



Vocal Command and Control of SAF Entities

Overview:

This project allows a human user to give spoken commands to Computer Generated Forces (CGF). Vocal commands are translated to natural language text and commands recognized using a keyword recognition technique. The recognized commands are translated again to a format that the Dismounted Infantry Semi-Automated Forces (DISAF) system can accept. The system demonstrated a basic capability that allowed use with a minimum of train-up required. The project was sponsored by STRICOM.

The Voice Federate (VF) system is composed of three major parts, as shown in Figure 1. The

main components are the Voice Recognition Federate (VRF), the Voice and Gesture recognition federate to DISAF interface (VGD), and DISAF.

The VRF interfaces with a human participant in the simulation by means of a microphone and headset. The VRF performs the speech-to-text (STT) and text-to-speech (TTS) functions. Both the STT and TTS functions are performed using Commercial Off The Shelf (COTS) products.

The VRF also maps text generated by the COTS STT software to a legal VF command. This command is then sent to the VGD using the High Level Architecture (HLA) communication services of the Run Time Infrastructure (RTI).

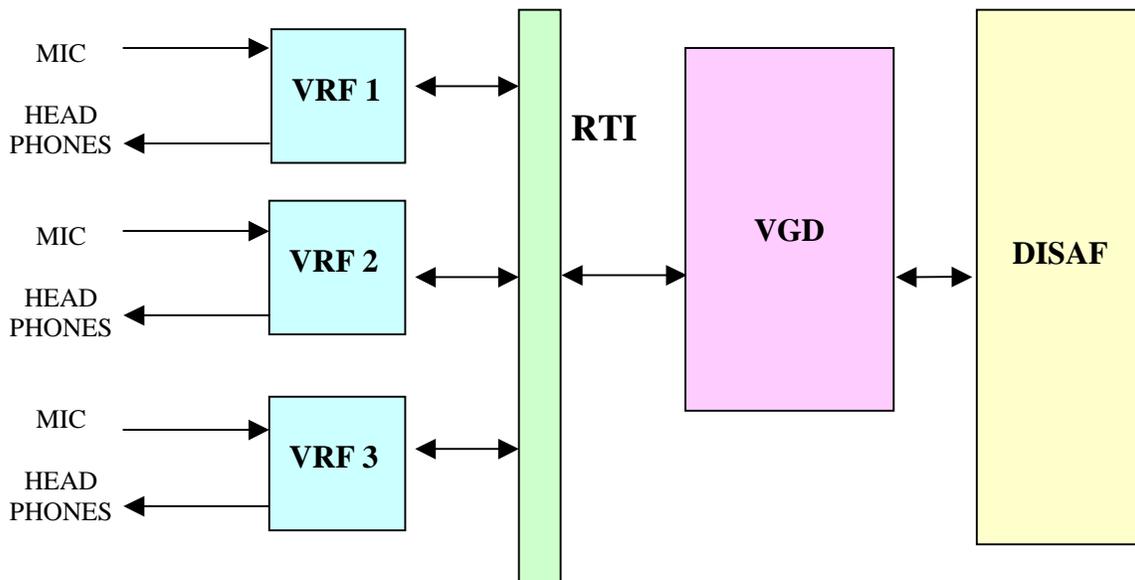


Figure 1 – Voice Interaction with SAF Main Components

executive program.

The VGD communicates between the VRF's using an HLA compliant protocol and communicates to DISAF using a TCP/IP compliant protocol. The VGD registers VRF's as they come online to route synthetic responses to the appropriate human user.

Although the VF can technically interface with any simulation able to communicate via TCP/IP the current VF operates with DISAF. IST has modified DISAF to allow the initiation of tasks commanded vocally and also to allow the generation of synthetic speech.

Accomplishment:

During the three years of the STRICOM Individual Combatant (IC) Science and Technology Objective (STO) program. This program accomplished the following key capabilities in order to enhance individual combatant synthetic environments. In addition, this program participated in an annual Culminating Event, which integrated this project with many others and obtained feedback from soldiers.

Synthetic Speech: Synthetic speech is classified as responsive speech and spontaneous speech. Responsive speech occurs when a synthetic entity generates speech in response to a request by a human participant in the simulation. Spontaneous speech occurs when a CGF entity generates speech without prompting by a human participant. Spontaneous speech is used by the CGF entities to inform the human participant of key events, such as new Opposing Forces (OPFOR) sightings ("spot reports").

Increased Accuracy: The accuracy of the speech recognition system was significantly improved over the duration of this program. There were two aspects to this effort. The first was to select and properly incorporate the COTS STT software most appropriate for VF needs. The second was to develop more robust parsing methods that allow spoken commands that contain minor variations and inconsistencies to be mapped to appropriate DISAF behaviors.

Natural Language Compatibility: The most challenging goal on this project was developing methods that allow human participants in the simulation to communicate with the CGF entities in natural fashion. The most time consuming

and difficult task was creating a method by which geographic locations within the virtual environment can be designated using voice alone. Techniques used were:

- Using cardinal compass points as part of move and fire commands
- Implementing a "follow me" vocal command
- Correlating natural references to landmarks within the synthetic environment.

Robust command parsing was also used to improve performance. The parsing software is able to support multiple command formats where appropriate to relieve the human participant of the burden of memorizing exact speech patterns for use in the simulation.

Natural Language Research is continuing at IST. The goal is to demonstrate the feasibility and effectiveness of allowing live participants in a simulation to converse naturally with their CGF counterparts. Due to the complexity of this task, the scope of the conversations supported by this new research will focus on verbal exchanges inherent in carrying out squad-level and fire team-level missions. CGF entities will have a far greater capability for generating synthetic speech, particularly with regard to requesting clarification when presented with ambiguous commands.

For more information, contact:

Don Washburn or Keith Garfield

University of Central Florida / IST

3280 Progress Drive

Orlando, FL 32826

Phone: (407) 882-1433 /-1342

E-mail: dwashbur@ist.ucf.edu

E-mail: kgarfiel@ist.ucf.edu

James Grosse

STRICOM - AMSTI - ET

12350 Research Parkway

Orlando, FL 32826

Phone: (407) 384-3872

E-mail: james_grosse@stricom.army.mil