Walking the walk/Talking the talk: Mission Planning with Speech-Interactive Agents

Benjamin Bell[1]; Philip Short[2]; Stewart Webb[2]
[1]CHI Systems, Inc; [2]Ael Ltd
bbell@chisystems.com; phil.short@baesystems.com; stewart.webb@baesystems.com

Abstract. The application of simulation technology to mission planning and rehearsal has enabled realistic overhead 2-D and immersive 3-D "fly-through" capabilities that can help better prepare tactical teams for conducting missions in unfamiliar locales. For aircrews, detailed terrain data can offer a preview of the relevant landmarks and hazards, and threat models can provide a comprehensive glimpse of potential hot zones and safety corridors. A further extension of the utility of such planning and rehearsal techniques would allow users to perform the radio communications planned for a mission; that is, the air-ground coordination that is critical to the success of missions such as close air support (CAS). Such practice opportunities, while valuable, are limited by the inescapable scarcity of complete mission teams to gather in space and time during planning and rehearsal cycles. Moreoever, using simulated comms with synthetic entities, despite the substantial training and cost benefits, remains an elusive objective. In this paper we report on a solution to this gap that incorporates "synthetic teammates" – intelligent software agents that can role-play entities in a mission scenario and that can communicate in spoken language with users. We employ a fielded mission planning and rehearsal tool so that our focus remains on the experimental objectives of the research rather than on developing a testbed from scratch. Use of this planning tool also helps to validate the approach in an operational system. The result is a demonstration of a mission rehearsal tool that allows aircrew users to not only fly the mission but also practice the verbal communications with air control agencies and tactical controllers on the ground. This work will be presented in a CAS mission planning example but has broad applicability across weapons systems, missions and tactical force compositions.

1. MISSION PLANNING, REHEARSAL GAPS

Mission planning and mission rehearsal are routinely performed using sophisticated automation and simulation technology. Planners, commanders and their personnel are now able to "fly-through" a mission, employing threat models and advanced visualization tools that can render accurate geospatial and terrain data. Such realistic simulations help prepare tactical teams for conducting missions in unfamiliar locales. For instance, detailed terrain data can prepare aircrew to recognize relevant landmarks and hazards, and threat models can provide a comprehensive glimpse of potential hot zones and safety corridors.

There is one aspect of mission performance that is critical to success which has remained beyond the reach of even the most advanced mission planning tools: verbal communication. Missions such as close air support (CAS) depend heavily on timely, succinct, correct and relevant spoken dialogue between air and ground elements. Joint Terminal Attack Controllers (JTACs) and CAS-rated aircrew typically train on live ranges to reach some criterial performance level. But once deployed, practice opportunities are severely limited by the inescapable scarcity of complete mission teams to gather in space and time during planning and rehearsal cycles.

2. POTENTIAL SOLUTIONS

Mission planning and rehearsal should allow users to practice the radio communication along with the other aspects of mission performance. In CAS, for instance, the air-ground coordination is critical to the success and safety of the mission and should be represented in walk-through/fly-through activities. Unfortunately this is seldom the practice, due largely to the separation in time and space of the respective staffs in the air and ground elements planning and rehearsing the mission.

In general there are two constructs for meeting this gap: (1) use of live confederates as role-players; and (2) software simulations of entities in the scenario.

2.1 The "Wetware" Option

Option 1 is the use of live personnel and requires no sophisticated technology. But there are cost and access penalties incurred by the use of live role-players:

1. When participants are drawn from the trainee ranks their time is spent on providing cues to keep the scenario moving rather than on effective mission rehearsal;
2. When drawn from the instructor ranks, role-playing interferes with performance

assessment, since instructors are called upon to divide their attention between evaluating mission success and role-play;

3. It creates variability that makes standardizing rehearsal difficult due to the human element influencing events in each scenario.
4. Costs arise from compensating, transporting and lodging role-players at dedicated facilities.
5. Availability is compromised because expert role-players can be exceedingly difficult to arrange, particularly for missions in new areas of operation or that employ novel tactics or recent equipment changes.

The consequence is that access to mission planning and rehearsal is measured and scheduled and conducted at specific facilities.

