

Robotic User Interface Enabled Interactive Dialogue with Intelligent Spaces

by
Rubaiyat Khan

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

June 2006
May 23, 2006

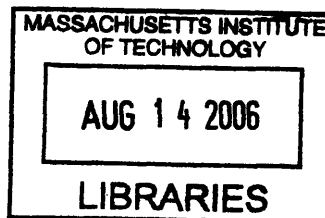
© Massachusetts Institute of Technology 2006. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and
distribute publicly paper and electronic copies of this thesis
and to grant others the right to do so.

Author.....
Department of Electrical Engineering and Computer Science
May 23, 2006

Certified by.....
Dr. Howard E. Shrobe
Principal Research Scientist, MIT CSAIL
Thesis Supervisor

Accepted by.....
Arthur C. Smith
Chairman, Department Committee on Graduate Theses



ARCHIVES

Robotic User Interface Enabled Interactive Dialogue with Intelligent Spaces

by

Rubaiyat Khan

Submitted to the

Department of Electrical Engineering and Computer Science

May 23, 2006

In Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

Abstract

Users can communicate with ubiquitous computing environments by natural means such as voice communication. However, users of the Intelligent Room at MIT CSAIL, a ubiquitous environment, have reported dissatisfaction communicating with the room due to the absence of a focal point and the room's inability to hold a dialogue. To enrich the user's interactive experience, we integrated a Robotic User Interface to the room, and augmented the room's natural language system to enable it to hold dialogues with users. The robotic teddy bear serves two purposes. First, it acts as the focal point of the room which users can address. Second, it enables the room to physically communicate with users by robotic gestures. We also incorporated a book recommendation system to illustrate the room's new ability to converse with users. These enhancements have heightened user experience in communicating with the Intelligent Room, as indicated by our user study.

Thesis Supervisor: Howard E. Shrobe
Title: Principal Research Scientist, MIT CSAIL

Acknowledgement

I would like to thank my thesis supervisor, Dr. Howard Shrobe, for all his help, mentorship, support, and insightful comments on my work.

Many thanks are also due to Kevin Quigley and Stephen Peters, past members of the AIRE research group, who went out of their way many times to help me solve any problems I faced during the course of this project, even after they had left AIRE. Current members of the AIRE research group, particularly Alice Oh, Gary Look, Harold Fox, and Max van Kleek, have also given me much useful feedback.

I would also like to thank all my friends for being so enthusiastic about my project, for giving me ideas about cool new things for the robotic teddy bear to do, and for giving me valuable advice about this work.

I would also like to thank my sister for always being there for me, for proof-reading my thesis to make sure that no last minute mistakes remain, and in general, for being a great sister. Lastly, I would like to dedicate this thesis to my parents Dr. Shazadi Nilufar and Dr. M. A. Mannan Khan. Thank you for always believing in me, and for encouraging me to continue my studies; without your support and dedication, this thesis work would not be possible.

Table of Contents

ABSTRACT	2
ACKNOWLEDGEMENT	3
LIST OF FIGURES.....	6
1 INTRODUCTION.....	7
1.1 MOTIVATION: LIMITATIONS OF HCI IN THE IROOM	7
1.2 THE SOLUTION: THESIS DESIGN GOALS	9
1.3 APPLICATION TO DEMONSTRATE ENHANCED INTERACTION	12
2 BACKGROUND	14
2.1 UBIQUITOUS COMPUTING SPACES.....	14
2.2 HUMAN-COMPUTER INTERACTION.....	16
2.3 AGENT BASED INTELLIGENT REACTIVE ENVIRONMENTS (AIRE) RESEARCH GROUP	17
3 THE INTELLIGENT ROOM	19
3.1 DESCRIPTION.....	19
3.2 THE ARCHITECTURE OF THE IROOM.....	21
3.3 PREVIOUS ATTEMPTS AT NATURALIZING HCI IN THE IROOM	22
4 RELATED WORK	24
4.1 ROBOTS AS AVATARS AND INTERACTION TOOLS	24
4.1.1 REA.....	24
4.1.2 KISMET.....	25
4.1.3 MEL.....	25
4.1.4 SAM	26
4.1.5 Look To Talk.....	26
4.2 SINGULARITY OF THE ROBOTIC BEAR – HOW THE ROBOTIC TEDDY BEAR IS DIFFERENT FROM OTHER HUMAN EMBODIED INTERACTION TOOLS.....	27
5 PROJECT OVERVIEW AND RESEARCH APPROACH	29
5.1 THE ROBOTS: ROBOTPHONE.....	29
5.2 ROBOT AS A FOCAL POINT FOR CONVERSATION WITH THE IROOM	32
5.3 ROBOTIC GESTURE.....	34
5.3.1 The RPDeviceControllerAgent	35
5.3.2 The TeddyLightConversationAgent	36
5.4 DIALOGUE WITH THE IROOM.....	40
5.5 APPLICATION TO ILLUSTRATE THE ENRICHED COMMUNICATION CAPABILITIES OF THE IROOM: A KNOWLEDGE BASED HISTORICAL BOOK RECOMMENDATION SYSTEM	41
5.5.1 Design.....	43
5.5.1.1 Inputs and Output.....	43
5.5.1.2 Classification of books.....	44
5.5.1.3 Knowledge Representation.....	46
5.5.1.4 The Domain.....	47
5.5.2 Implementation.....	48
5.5.3 Optimizations.....	51
5.5.4 Example	52

6	EVALUATION	57
6.1	USER STUDY	57
6.1.1	<i>Study Circumstances</i>	58
6.1.2	<i>Protocols for Study</i>	58
6.1.3	<i>Design of the Study</i>	59
6.1.4	<i>Data Analysis</i>	62
6.1.5	<i>Results</i>	63
6.2	EXPERT EVALUATION	69
7	LESSONS LEARNT	73
8	FURTHER WORK	76
9	CONCLUSION	79
	REFERENCES	81
	APPENDIX A: BOOKTEDDYLIGHTCONVERSATION.GRAM	87
	APPENDIX B: ROBOTIC USER INTERFACE STUDY - DIRECTIONS FOR PARTICIPANTS. “COMMAND AND CONTROL FIRST” VERSION.	89
	APPENDIX C: ROBOTIC USER INTERFACE STUDY - DIRECTIONS FOR PARTICIPANTS. “INTERACTIVE FIRST” VERSION.	94
	APPENDIX D: DATA ANALYSIS FOR USER STUDY	99

List of Figures

Figure 1. The original layout of the Intelligent Room.....	20
Figure 2. A RobotPHONE and the axis of rotation of its arms (reprinted with permission from [4])	30
Figure 3. Layout of the Intelligent Room after the RobotPHONE was integrated.....	32
Figure 4. The Robotic User Interface in the Intelligent Room.....	33
Figure 5. Left panel: Bear in resting position inside the Intelligent Room. Right panel: Bear looking and pointing towards the right front projector.....	34
Figure 6. The TeddyLightConversationAgent at work	37
Figure 7. Reliance relationships for the TeddyLightConversationAgent	39
Figure 8. The RUI pointing at pictures of books recommended by the system.....	42
Figure 9. Partial list of books in the knowledge base	45
Figure 10. Example rules from the book recommendation system.....	46
Figure 11. Transcript of a run of the BookTeddyLightConversationAgent	54
Figure 12. Inference tree for the BookTeddyLightConversationAgent	55
Figure 13. Robotic User Interface pointing at screen showing picture of recommended book.....	66
Figure 14. Summary of user study results.....	68

1 Introduction

As technology becomes more powerful and cheaper with every passing day, computer system designers are opting more and more for ubiquitous computing environments over traditional computer systems as ubiquitous environments are human-centric and allow users the mobility they desire. With this movement comes a need to make the interaction between users and their ubiquitous computing environments more natural and casual.

