education, 20*.

Johnson, W. L., & Valente, A. (2008). *Tactical Language and Culture Training Systems: Using Artificial Intelligence to Teach Foreign Languages and Cultures.* Paper presented at the IAAI 2008.

Johnson, W. L., & Valente, A. (2009). Tactical Language and Culture Training Systems: Using AI to Teach Foreign Languages and Cultures. *AI Magazine, 30*(2), 72-84.

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

Lampe, J. (2007). *Cultural Proficiency Guildelines (3.2).* Paper presented at the Plenary session of the Interagency Language Roundtable for the ACTFL Arabic Testing Consensus Project.

Lethin, C., & Muller, P. (2010). Future immersive training. *Warfighter Enhancement Activities, 4*(1), 2-4.

Marsella, S., & Gratch, J. (2009). EMA: A Model of Emotional Dynamics. *Journal of Cognitive Systems Research, 10*(1), 70-90.

MCCLL (2008). Tactical Iraqi Language and Culture Training System. *MCCLL 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*: The Regional and Cultural Competence Assessment Working Group (RCCAWG), Defense Equal Opportunity Management Institute.

McRorie, M., Sneddon, I., Sevin, E. d., Bevacqua, E., & Pelachaud, C. (2009). *A Model of Personality and Emotional Traits.* Paper presented at the International conference on Intelligent virtual agents IVA'09, Amsterdam.

Mead, G. H. (1967). *Mind, Self, and Society; from the Standpoint of a Social Behaviorist*. Chicago, Ill: University of Chicago Press.

Melo, C. M. d., & Gratch, J. (2009). *Expression of emotions using wrinkles, blushing, sweating, and tears.* Paper presented at the Intelligent Virtual Agents, Amsterdam.

Roque, A., & Traum, D. (2007). *A Model of Compliance and Emotion for Potentially Adversarial Dialogue Agents.* Paper presented at the the 8th SIGdial Workshop on Discourse and Dialogue, Antwerp, Belgium.

Rose, C. P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., & Weinstein, A. (2001). *Interactive Conceptual Tutoring in Atlas-Andes*. Paper presented at the AI in Education 2001.

Rosis, F. d., Pelachaud, C., Poggi, I., Carofiglio, V., & Carolis, B. D. (2003). From greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *Int. J. Hum.-Comput. Stud., 59*(81-118).

Sagae, A., Johnson, W. L., & Bodnar, S. (2010). *Validation of a Dialog System for Language Learners.* Paper presented at the The 11th annual SIGdial Meeting on Discourse and Dialogue, Tokyo, Japan.

Sagae, A., Wetzel, B., Valente, A., & Johnson, W. L. (2009). *Culture-Driven Response Strategies for Virtual Human Behavior in Training Systems.* Paper presented at the SLaTE-2009, Warwickshire, England.

Samtani, P., Valente, A., & Johnson, W. L. (2008). *Applying the SAIBA framework to the Tactical Language and Culture Training System*. Paper presented at the AAMAS 2008 Workshop on Functional Markup Language (FML).

Schutz, A. (1970). Alfred Schutz on Phenomenology and Social Relations. , ed. In H. R. Wagner (Ed.), *Selected Writings*. Chicago: University of Chicago Press.

Silverman, B. G., Johns, M., Weaver, R., O'Brien, K., & Silverman, R. (2002). Human Behavior Models for Game-Theoretic Agents. *Cognitive Science Quarterly, 2*(3-4, Fall 2002), 273-301.

Surface, E. A., Dierdorff, E. C., & Watson, A. M. (2007). *Special Operations Language Training Software Measurement of Effectivenss Study: Tactical Iraqi Study Final Report*: SWA Consulting.

Traum, D., Marsella, S., Gratch, J., Lee, J., & Hartholt, A. (2008). *Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents.* Paper presented at the the 8th International Conference on Intelligent Virtual Agents.

Traum, D. R., & Hinkelman, E. A. (1992). Conversation Acts in Task-Oriented Spoken Dialogue. *Computational Intelligence, 8*(3), 575--599.

Valente, A., Johnson, W. L., Wertheim, S., Barrett, K., Flowers, M., LaBore, K., et al. (2009). A Dynamic Methodology for Developing Situated Culture Training Content. Alelo TLT, LLC.

Vilhjalmsson, H., & Marsella, S. (2005). *Social Performance Framework.* Paper presented at the the AAAI Workshop on Modular Construction of Human-Like Intelligence.

Vilhjalmsson, H., & Thorisson, K. R. (2008). *A brief history of function representation from Gandalf to SAIBA.* Paper presented at the The Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Workshop on Functional Markup Language, Estoril, Portugal.