2.2 The Software Option

Option 2 is to employ software simulations of entities in the scenario in lieu of live role-players. We are exploring this option by introducing intelligent, interactive agents into a mobile mission planning package. We commenced this investigation by defining the core capabilities needed for synthetic teammates. To provide interaction effectively for mission planning and rehearsal, our analysis revealed that synthetic teammates must possess the following capabilities:

1. simultaneous execution of: taskwork (e.g., flying the aircraft, working the console); teamwork (interacting with other members of the team); and measurement (for subsequent analysis and feedback);
2. interaction via spoken language (required for rehearsing mission communications);
3. modulating behaviors to replicate various error modes, to allow for varying the proficiency of the synthetic team members (important for playing out contingencies and stress-testing the plan).

We expect that the above generic requirements extend well beyond conventional computer-generated forces (CGFs), semi-automated forces (SAFs), and game-based artificial intelligence, or "AI"s – largely scripted entities with limited abilities to respond to events beyond a predefined range of simple behaviors. CGF/SAF technologies do have an important role to play, but for our purposes they fall short of addressing specific needs that remain unmet. To meet these needs, we are employing cognitive modeling using CHI Systems' computational development tool, iGEN®, for encapsulating human expertise and behavior in synthetic agents (Zachary, LeMentec & Ryder, 1996). Sophisticated agents, such as those which may be built using iGEN, can provide dialogue-capable synthetic teammates to reduce reliance on human role-players and make mission planning and rehearsal more accessible, less costly, and more standardized.

2.3 Previous Work: On-Demand Team Training

Mission planning and rehearsal each share a simulation dimension with training, where this technique has received the most attention. We first integrated the cognitive modeling approach with full speech interaction for a US Navy program called Synthetic Cognition for Operational Team Training (SCOTT) (Zachary, *et al.*, 2001). SCOTT is a simulation-based practice and training environment in which a single human crewmember of an E-2C tactical crew can train in cross-platform coordination skills by interacting verbally with synthetic teammates, both on and off the E-2C. More recently, we developed Synthetic Teammates for Realtime Anywhere Training and Assessment (STRATA), a Close Air Support (CAS) trainer built on the progress made under SCOTT but using more sophisticated cognitive modeling and more advanced speech technologies (Bell, Johnston, Freeman & Rody, 2004). The emphasis in STRATA was to validate "on-demand team training" by making the instructor and the other CAS team members, such as the Forward Air Controller, entirely optional. Most recently, we developed the Virtual Interactive Pattern Environment and Radiocomms Simulator (VIPERS). VIPERS offers users opportunities for guided practice and feedback in radio communications skills and decision making in a simulated pattern environment (Bell, Ryder & Pratt, 2008). The format of this practice is simulation-based training with intelligent software agents performing in both tutoring roles and synthetic teammate roles, in a laptop-based portable application for anytime/anywhere training. Specifically, VIPERS provides three types of speech-interactive entities: (1) a synthetic instructor that provides coaching and feedback during scenarios and makes assessments to be used in a debrief; (2) a synthetic controller that maintains knowledge of all aircraft in the pattern and verbally responds to clearance requests and issues directives to all aircraft in the pattern; and (3) synthetic pilots/aircraft in the pattern behaving appropriately and making radio calls.

3. CAS MISSION PLANNING & REHEARSAL: AN EXPLORATORY STUDY

Work reported in this paper was aimed at applying some of the capabilities we had developed in the training domain to explore more realistic and more accessible mission planning and rehearsal tools. Our focus was on users in high OPTEMPO contexts, engaged in missions requiring a great deal of teamwork. We looked particularly at cases where teams are distributed and where verbal communication enjoys a key role in mission coordination, selecting CAS for this study. To accelerate our research, we employed a fielded mission planning and rehearsal tool, so that we could devote our attention to investigating the utility of speech-interactive synthetic teammates rather than on creating a suitable testbed. The tool we employed is called the Combined Arms Gateway Environment (CAGE).