This chapter describes the motivation behind this project. Limitations of the current state of human computer interaction in the Intelligent Room (IRoom), an ubiquitous computing environment at the Massachusetts Institute of Technology, cause user dissatisfaction. This project augmented the Intelligent Room's human computer interaction system to make interaction richer and more interesting.

1.1 Motivation: Limitations of HCI in the IRoom

*Alas, how is't with you,
That you do bend your eye on vacancy
And with the incorporal air do hold discourse?
[Hamlet, Act 3, Scene 4]*

Users of the Intelligent Room, as Shakespeare puts it, "bend their eye on vacancy and hold discourse with the incorporal air". As part of making human-

computer interaction seem more natural, the Intelligent Room allows its users to issue voice commands by simply talking out loud in the room. This innovative form of human-computer communication has created a new problem – even though talking to a computer is more natural than typing on a keyboard, since the computer in the Intelligent Room is embedded in the environment, talking to the computer entails talking to air. Historically, this is considered aberrant and is not well accepted in society. Since conversing with one's surroundings is not something that comes naturally to humans, IRoom users have reported feeling uneasy talking to space without any localized focus. A physical representation of the environment that can act as a "listening part" would make many users feel more comfortable because they will be able to address something specific instead of talking to air.

The second problem with the present IRoom is that it currently follows a command and control architecture. The user issues commands and the room follows them. There is no way for the room to establish a dialogue with the user or otherwise confirm whether it has understood the user correctly. Thus, when the user points at a lamp and says "switch this on", the room has no way of confirming whether its understanding about which device the user wants switched on is correct or not. Therefore, errors are sometimes made, which makes the Intelligent Room look not so intelligent. This command and control architecture also means that it is not possible for the room to ask questions from

the user to infer the user's needs from the answers to these questions.

The third inadequacy in human-computer interaction in the IRoom is that the room does not have a bodily representation or an avatar that can physically communicate with users via gestures. This lack of a physical representation implies that the room cannot point towards objects, and that it has to rely on synthesized speech to acknowledge user input.

Thus, it is clear that the model of interaction between the IRoom and its users needed to be refined to make it more natural, to allow a dialogue to be established between the user and the room, and to allow the room to ask questions to confirm the user's needs.

1.2 The Solution: Thesis Design Goals

The goal of this project was to overcome the limitations mentioned above by adding a Robotic User Interface (RUI) to the Intelligent Room and augmenting the natural language system of the room to allow dialogues.

The RUI consists of a robotic teddy bear that was integrated with the existing infrastructure of the Intelligent Room. This bear now acts as a physical and visual

representation of the environment surrounding the user which can be addressed. The user is able to address the RUI instead of casting about in search of an anchor point on which to focus. The presence of this “focal point” in the form of a robotic teddy bear in the IRoom substantially enhances a user’s level of comfort in carrying on voice communication with the IRoom rather than talking to thin air.

In addition to responding to the user through computer generated speech, now, the room can also use hand and head gestures of the robot for communication. In certain situations, voice communication is simply not sufficient for getting meaning across. For example, if a new user unfamiliar with the room is looking for the right front projector, the room can now point to the projector in question using the teddy bear. When the room did not have access to the robot, it had to describe the position of the projector in words, which may not necessarily be easy to understand and follow. Something like “the right projector is located on the east wall of the room, 2 feet from the corner between the north and the east wall” can neither be easily parsed, nor easily understood. Therefore, the robotic bear’s role in enhancing user interaction with the room extends far beyond just acting as a focal point which users can address during voice communication.

Additionally, the teddy bear can substantially enhance a user’s experience in interacting with the room because its physical appearance emulates friendship, warmth, and trust. Also, since most users have had a teddy bear as a toy at

some point in their childhood, a robot in the shape of a teddy bear is more likely to make the user feel more comfortable and trusting of the IRoom. The bear can be programmed to greet a user when she first walks into the IRoom by waving its arms and saying hello. Such an action will immediately endear the bear, and thereby the IRoom, to the user. The presence of the teddy bear will also make interaction with the Intelligent Room more attractive to children. One feature of the teddy bear is that if the user says "I am bored", or "entertain me", the bear recites Edgar Allan Poe's famous poem "The Raven". This type of informal interaction with the toy like teddy bear has a quality of amusement in it that does not fail to entertain users and enrich their experience in using the room.

A scenario in which a computer-generated voice working together with physical gestures generated by the teddy bear helps the room communicate with users naturally and efficiently is as follows:

Alice is working in the IRoom and needs the floor lamps to be switched on. She walks into the room and says out loud "Switch on the lamps." Hearing this, the robot points its arm towards the lamps and asks Alice if these are the lights she wants switched on. Alice's answer to this question is acknowledged by a nod of the bear's head, and if Alice confirms her request, the lights are turned on. Thus, a dialogue using both audio and physical communication tools is established between the room and the user, and the user's answer to the room's questions provides confirmation.

Such a scenario reduces erroneous understanding of the user's command by the room.

Thus, the goal of this project was to integrate the teddy bear into the IRoom's

infrastructure so that it could effectively work as an avatar for the intelligent environment. In an effort to make interaction more natural, the natural language system of the room was also enhanced to enable the room to hold a conversation with the user.

1.3 Application to Demonstrate Enhanced Interaction

To illustrate all the enhancements made to the interaction capabilities of the IRoom, a knowledge-based system that recommends historical fictions to users depending on user preferences was added to the infrastructure of the room. This book recommendation system uses the room's conversational abilities to ask the users questions about their book preferences. Which question is asked next depends on the user's answer to the previous questions – making this a true dialogue between the room and the user, and not the replaying of some pre-recorded questions. The user's answers help deduce which book they may like by using an inference system, and the names of the books are then read aloud using synthesized speech, while pictures of the books are displayed using a projector. During the question-answer process, the teddy bear is used to acknowledge the user's answer by nodding its head. Once the book recommendations are made, the teddy bear also points to the screen to attract the user's attention to pictures of the books that the system recommended. This

book recommendation system therefore demonstrates how the teddy bear's gestures and nodding, and the ability of the room to hold dialogues with the user can substantially enrich a user's interaction experience with the room.

2 Background

This chapter describes the history of ubiquitous computing and human-computer interaction. It also discusses the Agent Based Intelligent Reactive Environments research group that built the IRoom where the work for this project took place.

2.1 Ubiquitous Computing Spaces

Most of today's computers are very powerful and are, therefore, capable of helping users solve many different types of problems. However, these computers require the user to have a considerable amount of knowledge about how to use the computer itself to accomplish the required task. Thus, using a computer becomes another problem that needs to be resolved before users can complete their original task. For example, to sum some numbers, the user needs to know how to operate spreadsheets or calculator software; if the user is unaware of how these software works, she will first have to learn the software before she can accomplish her original goal of summing the numbers. Today's technology, therefore, causes the user's focus to shift from the task itself to how the task can be accomplished. Thus, today's "machine-centered technology" often forces humans to serve the machines by making them learn to speak the computer's language. Additionally, humans need to learn to use the machine's interaction

tools such as keyboards. To take full advantage of today's technological advances, we need to build human-centric computers which users can control naturally and intuitively.

