The Intelligent Room

Michael Coen, Krzysztof Gajos, Marion Groh, Stephen Peters, Brenton Phillips, Nimrod Warshawsky, Luke Weisman, Kevin Wilson, Andy Chang, Deb Dasgupta, Nicholas Hanssens, Stephanie Chiou & Rodney Brooks

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The goal of our research is to create *Intelligent Environments* in which computation is invisibly and seamlessly used to enhance ordinary activity.

Motivation: The driving force behind the Intelligent Room [4, 2] is to make computation not only genuinely user-friendly but also essentially invisible to the user. Interaction with computers should be via mediums that people are naturally comfortable with. The user-interface primitives of these systems are not menus, mice, and windows, but gesture, speech, affect, and context. Their applications are not spreadsheets and word processors but intelligent environments and personal assistants.

Coupled with their natural interfaces is the expectation that these systems are not only highly interactive (i.e. they talk back when spoken to) but also that they are useful for ordinary activities. They enable tasks historically outside the normal range of human-computer interaction, by connecting computers to naturally occurring phenomena that have traditionally been outside the purview of contemporary user-interfaces.

Previous Work: Research in the emerging field of Intelligent Environments is still in an infant state. A comprehensive collection of recent work can be found in [1].

Approach: Our approach with the Intelligent Room has been to embed multiple computer vision [5] and speech understanding systems [4] in an ordinary office. The vision systems recognize the postures, gestures, and faces of people within the room, and the speech recognition/understanding systems accept fairly unrestricted spoken input within narrow domains.

By embedding user-interfaces this way, the fact that people, for example, tend to point at what they are speaking about is no longer meaningless from a computational viewpoint and we can build applications that make use of this information. In some sense, rather than make computer-interfaces for people, we are creating people-interfaces for computers.

Within this room we have created a wide variety of applications that explore high-level human-computer interaction. These include a prototype *command post of the future*, which allows people to interact with and manage large, online stores of information:

> **User**: "*Computer, activate the command post.*"
> **Room**: "*The command post is activated.*" Room configures hardware and software for the command post application.
> **User**: "*Show me a map of Europe. Track all available satellites.*"
> **Room**: Displays interactive map on one of two projected displays.
> **User**: Selects region on map with laser pointer. "*What is this country?*"
> **Room**: "*You are pointing at Iraq.*" Displays data sheet about Iraq on adjacent projected display.
> **User**: "*I need information*"
> **Room**: "*What's your question?*"
> [User can now freely dictate any question
> **User**: "*What are Iraq's ballistic missile capabilities?*
> **Room**: "*I'm asking the START system for the answer...*" Displays list of missile capabilities on adjacent projected display.

Difficulty: The greatest challenge in creating an embedded system like the Intelligent Room is dealing with the vagaries and complexities of the real-world. Prototype hardware systems, such as those used for vision and speech, tend toward being less than fully reliable, and the "best" methods for processing their data are hotly contested research problems. In response, we have developed multiple techniques to not only cope with limitations in the current state of the art but also redefine it, particularly in the room's speech understanding system.

Figure 1: Several cameras in the Intelligent Room and the objects of their attention. There are currently seven video cameras in the room, of which three can look around the room under computer control. (Cameras are circled in the picture.)

Impact: Our goal is to bring computation into the realm of ordinary human activities and enable natural human-computer interaction. The Intelligent Room is a first step toward creating the kind of naturally interactive applications that until now have only been found in science fiction.

Future Work: Currently, we are augmenting the room with new devices and applications, including: a tutorial system that learns how people can use the Intelligent Room and communicates this to new users; an in-depth integration of the room's state with the START speech understanding system, and a deployment of the system into a meeting environment.

Research Support: This material is based upon work supported by the Advanced Research Projects Agency of the Department of Defense under contract number F30602-94-C-0204, monitored through Rome Laboratory. Additional support was provided by the Mitsubishi Electronic Research Laboratories.

References:

[1] Coen, M. (ed.) Proceedings of the 1998 AAAI Spring Symposium on Intelligent Environments. AAAI Technical Report SS-98-02. Stanford, California. 1998.

[2] Coen, M. Design Principles for Intelligent Environments. Proceedings of the Fifteenth National Conference on Artificial Intelligence. AAAI-98. Madison, Wisconsin. 1998.

[3] Coen, M. The Future Of Human-Computer Interaction or How I learned to stop worrying and love My Intelligent Room. *IEEE Intelligent Systems*. March/April 1999.

[4] Coen, M.; Thomas, K.; Weisman, L.; Groh, M.; and Yee, A. A Natural Language Modality for an Embedded Multimodal Environment. To appear in *Proceedings of the 1st International Workshop on Managing Interactions in Smart Environments.* Dublin, Ireland. Lecture Notes in Computer Science. Springer. Heidelberg. 2000.

[5] Coen, M., and Wilson, K. Learning Spatial Event Models from Multiple-Camera Perspectives To appear in *Proceedings of the 25th Annual Conference of IEEE Industrial Electronics. IECON'99.*