3.1 Summary of CAGE

CAGE, developed by AeI, is a mission support tool that enables operators to plan, rehearse and then conduct platform specific or independent missions under a wide variety of operational conditions. The system can be configured to support the operational needs of any given operator or platform configuration. CAGE is based on an open architecture JAVA framework.

CAGE allows planners to employ the rehearsal capability to create routes, inspect and deconflict airspace, view corridors and define threat cones. Planners and mission personnel can view the mission in 2-D (top-down) and 3-D. The 3-D view provides dynamic lighting (sun, shade, moonlight) to assess the tactical implications of time of day and visibility effects (fog, haze, cloudbase) to project the visibility under the forecast weather conditions.

3.2 A Human-Centric Approach

Our analysis started with a human factors integration approach by considering what features would be required of a speech interactive agent for training, planning and rehearsal; and what the associated benefits were. This was to ensure that the use of such technology was driven by the needs of the warfighter, rather being implemented as a technology push simply for its own sake. This analysis yielded four required characteristics of a speech-interactive agents:

1. Real-time – includes element of time pressure on decision making and actions;

2. Unpredictable – able to include unanticipated / unexpected events;

3. Dynamic – able to respond to user actions;

4. Replicates the modality of real dialogue – user must process information in same way (e.g., cannot simply read prompts from a screen).

We also identified the following anticipated benefits:

1. Reduced instructor input – elements of automation mean that multiple users can train or rehearse concurrently on multiple systems, without the need for multiple instructors or mission commanders.

2. Increased combat readiness – availability of multiple, less costly systems reduces reliance on expensive, scarce simulators;

3. Reduced flying / simulator hours – system enables training that was previously only possible in the air or on a full mission simulator;

4. Reduced cost - as a result all of the above.

3.3 Needs Analysis

A high-level needs analysis was performed for a CAS scenario. This was a limited analysis, in alignment with the exploratory nature of this research, and so was focused specifically on voice interaction. This entailed performing a Hierarchical Task Analysis (HTA) for the scenario, and reviewing each relevant step[1] to identify:

- The objective for that step.
- How to gauge that the objective has been achieved, *i.e.* the measure of effectiveness (MoE);
- The required inputs for that step (what the instructor has to include over and above the synthetic agent component in order to accomplish the step);
- The specific benefits that the synthetic agent provides, which would not have been achieved by other means (*e.g.* by displaying the dialogue as text on a screen);
- What the technology must be able to do in order to provide the required benefit.

[1] By 'relevant step' we mean those steps that involve the user doing something, as the HTA also covers the actions of the Joint Terminal Attack Controller (*i.e.* the actor being 'played' by the synthetic agent).

The results of the HTA were captured against the following criteria (example outcomes shown in parentheses):

- Task: (*Look for described area and features*).
- Objective: (*Rapidly and accurately identify areas based on description of the visual scene*).
- MoE: (*Identify target within elapsed time parameters*).
- Required inputs: (*A representation of the visual scene that relates to the descriptions being provided*).
- Benefit: (*Synthetic agent allows natural interaction between user and JTAC, with correct sensory input (auditory) and output (speech)).*
- Requirement for agent: *(able to provide descriptions that relate to the visual scene provided).*

3.4 Technical Approach

To bound the scope of our initial experiment, we created a set of CAS scenarios, focusing on dialogue between the pilot and JTAC, allowing for alternative dialogue branches and error correction. The complexity of the scenarios determines the necessary sophistication of the grammar, synthesized voice, and agent model. For this exploratory effort, therefore, the scenarios were limited to specific phases of a representative CAS mission.

In order to efficiently introduce voice capabilities to CAGE, a TCP socket-based network protocol using XML-based messaging was employed to enable communication between CAGE and our existing speech-enhanced synthetic agent framework. The XML schema was directly derived from the High Level Architecture (HLA) interactions used in some of our previous work (e.g., Chapman, Ryder, Bell, Wischusen & Benton, 2004). A network-based API was chosen based on direct routine calls as this approach involved minimal modification of the existing systems, each of which was able to retain its modes of operation, largely independent of the others. CAGE is responsible for loading the dialogue information (in the form of an XML file) and sharing that information with the agent framework as the scenario progresses. Data is shared at a semantic level. Position data, for instance, is shared to allow the agent component to generate the appropriate synthetic speech. CAGE determines the pace of the exercise by sending the information at the desired time intervals based on user responses and progress through the mission.