Ubiquitous computing strives to make it easy for the user to communicate with computers naturally by eliminating the difference between what the user knows about computers and what she needs to know. It essentially converts the computer into an invisible servant that can understand the users in their natural environment, without any need for the users to undergo computer training. The computer thus becomes an extension of the users "unconscious" in order to achieve most tasks by simply using intuition. In essence, ubiquitous computing environments strive to make computing as omnipresent and natural as the oxygen we breathe. Mark Weiser, the father of the field of ubiquitous computing, says "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it" [5]. He describes the age of ubiquitous computing as the "age of calm technology," when technology recedes into the background of our lives [6]. Thus, the goal of the field of ubiquitous computing is to embed computers in our environments in such a way that users can use them without even thinking about them, let alone learning how to use them.

Current ubiquitous technology consists mostly of intelligent or reactive environments, also known as smart spaces. These are mostly situated in computer science research laboratories, where computing technology has been embedded into the environment. These spaces are equipped with different types of sensor mechanisms (such as vision sensors, sound sensors, motion sensors) which enable them to observe and collect data about what users of the space are doing. The computers embedded in these smart spaces can then understand and manipulate these data to assist users with their tasks. For example, vision sensors in a smart space may observe that its user has left the room, and can use this data to reason that it is no longer necessary to keep projectors and lights running, and can therefore switch these devices off.

2.2 Human-Computer Interaction

Human-computer interaction has been a topic of much interest among ubiquitous computing researchers. Traditionally, the requirements of the computer have always determined human-computer interaction. Users have always had to learn to communicate with computers in a way that the computers will understand, and to use tools that would never be used in a natural setting. These standard interaction modalities, such as mice and keyboards, may take a considerable amount of time and practice for the users to gain expertise in using.

In addition to using tools to communicate with a computer, humans also need to learn specific ways to talk to the computer so that it understands their goals. For example, to use a spreadsheet, the user needs to know the different parts of the graphical user interface (GUI) of the spreadsheet, and which component of the GUI needs to be clicked on to perform which calculation. Obviously, this is not what ideal communication with a computer should be, as the user's intentions should be implicit. Hence, it is important to replace these modalities with some other more natural means of communication with computers in ubiquitous computing environments.

New and innovative means of communication that are currently being used in many intelligent spaces include voice communication via close talking microphones and microphone arrays, visual communication through facial and hand gestures and communication via movement.

2.3 Agent Based Intelligent Reactive Environments (AIRE)

Research Group

AIRE is a research group in the Computer Science and Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. It created and maintains the Intelligent Room, on which the work of this thesis was based. AIRE

designs and constructs smart environments by augmenting spaces with basic perceptual sensing, speech recognition, and vision recognition. [16]

Most of AIRE's pervasive computing work is part of Project Oxygen, whose goal is to create human centric computing that is "pervasive, nomadic, adaptable, intentional, eternal, and powerful, yet efficient". Project Oxygen hopes to create reactive environments that Weiser had originally envisioned. [17] The role of the AIRE research group in Project Oxygen is to create Intelligent Environments.

3 The Intelligent Room

3.1 Description

Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Laboratory's Intelligent Room (IRoom), developed by the AIRE research group, is a fully functional conference room that is instrumented with ubiquitous technology. It is located on the second floor of the Ray and Maria Stata Center at MIT, and supports business meetings, conferences, entertainment, etc. It is equipped with white boards, multiple projectors, various video inputs (including three cameras for image processing and gesture recognition), a microphone array and close talking microphones for audio input, lamps operated via X10, video-multiplexers so that any output image can be shown on any display etc. All these audiovisual hardware are connected to the Metaglue System. [53]

Figure 1 shows the layout of the Intelligent Room before the robotic teddy bear was integrated into it.

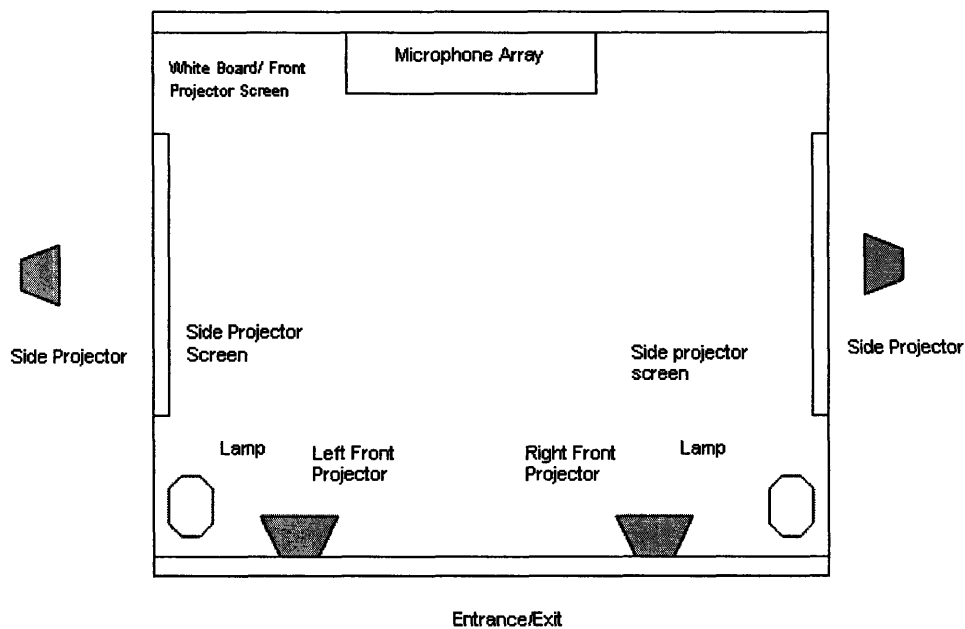


Figure 1. The original layout of the Intelligent Room

As Rodney Brooks, the director of the Computer Science and Artificial Intelligence laboratory at MIT says, “[In] the Intelligent Room, the computer is drawn out into the world of people, and forced to operate there, listening to what people say, watching what they do and how they move, keeping track of what is going on, and trying to be helpful when there is something that they can do. In the Intelligent Room the person is in charge; the computer has to play by human rules.” [19]

The IRoom is essentially an intelligent environment. Different features of the room can be used by natural human-computer interaction tools like issuing voice commands. For example, to switch on a display or to open a computer application, the user just needs to say his intentions out loud so that microphones can record his voice command and pass it on to the system which then carries out the required task.

Early origins of the IRoom can be traced back to the early 1990s when MIT researchers started developing agent based smart spaces. Some of this work evolved into research in human-computer interaction. These efforts later converged to create the Intelligent Room project, which still continues in the AIRE research group. [19, 12, 13, 14]

3.2 The Architecture of the IRoom

The software infrastructure of the Intelligent Room is provided by Metaglué, a multi-agent system implemented in the Java programming language [1, 3]. Each agent in Metaglué is simple software that interacts with similar independent agents to carry on a particular task in the IRoom. All agents that serve a common purpose belong to the same society. The idea of society of agents is taken from

Minsky's revolutionary idea of the "society of mind" [3]. For example, an agent is responsible for the proper functioning of each device in the room. A lamp agent might be in charge of switching a lamp on and off, while a projector agent might be in charge of controlling a projector, while yet another agent might decide when a lamp and/or projector needs to be switched on and off.

Metaglué allows remote agents to interact with each other through Java's remote method invocation (RMI) capabilities. Agents can also communicate with each other via the publish/subscribe model. Other benefits of using the Metaglué architecture include mechanisms for resource discovery and management [43], robust recovery mechanism for failed components [3], persistent storage, multimodal input and output through speech gesture and graphical user interfaces [51], user customization etc.