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, February 2008*.

Weber, M. (1994). *Sociological Writings*. New York: Continuum.

ADDITIONAL READING

Arvizu, S., & Saravia-Shore, M. (1990). Cross-Cultural Literacy, An Anthropological Approach to Dealing with Diversity. *Education and Urban Society, 22*(4), 364-376.

Eysenck, H. J. (1976). *The Measurement of Personality*. Lancaster: Medical and Technical Publishers.

Hernandez, A., Lopez, B., Pardo, D., Santos, R., Hernandez, L., Relano, J., et al. (2008). *Modular definition of multimodal ECA communication acts to improve dialogue robustness and depth of intention.* Paper presented at the The Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Workshop on Functional Markup Language, Estoril, Portugal.

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., & Sagae, A. (2010a). *Adaptive Agent-Based Learning of Intercultural Communication Skills*. Paper presented at the Demonstration sessions of the User Modeling and Personalization Conference (UMAP 2010).

Johnson, W. L., & Sagae, A. (2010b). *Adaptive Agents for Promoting Intercultural Skills.* Paper presented at the Adaptation and Personalization in E/B-Learning Using Pedagogic Conversational Agents (APLeC), workshop of UMAP 2010, Big Island, HI, USA.

Johnson, W. L., & Valente, A. (2008). *Collaborative Authoring of Serious Games for Language and Culture*. Paper presented at the SimTecT 2008.

Kumar, R., Sagae, A., & Johnson, W. L. (2009, July 2009). *Evaluating an Authoring Tool for Mini-Dialogs.* Paper presented at the AIED 2009 (Poster), Brighton, UK.

Lazarus, R. (2000). *Emotion and Adaptation*. New York, NY, USA: Oxford University Press.

Mascarenhas, S., Dias, J., Afonso, N., Enz, S., & Paiva, A. (2009). *Using rituals to express cultural differences in synthetic characters*. Paper presented at the AAMAS 2009.

Matsuda, N., Cohen, W. W., Koedinger, K. R., Stylianides, G., Keiser, V., & Raizada, R. (2010). *Tuning Cognitive Tutors into a Platform for Learning-by-Teaching with SimStudent Technology*. Paper presented at the the International Workshop on Adaptation and Personalization in E-B/Learning using Pedagogic Conversational Agents (APLeC), Kona, HI.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology, 46*, 961-978.

Muller, P., CDR Dylan Schmorrow, U., & Buscemi, T. (2008). The infantry immersion trainer: today's holodeck. *Marine Corps Gazette, September 2008*.

Traum, D., Marsella, S., Gratch, J., Lee, J., & Hartholt, A. (2008). *Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents.* Paper presented at the the 8th International Conference on Intelligent Virtual Agents.

Wang, N., & Johnson, W. L. (2008). *The Politeness Effect in an Intelligent Foreign Language Tutoring System*. Paper presented at the the 9th International Conference on Intelligent Tutoring Systems.

KEY TERMS & DEFINITIONS

Intercultural competence: general skills for handling intercultural situations; in this work we focus on communicative knowledge, skills, and attitudes that apply in situations requiring interpersonal communication across cultures

Cultural environment: a setting for interpersonal communication that includes a variety of factors, including the following general categories: physical environment; political, economic, and social structures common within the culture; and typical personal perspectives (e.g. attitudes toward time, personal relations, work, and the role of the individual within the community)

Role-player: a human acting as a character in a live simulation, as in military wargaming; Virtual Role-Players are conversational agents in a computer-based simulation

Immersive environment: a setting for study that is an engaging model of the environment where the learner will use the skills being taught; in this work we use “immersive” to describe the 3-D game where learners operate in a simulated world, and where all characters speak only in the language being studied

Game engine: the software responsible for rendering the simulated environment where conversational agents operate; examples presented in this work include VBS2, RealWorld, and Unreal

Training scenario: an engagement with conversational agents in a virtual environment that is designed to train the human participant in a set of skills; in particular, a task-specific engagement that simulates a situation the user expects to face in real life

Scene: one stage of a training scenario, perhaps involving a fixed set of characters and sub-tasks

Avatar: the visual representation of a conversational agent, often an animated 3-D rendering of the character being played by the agent in a simulated world

Authorability: the ease with which behavior of a conversational agent can be created, in particular by subject matter experts who may be non-programmers; this is a measure of how mature the agent technology is, including authoring tools

---

[i] http://clipsrules.sourceforge.net/
[ii] http://protege.stanford.edu/