The speech-enabled agent framework consists of modules to synthesize and recognize speech, an agent implementation to respond appropriately, and a messaging framework called the Socket Executive to mediate communication among the modules and with CAGE over TCP (see Figure 1).

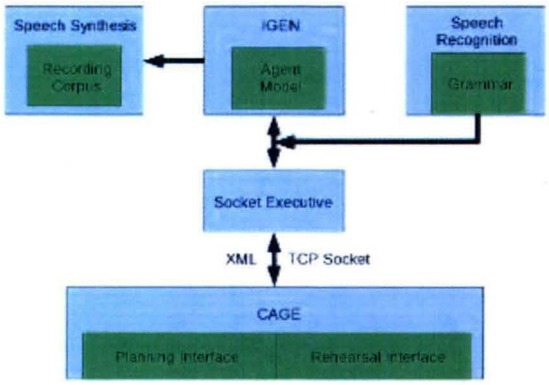

Figure 1: Architecture integrating CAGE, iGEN

We built an iGEN model to play the role of the ground-based observer - the Joint Terminal Attack Controller (JTAC) - and exported it to XML. The speech dialogues from the scenarios were divided into phrases spoken by the user and those spoken by the synthetic teammate. Speech synthesis was accomplished using the Festival speech synthesis engine (Black & Taylor, 1997) and Limited Domain Synthesis (LDOM) (Black & Lenzo, 2000), which uses phonemes derived from recordings to synthesize speech. This approach provides speech that is tactically realistic and based on concatenated recordings of domain experts; but also preserves the capability to dynamically generate speech in real-time, voicing variables such as coordinates, call signs and mission times.

One requirement of the LDOM approach is that recorded samples be collected for any lexical token in the vocabulary. This is a minimal requirement since word pair, tuples and longer phrases are permitted as well. We enhanced realism by recording phrase variants similar to those expected during mission planning and rehearsal. By carefully examining the dialogues, and constructing phrases covering the expected vocabulary including all possible numerals, call-signs, and directions, a corpus of phrases was created and then recorded by a domain expert. The recordings were volume normalized, broken into phonemes, and indexed for use by the Festival engine at runtime. Additional recordings made to accommodate revisions to the vocabulary were incorporated into the previous

corpus. Pauses were inserted into some of the communications (e.g., reading coordinates) to more realistically capture the manner in which such phrases are spoken operationally.

The user-spoken dialogue was represented semantically as a string identifier followed by any variables (e.g., coordinates, directions) optional words or phrases, and modifiers (such as "not"). These forms were coded into a Backus-Naur Form (BNF) grammar for the speech recognition engine. These semantic representations were also encoded into the grammar so that when a user-spoken phrase is recognized, the engine would return not only a plain-text representation but also the semantic frame with optional parameters included. This enabled the agent to more easily understand the recognized speech.

The socket executive uses a publish-subscribe mechanism to distribute information among the synthetic teammates, and constructs XML messages for communicating with CAGE over TCP. The synthetic teammate framework and components were built with native C++ code.

4. RESULTS

4.1 Scenario Definition

The implemented scenario demonstrates a mission rehearsal with a user in the role of lead CAS pilot, interacting with a synthetic JTAC. When a scenario is started, the components load their required data (CAGE loads its scenario data, the speech components load the grammar and voice data, and iGEN loads the JTAC model) and each initializes the appropriate communication channels. The user selects a call sign from a set of nominal identifiers and two-digit suffixes. The user then begins the mission and initiates communication by checking in with the chosen call-sign. Figure 2 shows a representative display at this point in the mission, with a 3-D view on the left and the 2-D view on the right.

