QuickSet: A Multimodal Interface for Distributed Interactive Simulation

Phil Cohen, Liang Chen, Josh Clow, Michael Johnston, David McGee, Jay Pittman, Ira Smith

Center for Human-Computer Communication

Oregon Graduate Institute of Science and Technology

P. O. Box 91000

Portland, OR 97291

http://www.cse.ogi.edu/CHCC

Email: pcohen@cse.ogi.edu

Introduction

We demonstrate QuickSet, a wireless, handheld, collaborative, system that can be used to control distributed interactive simulations based on the ModSAF simulator [3] and a 3-D terrain visualization system called CommandVu. Together with the CommandTalk spoken interaction component from SRI International [4] these form the LeatherNet system [1]. With QuickSet, users can formulate a military scenario by creating, positioning, and editing units, supplying them with behavior, specifying fortifications, objectives and other points, etc. In contrast to the original ModSAF GUI, users of QuickSet can employ multiple input modalities, including speech, gesture, and direct manipulation, to suit the task and situation at hand.

QuickSet operates on a 3-lb. handheld 100MHz 486 PC (Fujitsu Stylistic 1000), as well as desktop PC's, employing wireless LAN communications, color screen, microphone, pen stylus, onboard speech recognition, gesture recognition, natural language processing, and multimodal integration. These components communicate with one and the other components of LeatherNet via the Open Agent Architecture [2]. Figure 1 shows collaborative use of the system for creating units, areas, lines, and minefields.

QuickSet can run standalone on the handheld, with a "stub" component emulating ModSAF, and can under user control, upload its scenario to ModSAF to be simulated. Other users can do likewise, with the net effect being an accumulation of entitites and behaviors. For example, this capability can be used for one party playing "blue" and the another "red." When connected online, the handhelds track the ongoing simulation, and therefore are collaborative, in that each can see and effect the course of the simulation. Users can also couple their handheld systems, so that panning and zooming of one will cause identical activity on the others.

Figure 1: Quickset: Multimodal, collaborative simulation set-up.

Multimodal Interaction

Empirical work in our laboratory has investigated the use of speech-only, pen-only, and combined pen-voice interfaces in a variety of tasks, including map-based tasks, using high-fidelity Wizard-of-Oz simulations 5, 6]. These studies have demonstrated that speech-pen interfaces, as compared with unimodal interfaces, yield significantly faster task performance, with significantly fewer user errors and disfluencies. Post-experimental interviews confirm that subjects strongly prefer multimodal to unimodal interaction.

Each modality has characteristic strengths and weaknesses. For example, speech enables creation of new entities, selection of entities not currently visible on the screen, or among a large number of entities shown on the screen. In addition, it is frequently useful for assigning tasks or procedures to entities, and for creating rules. However, speech inut is occasionally unusable, for example, in high noise conditions. Pen input is often more convenient (and more accurate) for indicating that are currently in view on the screen. It also allows specification of irregular lines that might indicate routes or boundaries of areas such as swamps, landing zones, assembly areas, etc. Pen input also has the advantage of being usable in ublic situations where concerns for privacy and security arise. Multimodal interaction provides the user with the ability to capitalize on both sets of advantages, using whatever modality meets the needs of the moment. In addition, experimental subjects often switch modes to correct errors, such as writing a word that the speech recognition system fails to recognize.

The heart of QuickSet is its synergistic integration of modalities. Not only can users employ speech and gesture simultaneously, their integration often corrects recognition errors. For instance, the user might speak the command "M1A1 platoon follow this route" while concurrently drawing a route with the pen to some objective. Depending on the shape of the route, the gesture recognizer might recognize the ink as a route, an area, a tap, a letter or digit, a map symbol, or an editing gesture. Likewise, the speech recognizer might produce other interpretations of the speech. Through a generalized unification procedure and a common semantic representation, the best joint interpretation that makes sense for the current task is selected. For example, the task semantics for route-following expects a linear object as a route. Thus, the multimodal interpreter will select a gesture roughly linear in extent provided the joint interpretation scores sufficiently highly.

Agent-Based Architecture

The Open Agent Architecture [2] is a distributed agent framework that fosters interoperation among agents running on PC's or Unix platforms located anywhere on the Internet. The agents communicate with a simple Horn clause content language. Agents register their capabilities with a "facilitator"agent, who stores relevant information on a global blackboard. Agents can read from and write to the blackboard, can broadcast information, can request that goals be solved, and can post triggers (i.e., requests for conditional actions) on the blackboard. Goal satisfaction often is accomplished asynchronously, while synchronous solving of goals essentially results in a distributed Prolog-like execution.

Because multiple user interfaces can subscribe to the same events, and can offer the same capabilities, the basics of a heterogeneous collaboration system are immediately available. For example, the current system not only incorporates PC-based QuickSet, ModSAF, and 3-D virtual reality interfaces simultaneously, it was also relatively straightforward to include multimodal web-based interface that communicates with the agent architecture and through it to ModSAF.

Field Testing

The system is being tested by NRaD in San Diego, and by the US Marine Corps at 29 Palms, California. It was also tested during the Royal Dragon Exercise in June of 1996 by the 82nd Airborne Corps at Ft. Bragg. Field testing emphasized the importance of multimodal input when speech recognition became unusable while confronted with low-altitude F-15 noise. Finally, the multimodal interface will be combined with a GUI from Ascent Technologies and deployed for use in Exercise Initialization for DARPA in December. Further information about QuickSet can be found in [7].

Acknowledgements

This work is sponsored by the Information Technology and Information Systems offices of DARPA under contract number DABT63-95-C-007, and also in part by ONR Grant, N00014-95-1-1164. It has been done in collaboration with the US Navy's NCCOSC RDT&E Division (NRaD) and SRI International.

References

1. Clarkson, J. D. and Yi, J., LeatherNet: A synthectic forces tactical training system for the USMC Commander, *Proceedings of the Sixth Conference on Computer Generated Forces and Behavioral Representation*, Univ. of Centeral Florida, Orlando, 1996, 275-281

2. Cohen, P. R., Cheyer, A., Wang, M., and Baeg, S. C. An Open Agent Architecture, *Working notes of the AAAI Spring Symposium Series on Software Agents*, Stanford, CA 1994, 1-8.

3. Courtmanche, A. J., and Ceranowicz, A. ModSAF Development Status, *Proceedings of the Fifth Conference on Computer Generated Forces and Behavioral Representation*, Univ. of Central Florida, Orlando, 1995, 3-13.

4. Moore, R. C., CommandTalk: Spoken language interface to the LeatherNet System, ARPA Software Technology and Intelligent Systems Symposium, Chantilly, Virginia, 1995, 28-31.

5. Oviatt, S. L. Multimodal interfaces for dynamic interactive maps. *Proceedings of CHI'96 Human Factors in Computing Systems*, ACM Press, NY, 1996, 95-102.

6. Oviatt, S. L., Cohen, P. R., and Wang, M. Q., Toward interface design for human language technology: Modality and structure as determinants of linguistic complexity, *Speech Communication 15, 3-4* , 1994, 283-300.

7. Pittman, J., Smith, I., Cohen, P. R., Oviatt, S. L., and Yang, T. C., QuickSet: A multimodal interface for military simulation, *Proceedings of the Sixth Conference on Computer Generated Forces and Behavioral Representation*, Univ. of Centeral Florida, Orlando, 1996, 217-224.