3.3 Previous Attempts at Naturalizing HCI in the IRoom

As described earlier, interaction between the users and the IRoom is comparatively much more natural than that between a traditional computer and its user. The most widely used means of interaction in the Intelligent Room is voice conversation using a close-talking microphone. The natural language

system of the room is capable of understanding the user's speech and then reacting to it.

However, not all users find using a close-talking microphone very convenient as it requires the user to be in close proximity with the microphone. To allow the user to walk around without being tied to a microphone, the Large AccOUstic Data array (LOUD) microphone array was created and integrated with the Intelligent Room. This microphone array tracks the user and selectively records her voice (ignoring other noise in the room) so that her voice commands can be clearly understood. [59]

Other attempts to make human computer interaction more natural include facial and hand gesture recognition by the room. The room is also capable of understanding pointing. If a user points at a lamp with her finger, video cameras in the room pick up the horizontal protrusion due to the extended hand. Similarly, if a user points at a part of a projected presentation with a laser pointer, the video cameras pick up the bright red color of the laser pointer. The user can then refer to the object she is pointing to as "it" or "this", and the room is able to understand the user. Thus, the room will be able to correctly switch on one of the two floor lamps if the user points at one and asks the room to "switch this on" [19]. Traditional communication tools like keyboard and mice are also present in the room.

4 Related Work

4.1 Robots as Avatars and Interaction Tools

Robots have been used for human-embodied interaction in other intelligent systems. Some of these systems are described below.

4.1.1 REA

The Gesture and Natural Language research group at the MIT Media lab have developed autonomous agents that are capable of having a real-time face-to-face conversation with humans. These agents look like humans and can communicate both verbally and non-verbally. They provide a means of natural interaction with computers since humans already know how to speak with other humans, and therefore, do not need to undergo training to interact with a computer interface that looks and speaks like a human. The REA (Real Estate Agent) is such an agent. It is a computer generated humanoid that uses speech, eye gaze, body posture, hand gesture, and facial displays to communicate with users. [21, 22]

4.1.2 KISMET

The Sociable Machines Project at the Humanoid Robotics Group at the MIT Artificial Intelligence Laboratory has developed an “expressive anthropomorphic robot” called Kismet that interacts with people naturally in face-to-face scenarios. Kismet obtains information about the world through various visual and auditory channels and communicates with users using facial expression, body posture, gesture, gaze direction and voice. Kismet was built to “enter into natural and intuitive social interaction with a human caregiver, and to learn from them”. The interaction between Kismet and its caregiver was modeled after a human parent-infant relationship, and work in the Kismet project was inspired by “infant social development, psychology, ethology, and evolution. Kismet has been designed to support several social cues and skills that could play an important role in socially situated learning with a human instructor”. [23, 24, 60]

4.1.3 MEL

MEL is an interactive robotic penguin created at Mitsubishi Electronic Research Laboratory. MEL can imitate a “human’s conversational gaze behavior during collaborative conversation” [26]. In addition to speaking, MEL also gestures with its wings, and uses its head to track users and to look at objects under

discussion. In addition, MEL can also open and close its beak and flap its wings up and down. The conversational and collaborative capabilities of MEL are provided by the Collagen™ middleware for collaborative agents and IBM ViaVoice. MEL was used to “build a visual grounding module for interactive dialogue systems.” [25]

4.1.4 SAM

SAM is an expressive and responsive user interface in the IRoom. SAM currently consists of a “minimal depiction of an animated face” that can be displayed on a projector screen; SAM can express various human emotions such as anger, confusion, surprise, and worry, and therefore, show the “emotional state” of the Intelligent Room. [16]

4.1.5 Look To Talk

Look To Talk (LTT) is a “gaze-aware interface that directs an utterance spoken by the user to a software agent in a multi-user collaborative environment”. LTT was developed for the IRoom at MIT CSAIL, and uses SAM as an avatar for the IRoom. LTT makes sure that the IRoom speaks to an user only when the user is facing SAM. If the user is facing another user in the IRoom, LTT assumes the

user is talking to the second user and not to the room. In such a situation, the speech recognition system of the environment remains inactive, making sure that the room does not respond to utterances that are really directed at other humans and not at the environment. [16, 58]

4.2 Singularity of the Robotic Bear – How the Robotic Teddy Bear is Different from Other Human Embodied Interaction Tools

The Robotic teddy bear integrated to the IRoom is somewhat different from all the projects described above. First, the teddy bear is a three dimensional physical object that users can actually touch. Thus, it is very different from REA (a virtual humanoid) and SAM (an animated face). One advantage of the teddy bear being a physical object as opposed to a virtual interface is that it is always present in the IRoom. SAM, for example, needs the projector to be turned on so that it can be displayed. The teddy bear is always present and the user can interact with it without switching on any other hardware or software in the IRoom. Thus, it is constantly available, and not dependent on anything else for its use. Even though Kismet has a physical representation, Kismet consists of a face, and is not capable of pointing with its arms. However, Kismet can create facial expressions, which the teddy bear cannot. The extent of the teddy bears

movement includes moving its arms and nodding and moving its head. Of all the human embodied robots mentioned so far, the teddy bear is most similar to MEL, since both MEL and the bear can move their heads and arms. However, MEL cannot be used as an input device. The teddy bear's arms or head can be moved to provide haptic input to the system.

5 Project Overview and Research Approach

To accomplish the goals of enriching user interaction with the Intelligent Room, a Robotic User Interface (RUI) was added to the room, and the natural language system of the room was augmented to allow dialogue between the room and its users. This chapter describes in detail the research approach taken.

5.1 The Robots: RobotPHONE

The robotic teddy bear that was integrated with the IRoom is known as RobotPHONE. It was developed by the Inami Laboratory of the University of Electro Communications in Tokyo, Japan. The teddy bear is a RUI that can be used for physical communication since the robot can be programmed to rotate its arms and head by programmer specified angles. The robot can also be used as a haptic input device; the programmer can physically move the robot's arms and head; this motion can then be stored in the robot's memory and later played back. Figure 2 shows how the robots arms can be moved to point at different directions. The robot's head can also move side by side and up and down.

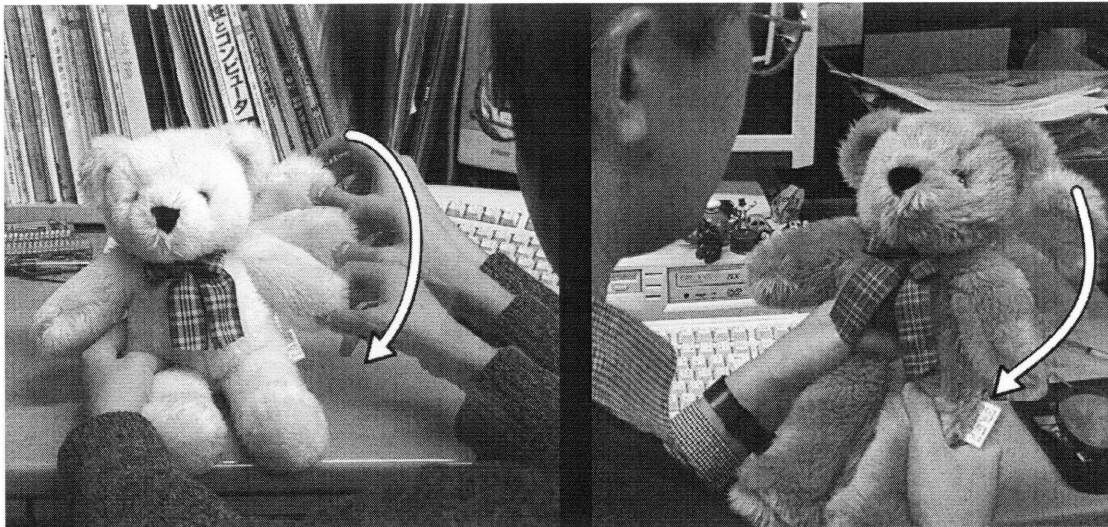


Figure 2. A RobotPHONE and the axis of rotation of its arms (reprinted with permission from [4])

The fact that the RUI is shaped like a teddy bear is important since such a human like shape (head, two arms, two legs, torso) allows it to achieve human like motions. For example, it can move its head to “look” at different directions and move its arms to “point” at different directions. RobotPHONE was chosen for integration into the IRoom because of its ability to look and point like humans. An additional advantage to the robots being teddy bear shaped as opposed to machine like robots featured in science fiction movies is that users find this shape familiar and hence comforting and friendly to use. In fact, many users of the Intelligent Room have commented that the teddy bears are “cute” and that they like interacting with them.