Figure 2: Representative display in CAGE

The JTAC agent transmits a 9-line brief, based on information given to it by CAGE (the user can request a re-transmit at any point during the mission). The user then repeats the 9-line and the read-back is checked by the synthetic JTAC for accuracy. If an error is found in the readback, the user is asked to repeat any incorrect portions of the communication until it is correct (and only the incorrect portions). The JTAC agent then directs the user to the target, who must read back the targeting information, which is again checked for accuracy. Following an accurate read-back, the JTAC clears the user for attack. After attack the JTAC responds with a battle damage assessment, and the user signs off. During each exchange the JTAC waits for the appropriate response from the user, and asks the user to repeat any communication that is incorrect or unrecognizable. A representative transcript is shown in Figure 3.

User: Widow 76 this is Vader 28 checking in as fragged
JTAC: Vader 28, Widow 76 Loud and clear, this is a Type 1 control, call ready to copy.
User: Vader 28 Type 1 control, ready copy
JTAC: IP U278, Heading 055 magnetic, Distance 9260 meters, Elevation 70 feet. Target is a Helicopter parked on western edge of dispersal. Location North 51 00.89 West 002 38.01. Mark Laser 1111 LTL 355 Magnetic. Friendlies 1000 South, Egress North to Bad Wolf. Advise when ready for remarks
User: Ready to copy remarks
JTAC: Final attack heading 055 through 030
User: Elevation 70 feet , Location North 51 00.89 West 002 38.01. Friendlies 1km South. Laser 1111 LTL 355 magnetic. Attack heading 055 through 030 magnetic
JTAC: Readback correct, report leaving IP
User: Leaving IP, abort alfa romeo sierra
JTAC: Widow 76, abort alfa, romeo, sierra your target is one of 2 helicopters on the western edge of a dispersal.
User: Helicopter, western edge, dispersal. Vader 28 leaving IP.
JTAC: Short of target, airfield
User: Short of target, airfield
JTAC: North of runways, group of 8 hangars. From there, 12 o'clock 500, further set of 3 hangars, North East corner airfield. Laser on. Friendlies to South of all runways.
User: Contact 10 seconds. Further 3 hangars Laser on. Visual friendlies
JTAC: Right of hangars is large dispersal, in sunlight, target is helicopter on right hand side
User: Contact Target, left of target further helicopter against building.
JTAC: Affirm, cleared hot
User: In hot. Rifle away. Terminate
JTAC: Terminate, Vader 28, widow 76, Delta Hotel, helicopter destroyed, End of mission.
User: Target destroyed, Delta Hotel, End of Mission.

Figure 3: Representative dialogue between aircraft (user) and JTAC agent

4.2 Synthetic Teammate Interactions

An important design consideration is the degree of variability in whether user utterances are treated as "legal". Too restrictive an approach erroneously emphasizes syntax over semantics, frustrates users, and undermines mission planning and rehearsal objectives. Too accommodating an approach not only adds complexity to the recognition process but could introduce non-doctrinal phraseology.

There is no quick-fix solution; striking a proper balance depends on thoughtful, comprehensive consultations with subject matter experts, guided by a cognitive task analysis methodology (e.g., Zachary, Ryder & Hicinbothom, 2000). For our exploratory study we employed a CAS-rated RAF pilot and implemented logic in the JTAC agent that permits lexical and syntactic variations based on the tactical context. Each communication spoken by the user can thus be phrased in different ways; optional wording can be omitted and some alternate wordings are accepted.

This flexible grammar, combined with the selective requests for read-back (i.e., only incorrect portions of the 9-line need be repeated) afford a transparent dialogue capability. The work reported here was speaker-independent – no training to a specific voice was required. Our testers consisted of both U.K. and U.S. speakers with no noticeable differences in recognition rates among them.

Initial results showed that there was an immediate benefit to being able to practice techniques as they would be performed for real while remaining in a benign environment. For early-stage training, this removes the stress of the real situation in order to put the trainee at ease; for planning and rehearsal the realism is sufficient to provide the necessary situational awareness to adequately exercise the plan and measure an individual's performance in executing it.

Early feedback from end-users also indicates the scaleability of this technology. There is significant potential to increase the richness of the training experience, including using the synthetic agent to increase the user's exposure to operational stress; to augment the simulated environment with more diverse players and to provide voice interaction in situations where it is not currently available.

