

How natural inputs aid interaction in graphical simulations?

Luc JULIA, Adam CHEYER, John DOWDING, Harry BRATT, Jean Mark GAWRON,
Elizabeth BRATT and Robert MOORE
SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, U.S.A.

Abstract. In this paper, we're showing how we are improving the interface to a graphical simulation by adding natural inputs, such as Speech and Gesture Recognition, and Natural Language Understanding using an Agent based Architecture.

1. Overview

ModSAF provides both 2D and 3D (CommandVu, Fig.1) representations of a field where the users, the commanders, are issuing military-type commands.

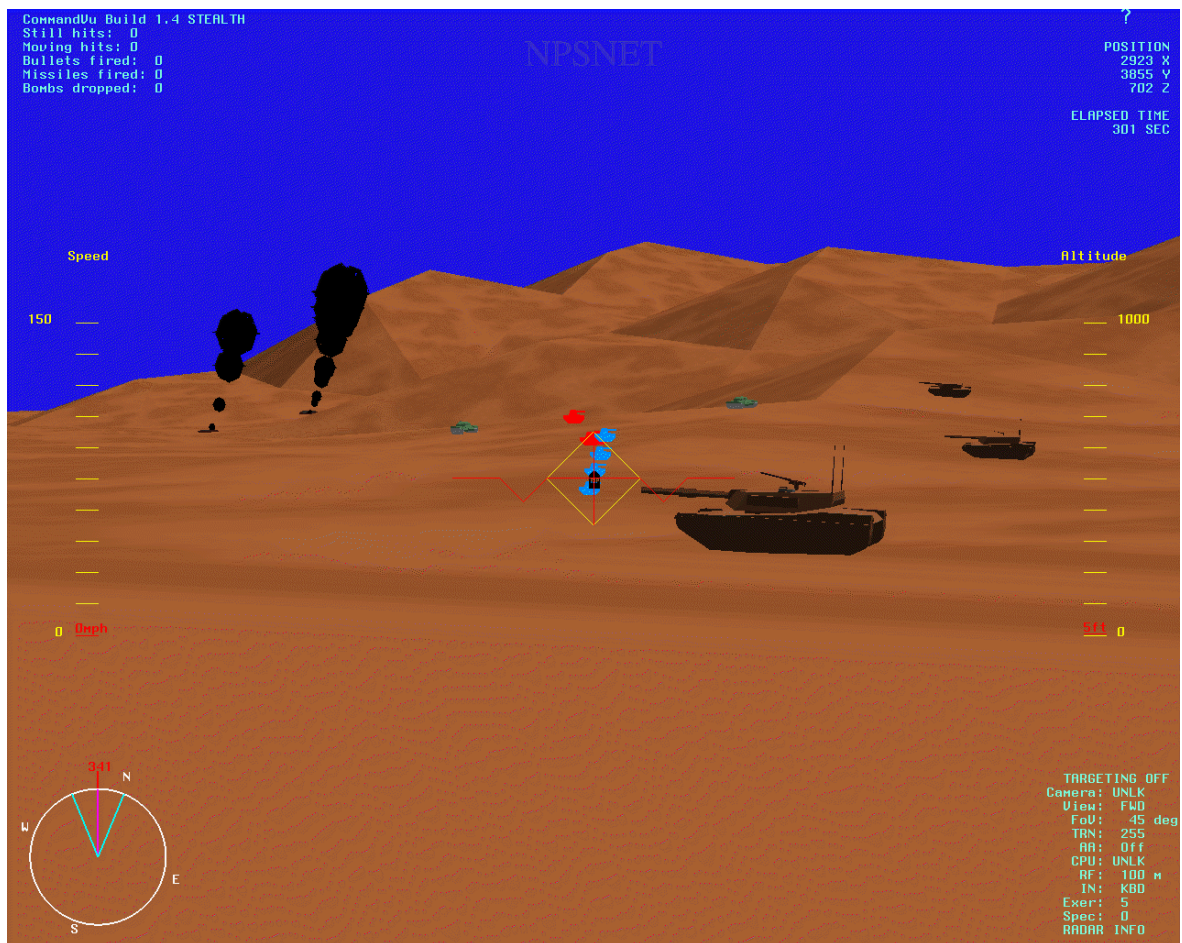


Figure 1: CommandVu, ModSAF 3D representation

2. Implementation

A mouse-and-menus interface, although easy to use for simple applications, is awkward, slow and difficult to learn for complex command-and-control operations in battlefield simulation systems. Giving a single command to a simulated unit can require performing many mouse clicks and negotiating multiple levels of menus. That is why SRI International first addressed this problem by developing CommandTalk [1], a spoken-language interface that allows commanders to use natural, spoken English commands to control simulated forces. SRI's Open Agent Architecture (OAA, Fig. 2, [2]) was used to rapidly put the needed pieces together, i.e. the Nuance speech recognizer [3], the Gemini natural language (NL) parsing [4] and interpreting system and the ModSAF simulator, in order to make them talk to each other.