In research conducted in Japan, these robots were used as Robotic User Interfaces (RUI) to provide haptic feedback to remote users. Two users at

different ends of a network connection could each use a RobotPHONE to transmit gestures to the opposite end. So if Alice and Bob both have RobotPHONEs connected to a network and Alice waves the right arm of her RobotPHONE, her RobotPHONE works as an input device and records the movement data in a motion file (files with “.mot” extension). It then transmits the motion file to the other end of the network. Bob’s RobotPHONE, upon receiving the motion file recorded by Alice’s RobotPHONE, can then act as a display device and play it back. This will result in Bob’s RobotPHONE also waving his right arm in the exact same way as Alice’s did. As a result, the bears can act as physical avatars for users. [4]

In the Intelligent Room, instead of working as physical extensions for other users, the RobotPHONE was utilized as a physical extension of the IRoom itself. In the summer of 2005, researchers from the Inami laboratory collaborated with the AIRE group, as a result of which the robots could be used in the IRoom.

5.2 Robot as a Focal Point for Conversation with the IRoom

The robotic teddy bear that was added to the infrastructure of the room is located right next to the central whiteboard that doubles as the screen for the main projector (the right front projector) of the room. Figure 3 shows the new layout of the Intelligent Room after the RUI was integrated.

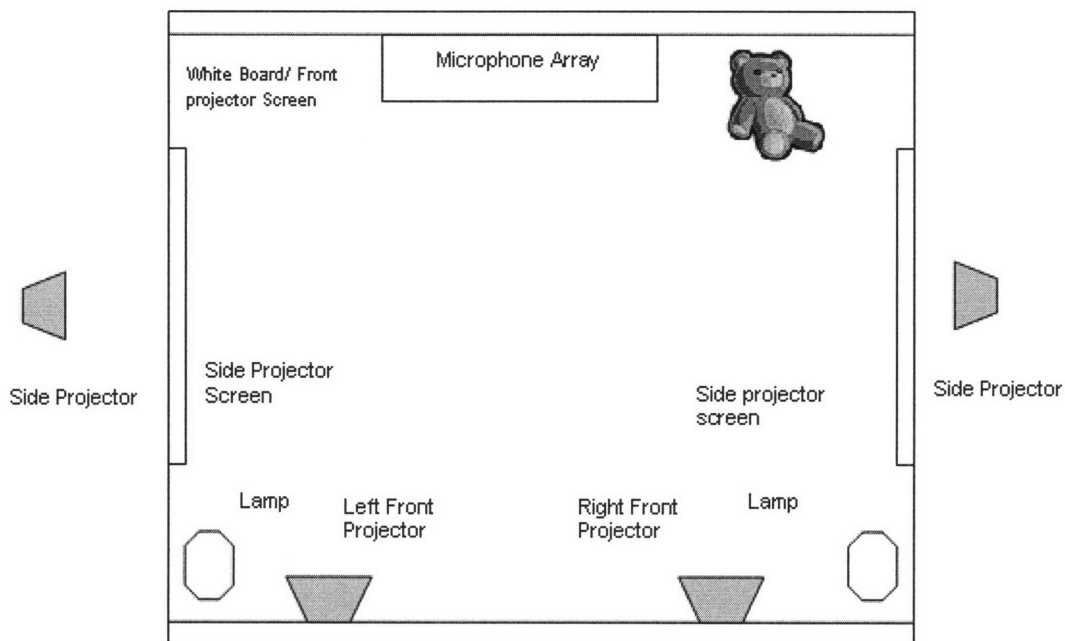


Figure 3. Layout of the Intelligent Room after the RobotPHONE was integrated

Figure 4 shows a picture of the robotic teddy bear in the Intelligent Room.

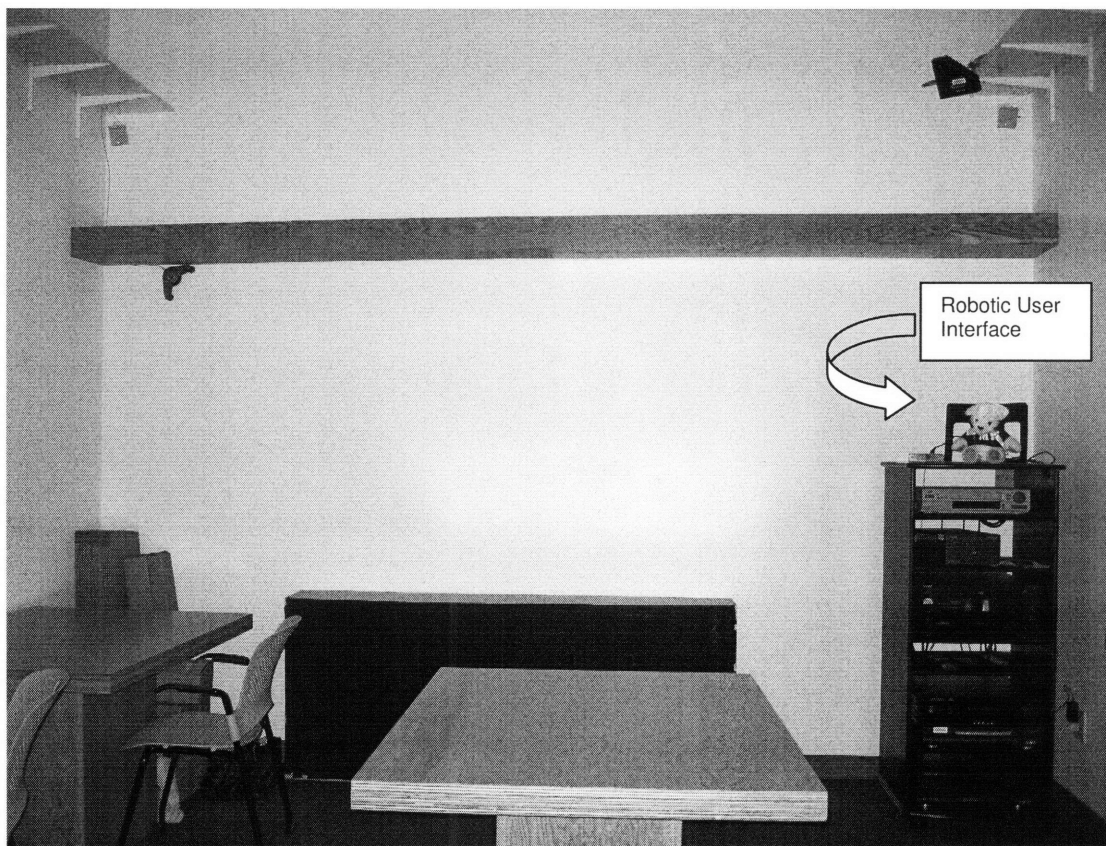


Figure 4. The Robotic User Interface in the Intelligent Room

The microphone array is below the central whiteboard, and the robot is situated just beside it. Users normally face the whiteboard and the microphone array while using the room; therefore this location is ideal for the teddy bear as it places the bear in the focal point of the room and users automatically find them addressing the bear. The robot itself faces towards the door and is situated at the eye level of standing users so that the user's eyes fall on it as soon as she walks into the Intelligent Room.