4.3 Broader Implications

The investigation reported here provides preliminary support for the utility of speech-interactive synthetic teammates in the mission planning and rehearsal domain. We recognize that our results are based on a limited scenario, and we are currently planning to develop more comprehensive, complex scenarios, which will require behavioral, speech and grammar components with additional sophistication.

To achieve the performance reported here in richer scenarios, we require more robust speech recognition and discourse management. We will address this by employing a dynamic grammar, where an intelligent agent activates and de-activates sub-grammars as the tactical situation changes, an approach we have reported in previous work (Bell, Johnston, Freeman & Rody, 2004). Our work has indicated that there is significant training benefit to be gained from using speech interactive agents through increased richness or improved efficiency of the training environment (Bell, Ryder & Pratt, 2008).

We are also expanding the reach of this approach through integration of the capabilities reported here with a more sophisticated testbed called the Distributed Synthetic Air Land Training (DSALT) facility operated by the UK MOD. Results from that experiment will provide a firmer foundation for assessing the utility of speech-capable synthetic teammates for training, mission planning and rehearsal.

5. CONCLUSION

New simulation capabilities that extend the benefits of synthetic training can yield parallel advances in mission rehearsal and mission planning. For missions that rely on effective communication and coordination, though, the verbal exchange among tactical teammates is trained, planned and rehearsed only if and when suitable role-players are available, co-located in time and place.

In this paper we have introduced speech-interactive synthetic teammates as a capability that overcomes these limitations and provides on-demand team simulation. Using CAS as a mission representative of the need for effective tactical communication and coordination, we present a mission planning and rehearsal system that is augmented with a synthetic JTAC agent. This added capability allows commanders and aircrew to plan and fly through a CAS mission while communicating verbally with the synthetic JTAC.

By employing the knowledge encapsulated in an intelligent agent, we can overcome many of the challenges faced in human-computer dialogue, and continue to enrich synthetic training while migrating the benefits of this approach into the realms of mission planning and rehearsal.

REFERENCES

Bell, B., Johnston, J., Freeman, J., & Rody F. (2004). STRATA: DARWARS for Deployable, On-Demand Aircrew Training. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (I/ITSEC), December, 2004.

Bell, B., Ryder, J., and Pratt, S. (2008). *Communications and coordination training with speech-interactive synthetic teammates: A design and evaluation case study.* In D. Vincenzi, J. Wise, P. Hancock and M. Mouloua (Eds.), Human Factors in Simulation and Training. CRC Press. Boca Raton.

Black, A.W. and Taylor, P.A. (1997). The Festival Speech Synthesis System: System documentation. Technical Report HCRC/TR-83, Human Communication Research Centre, University of Edinburgh, Scotland, UK, 1997.

Black, A.W. and Lenzo, K. (2000). Limited Domain Synthesis. In *Proceedings of the Sixth International Conference on Spoken Language Processing* (ICSLP 2000). Beijing, China October 16-20, 2000.

Chapman, R. J., Ryder, J., Bell, B., Wischusen, D. & Benton, D. (2004) STRATA (Synthetic Teammates for Real-time Anywhere Training and Assessment): An integration of cognitive models and virtual environments for scenario based training. In *Proc. of Human Factors & Ergonomics Society 49th Annual Mtg.* October, 2004, New Orleans, LA.

Zachary, W., LeMentec, J.C., & Ryder, J. (1996) Interface agents in complex systems. In C. Ntuen & E. Park (Eds.), *Human Interaction with Complex Systems: Conceptual Principles and Design Practice.* Norwell, MA:Kluwer Academic Publishers, pp 35-52.

Zachary, W., Ryder, J. M., and Hicinbothom, J. H. (2000). Building Cognitive Task Analyses and Models of a Decision-Making Team in a Complex Real-Time Environment. In Chipman, Shalin & Schraagen, Eds. *Cognitive Task Analysis.* New Jersey: Lawrence Erlbaum, 2000. 365-384.

Zachary, W. Santarelli, T., Lyons, D., Bergondy, M. and Johnston, J. (2001). Using a Community of Intelligent Synthetic Entities to Support Operational Team Training. In *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation.* Orlando: Institute for Simulation and Training. pp: 215-224.