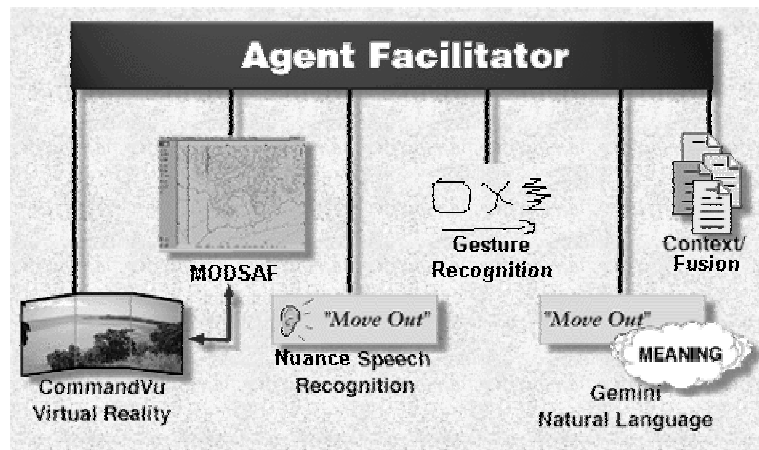


Figure 2: OAA configuration for ModSAF

3. Experience

We also developed, using OAA, another map based system, Multimodal Maps (MMAF, [5]) which provides, in addition to the speech, other natural modalities such as 2D gestures (both deictic and emblematic, [6]) and handwriting [7]. We brought in the architecture, very similar to CommandTalk, a key component that handles the fusion of modalities, the potential ambiguities and redundancies. Using the parallel outputs of the NL (processing both speech and handwriting inputs) and the gesture agents, the fusion agent is trying to build a valid command by filling slots according to the Verb-Object(s)-Variable(s) model (VO*V*), Fig. 3.

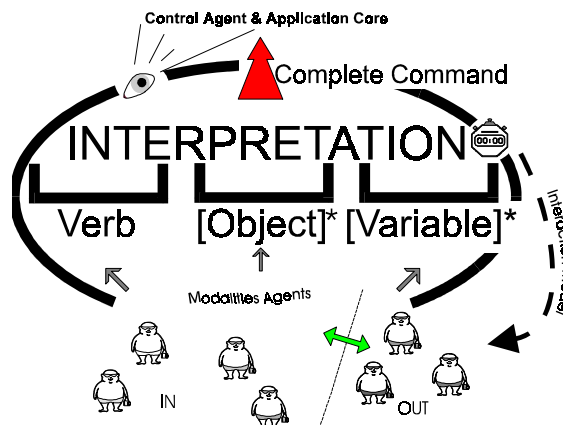


Figure 3: VO*V* Model

This interpretation agent also provides a multimedia mechanism in order to prompt the user if a command is not valid or if the ambiguities are impossible to solve within the local contexts.

4. Improvements

Adding to CommandTalk this new set of agents, which is effortless in the plug and play OAA environment, allows the user to issue synergistic, natural multimodal commands such as: Create forces and control measures - "Create an M1 platoon here" + deistic gesture, or a crossout gesture to delete an unit. Assign missions to forces - "First platoon, on my command, advance in a column to Checkpoint 1, using that path" + drawing of the path. Modify missions during execution - "Change formation to echelon right". Control system functions, such as the display: "Center on Objective Alpha", or right arrow to scroll the map or the 3D view. The vision shown in Fig. 4, is a step closer, a body motion/gesture recognition agent has to be brought to the community in order to achieve this goal.

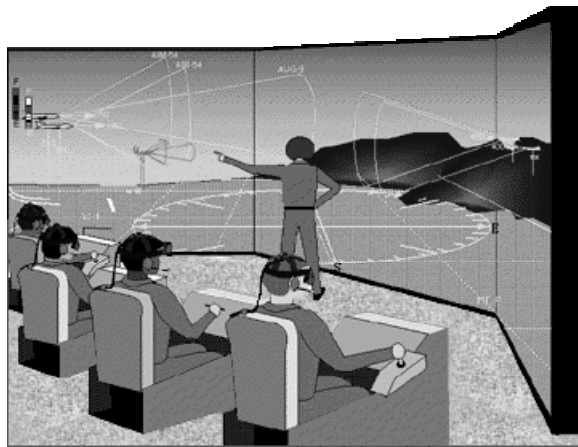


Figure 4: Multimodal Control

According to numerous studies, including ours [8], speech and gesture interfaces provide a faster, easier-to-learn and far more natural way to interact with simulated systems as soon as the response time is acceptable, not far from what the user expects in “reality”.

References

- [1] R. Moore, J. Dowding, H.Bratt, J.M. Gawron, Y. Gorfu and A. Cheyer. CommandTalk: A Spoken-Language Interface for Battlefield Simulations. Fifth Conference on Applied Natural Language Processing, Washington, D.C., 1997.
- [2] P. Cohen, A. Cheyer, M. Wang and SC Baeg. OAA: An Open Agent Architecture. AAAI Spring Symposium, Stanford, California, 1994..
- [3] <http://www.nuance.com>. Nuance Communications is a SRI's spin-off.
- [4] J. Dowding, J.M. Gawron, D. Appelt, J. Bear, L. Cherny, R. Moore and D. Moran. Gemini: A Natural Language System for Spoken-Language Understanding. 31st Annual meeting of the Association for Computational Linguistics, Columbus, Ohio, 1993.
- [5] A. Cheyer and L. Julia. Multimodal Maps: An Agent-based Approach. International Conference on Cooperative Multimodal Communication (CMC/95), Eindhoven, The Netherlands, 1995.
- [6] L. Julia and C. Faure. Pattern Recognition and Beautification for a Pen Based Interface. ICDAR'95, Montreal, Canada, 1995.
- [7] <http://www.cic.com>. Communication Intelligence Corporation is a SRI's spin-off.
- [8] A. Cheyer, J.C. Martin and L. Julia. A Unified Framework for Constructing Multimodal Experiments and Applications. International Conference on Cooperative Multimodal Communication (CMC/98), Tilburg, The Netherlands, 1998.