5.3 Robotic Gesture

Once the robot was integrated to the IRoom, it could act as a physical extension of the room. Head and hand gestures of the robot are now used by the room to look at different directions and to point at different objects. Figure 5 shows the robot in two different positions. In the left hand panel, it is not being used by the IRoom, and is sitting in its resting position. In the right hand panel, the robot is looking and pointing towards the right front projector of the room in response to the user query “show me the right front projector”.

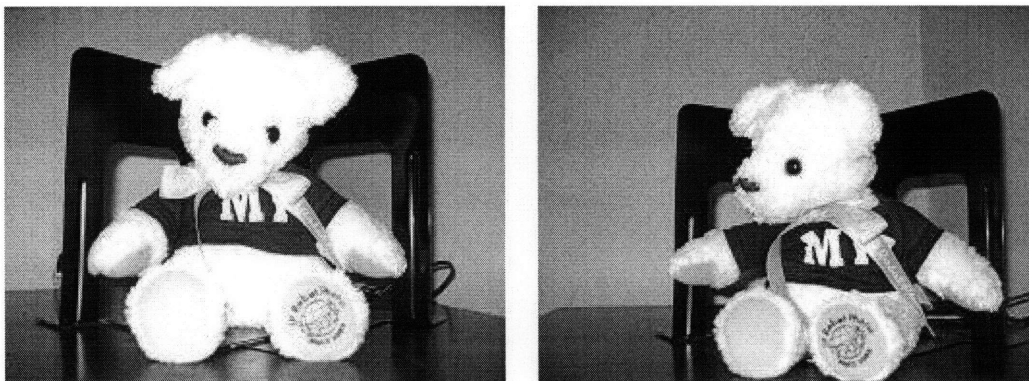


Figure 5. Left panel: Bear in resting position inside the Intelligent Room. Right panel: Bear looking and pointing towards the right front projector

The teddy bear can also move its head in a nodding motion. Its head nods are now being used in the Intelligent Room to acknowledge user utterances in response to questions asked by the room.

All the software that controls hardware in the Intelligent Room consists of Metaglué agents. Therefore, to incorporate the robotic teddy bear with the Intelligent Room, Metaglué agents needed to be written to control the robot and to make it react to user inputs. Metaglué is an extension of the Java language; therefore, software that runs in the IRoom has to be written in Java.

5.3.1 The RPDeviceControllerAgent

A Metaglué agent was written to manipulate the robots by calling native methods provided by the manufacturer of the robots. This agent, called the `RPDeviceControllerAgent`, has the following functionalities:

1. It can make the robot move. This agent is capable of moving both the head and the arms of the robot. The head and each arm have two axes of rotation, giving the robot a total of six degrees of freedom in movement. Therefore this agent can make the robot point at different objects (by moving its arms) and look at different directions (by moving its head).
2. This agent can record motion files (with the extension “.mot”) when the programmer physically moves the robot’s arms and head using the haptic feedback system provided by the manufacturer.
3. This agent can play back motion files which cause the robot to move its heads and arms in exactly the same way as the time when the motion file was recorded.

To enable the robot to point at different objects in the room, motion files were recorded for playback. These files were recorded while the programmer manually moved the robot's arms and head. Therefore, the robot points at the door when a user asks "show me the door" by playing the door.mot file. This file was created by the programmer by physically moving the arms and the head of the robot in the direction of the door and then recording this motion using the record() method in the RPDeviceControllerAgent.

An alternative way of pointing and looking at objects in the room was tried and rejected. This alternative technique used the robot's built-in movement methods to rotate its arms and head by programmer specified angles. However, this scheme resulted in jerky robotic movements and was therefore not used.

5.3.2 The TeddyLightConversationAgent

The TeddyLightConversationAgent listens for user inputs and can hold a dialogue with the user. It responds to a user's voice command in two ways. First, it uses the RPDeviceControllerAgent to move the teddy bear's arms and head if needed. This use of the RPDeviceControllerAgent allows the room to make the robot look at different directions, and point at objects. Secondly, the TeddyLightConversationAgent uses the SpeechOutAgent to generate

speech in response to the user's commands. Figure 6 shows some examples of the `TeddyLightConversationAgent` responding to user voice utterances by using both robotic gesture and voice communication.

User Says	Room Says (SpeechOutAgent)	Robot's Response (RPDeviceControllerAgent)
Hello	Hello to you too	Moves head slightly to acknowledge greeting
Teddy	Hi there	Moves head slightly to acknowledge greeting
Show me the right lamp	The right lamp is in that direction	The head looks towards the right lamp. The right arm points towards the right lamp.
Show me the door	The door is in that direction	The head looks towards the door. The right arm points towards the door

Figure 6. The `TeddyLightConversationAgent` at work

A "hello" from the user elicits a "hello to you too" from the robot; a "teddy" or "bear" from the user causes the bear to greet her by moving its head and saying "hi there"; finally, a "show me the door" will prompt the robot to point towards the door (or any other object in question) and reply "the door is in that direction".

The `TeddyLightConversationAgent` performs its function by continuously listening for the user's voice commands when it is running. If the user says something, the agent matches the user's utterance against a grammar file, `TeddyLightConversation.gram`, and checks if the user's speech is included

in that file. If it is, then the agent looks at its `acceptedTagResults()` method to find out how it needs to respond to the user's voice command.

For example, if the user utters "show me the door", the `TeddyLightConversationAgent` will find a tag in `TeddyLightConversation.gram` called `<whereisdoor>` that is activated by this particular utterance. It will then invoke its `acceptedTagResults()` method. A snippet of java code from this method is shown below:

```
public void acceptedTagResult (Hashtable tags) throws
RemoteException{
.
.
.
    if (tags.containsKey("whereisdoor")){
        speechOut.saySafe("the door is in that direction");
        rpDeviceController.point("door.mot");
    }
.
.
.
}
```

This method tells the robot that the `door.mot` file needs to be played back by the `RPDeviceControllerAgent` to make the robot look and point at the direction of the door. It also makes the `SpeechOutAgent` say out loud "the door is in that direction" in reply to the user's utterance. Thus, the `TeddyLightConversationAgent` improves interaction between the room and its users by making the robot move and by answering the user using computer

generated speech.

To use both the `RPDeviceControllerAgent`, and the `SpeechOutAgent`, the `TeddyLightConversationAgent` needs the Metaglu primitive `reliesOn()` to create handles to the `RPDeviceControllerAgent` and the `SpeechOutAgent`. Figure 7 shows the reliance relationship between the `TeddyLightConversationAgent`, the `RPDeviceControllerAgent`, and the `SpeechOutAgent`.

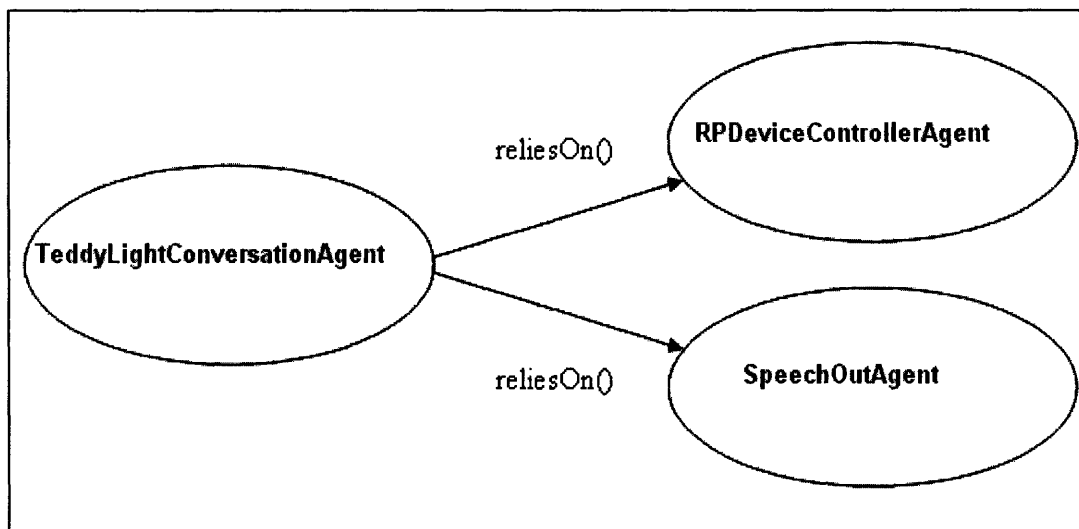


Figure 7. Reliance relationships for the `TeddyLightConversationAgent`

The Intelligent Room uses IBM ViaVoice for speech communication. IBM ViaVoice can only run on Windows machines; therefore `TeddyLightConversationAgent` has to run on one of the Windows machine in the IRoom. Currently, it is tied to *leifr.csail.mit.edu*. The design of `TeddyLightConversationAgent` was inspired by the `HAL2001Agent` which can respond to simple voice commands by using computer generated speech.

5.4 Dialogue with the IRoom

The `TeddyLightConversationAgent` can also hold a dialogue with the room, enabling it to confirm a user command. For example, when the user says “switch on the lamps”, the room uses the `SpeechOutAgent` to ask the user “are you sure you want the lights switched on?” Only if the user replies with a yes to this question will the room switch on the lights. If the user answers no, the room will not switch on the lights. If the user does not reply, the room will not do anything about the light but continue to listen for other user inputs.

To accomplish this, certain voice commands have been marked in the `TeddyLightConversation.gram` grammar file as commands that need user confirmation. When the room hears such a command, it then starts to listen for a yes or a no. When it hears a yes or a no, it considers these to be the user’s

confirmation. In case it hears something else, it matches it against the grammar file to check if it is a command. In case the new utterance is indeed a command, this agent responds to the new command, while waiting for confirmation for the previous command.

5.5 Application to Illustrate the Enriched Communication Capabilities of the IRoom: A Knowledge Based Historical Book Recommendation System

To better illustrate the room's enhanced interaction capabilities that include the ability to hold a conversation with the user, acknowledging user input by nodding the teddy bear's head, and pointing at objects using the teddy bear's arms, a knowledge base consisting of historical fiction books was created. Another agent, the `BookTeddyLightConversationAgent` was written to allow the IRoom to recommend books to the user according to user preferences. This system is the engine that drives dialogue in the room. It finds out user preferences by asking questions and then listening for answers. The order and number of questions asked depend on the user's answer to the previous questions, making the conversation more than the simple play back of pre-recorded question statements. When the IRoom hears a reply from the user to a question that it asked, it acknowledges the reply by nodding the teddy bear's head. The room

then uses the data gathered from the answers to infer books that the user might enjoy. It communicates the book recommendations to the user using verbal, visual, and physical cues. It says out the names of the recommended books using the `SpeechOutAgent`. It also displays the pictures of the books on a screen using the `ImageViewerAgent` written for this purpose. It then attracts the user's attention to the pictures on the screen by making the bear look and point at the direction of the screen as illustrated in figure 8.



Figure 8. The RUI pointing at pictures of books recommended by the system

The following is a detailed description of the book recommendation system, design choices made for the system, and its implementation.

5.5.1 Design

The books in the knowledge base are categorized according to the time period of the events they describe, the setting of the events, and the category of the book's subject matter. The recommendation system is subjective, and its expertise reflects knowledge of a wide variety of books, their thematic aspects and historic settings.

5.5.1.1 Inputs and Output

The book recommendation system takes as input information about the user's reading preferences, including whether the book should be fictional, light reading, set in a foreign country, based on war, etc. Input is received by asking the user questions using the `SpeechOutAgent`, and then listening for the user's answer and matching them with the `BookTeddyLightConversation.gram` grammar file. This grammar file is shown in Appendix A. The system then outputs an ordered list of books that the user might enjoy. This list is read out loud to the user using the `SpeechOutAgent`. The list is also printed out in the console since

the user may want to take a look at the list again later without going through the entire inference system. Pictures of all the books on the list are also displayed using the front projector, and the bear points towards the screen to draw the user's attention towards it.

5.5.1.2 Classification of books

Each book chosen for the knowledge base was classified according to the category of topic it covered, the setting, and the time period of the events described in the book. For simplicity, books were divided into four categories: *drama, romance, mystery and fiction based on war*. The books were also further classified according to whether they were set in the middle ages, or in the eighteenth to twentieth century, and whether they are based in the Americas or in foreign countries. Simple single step inference rules were then used to determine which aspects of the books the user may appreciate, and then books containing these aspects were recommended to the user. Figure 9 shows a partial list of the books after they have been categorized.

- *The Thralls Tale* by Judith Linbergh: American, Drama, MiddleAges
- *The Carribbean* by James Michener: American, Drama, MiddleAges
- *The Teleportation of An American Teenager* by Andrew Rodriguez: Foreign, Romance, MiddleAges
- *Bone Walker* by Kathleen Gear: American, Mystery, MiddleAges
- *The Summoning God* by Kathleen Gear: American, Mystery, MiddleAges
- *The Last of the Mohicans* by James Cooper: American, Drama. 18th-20th Century
- *A Million Nightingales* by Susan Straight: American, Drama. 18th - 20th Century
- *Thunder At Gettysburg* by Patricia Gauch: American, Fiction Based On War, 18th – 20th Century
- *The Armies of the Night* by Norman Mailer: American, Fiction Based On War, 18th – 20th Century
- *Gone With The Wind* by Margaret Mitchell: American, Romance, 18th – 20th Century
- *Into the Wilderness* by Sarah Donati: American, Romance, 18th – 20th Century
- *Too Soon For Flowers* by Margaret Miles :American, Mystery, 18th – 20th Century
- *The Dante Club* by Matthew Pearl : American, Mystery, 18th – 20th Century
- *Baudolino* by Umberto Eco : Foreign, Drama, Middle Ages
- *In the Company of the Courtesan* by Sarah Dunant: Foreign, Drama, Middle Ages
- *The Queen's Fool* by Philippa Gregory: Foreign, Romance, Middle Ages
- *The Name of The Rose* by Umberto Eco: Foreign, Mystery, Middle Ages
- *Six For Gold* by Mary Reed: Foreign, Mystery, Middle Ages
- *The Hidden Diary* by Mary Antoinette: Foreign, Drama 18th – 20th Century
- *Outlander* by Diana Gabaldon: Foreign, Drama 18th – 20th Century
- *Austerlitz* by W. G. Sebald: Foreign, Drama 18th – 20th Century
- *Night* by Elie Wiesel: Foreign, Fiction Based On War, 18th – 20th Century
- *Cat And Mouse* by Guenter Grass: Foreign, Fiction Based On War, 18th – 20th Century
- *Marrying Mozart* by Stephanie Cowell: Foreign, Romance, 18th – 20th Century
- *A Kiss From Maddalena* by Chrisopher Castellani: Foreign, Romance, 18th – 20th Century
- *Dark Assassin* by Anne Perry: Foreign, Mystery, 18th – 20th Century
- *The Lamplighter* by Anthony O Neil: Foreign, Mystery, 18th – 20th Century

Figure 9. Partial list of books in the knowledge base

5.5.1.3 Knowledge Representation

A rule based system was chosen for representing the knowledge and inference structure for this system. A rule based system is most appropriate for this book recommendation system as rules can be created for basic factors that can narrow down the range of interest to certain book topics. Also, the knowledge of how to evaluate the different aspects of the book could be expressed with independent single step inferential rules. Figure 10 shows some of the simple rules that help the system determine the user's preferences and then recommend a book according to these preferences.

Example of relevant knowledge: rules

- murder AND robbery -> crime
- espionage AND violent conflict -> suspense
- James Bond movies and books -> espionage
- fiction AND (crime OR suspense) -> book category is mystery
- cannons AND gun-powder -> medieval technology
- barbarian invasions AND knights AND medieval technology -> middle ages determined
- date of event of interest between 399 CE and 1499 CE -> year falls in middle ages
- middle ages determined OR ((NOT middle ages determined) AND year falls in middle ages) -> time period middle ages
- fiction book AND category is mystery AND foreign setting AND time period is middle ages -> *THE-NAME-OF-THE-ROSE-BY-UMBERTO-ECO*

Figure 10. Example rules from the book recommendation system

5.5.1.4 The Domain

The domain of historical fictions was chosen since it is narrow enough that it can be handled by a rule based system. It also asks enough questions from the user to demonstrate the Intelligent Room's capability of holding dialogues with its users. Also, the structure of the rules and the way the new natural language system of the room is written ensures that the question the room asks the user depends on the user's answer to previous questions, thus creating a true conversation between the room and the user and not merely some sort of replay of prerecorded conversations.

The domain chosen is subjective and heuristic. However, the purpose of the knowledge based system is to illustrate the room's capability to hold a dialogue with its users, both verbally and physically using gesture; therefore, the fact that the domain chosen is subjective is immaterial.

The author of this thesis considers herself to be an expert in this domain since she often selects historical fictions to read based on her own preferences, and articulated most of the rules and the heuristics behind the book recommendations herself. However, she did discuss the problem with friends, and used Amazon.com and BarnesAndNoble.com as resources from which to

find more information about the thematic setting and content of most of the historical fictions in the database.

Since the overall task was well suited to a rule-based representation, any other alternate representations were not considered for the purposes of this project.

5.5.2 Implementation

The rule based historical fiction recommendation system is essentially a Metaglu agent. Therefore, it was written in Java. Two alternative implementation languages were considered and rejected. Joshua Lisp would have been a good choice since it is specialized for the purpose of implementing rule based systems. Its advantages include the fact that rules can be backward chained or forward chained, and it allows the programmer to associate certainty and importance factors for each rule. It also has the ability to explain why the system made a particular inference. All of these are important properties that would have made the historical fiction recommendation system in the Intelligent Room more useful for its users. However, the system needs to be able to call on the `SpeechOutAgent` to ask the user about her preferences, and the natural language system of the room needed to be able to listen to the user's speech. It also needs to be able to call on the `RPDeviceControllerAgent` for nodding

and pointing. The amount of complexity that would have resulted from using Joshua Lisp and interfacing it with Metagluue seemed unnecessary for a simple knowledge based system whose main function is to illustrate the room's new and enhanced interaction capabilities. For very similar reasons such as difficulty and complication in interfacing, JESS, another expert system building tool that usually works very well with Java, was also rejected.

Even though the Metagluue agent which contains the knowledge base and the rules for inference, the `BookTeddyLightConversationAgent`, was written in java, it emulates Joshua's backward chaining inference technique. All rules are written as java methods with if/else branches. Thus, the rule

if A -> B

is written as

```
public static void rule1() {
    if(A) {
        B = true;
    }
    else{
        B = false;
    }
}
```

Appropriate method invocations at the right place at the right time in the course of the run of the book recommendation system ensure that whenever the value of a predicate needs to be checked, all rules that can determine the value of that particular predicate are fired. For instance, if it is necessary to check whether the

book category the user is interested in is *drama*, in addition to calling the method that deduces whether book category is *drama*, the methods that deduce whether book category is *romance*, *mystery*, or *fiction based on war* are also invoked. Thus the system emulates Joshua's reasoning and inference structure.

Invoking all the methods that determine the value of a predicate in succession means that the system asks all related questions in sequence. This makes the user think that the system is following a line of reasoning, and therefore, the system seems more intelligent to the user. In addition to making the system seem smarter, this structure makes sure that the user can follow the intention behind the questions that room asks since related questions are asked in succession instead of haphazardly. Thus, the `BookTeddyLightConversationAgent`, even though it is implemented in java, emulates many of Joshua's reasoning structure. However, as the system is currently written, it is not possible to "tell it a fact". The user has to wait for the system to ask a relevant question.

5.5.3 Optimizations

To make the system efficient in terms of the number of questions it asks, whenever a rule has an OR clause and a predicate evaluates to be true, the rest of the OR clause is not evaluated. Thus, for the following rule:

```
If ((user is interested in cannons) OR (user is
    interested in gun powder) ->
    user is interested in medieval technology
```

In case `user is interested in cannons` evaluates to true, the rest of the predicate is skipped, and the value of `user is interested in medieval technology` is set to true. Thus, if the user is interested in cannons, the system will not ask her about her interest in gun powder. Such an algorithm ensures that the room does not ask unnecessary questions of the user when it already has enough information to make an inference.

Similarly, if a rule contains an AND clause, and a predicate evaluates to false, the rest of the AND clause is skipped, ensuring that only the minimum number of questions are asked of the user, making the book recommendation system quicker, and more intelligent. These optimizations make the system efficient and truly interactive since the order and number of questions is determined according to the user's response at runtime.

5.5.4 Example

When the `BookTeddyLightConversationAgent` runs in the Intelligent Room, it listens for the phrase “I want a book” from the user. If the user speaks this sentence, the agent goes into the book recommendation system and starts firing the rules available to it. When it needs information from the user, it formulates questions and uses the `SpeechOutAgent` to ask these to the user. The `BookTeddyLightConversationAgent` uses the `Metaglu` primitive `reliesOn()` to create a handle to the `SpeechOutAgent`.

After asking a question, the system waits and listens for the user’s answer. Every time the user answers one of its questions, it uses the information to infer as many facts as it can, and then keeps on asking the user for more information until it knows which books to recommend to the user. The rules that determine whether the user is interested in fiction are fired first. The rules that determine whether the user is interested in foreign settings are subsequently fired, followed by those that determine which book category the user is interested in and those that determine the user’s time period of interest. Once the system has enough information, and has inferred the setting, category and time period, it looks at all the books in the database, and recommends those that match the user’s preferences.

A transcript of a conversation between a user and the Intelligent Room is shown in Figure 11. Every time the user responded to a question, the teddy bear nodded its head in acknowledgement. At the very end, when book recommendations were made, pictures of the books were displayed on the screen and the robot pointed towards them.

```

    User: Computer, Stay awake
IRoom: With pleasure
    User: I want a book
IRoom: Are you interested in fiction?
    User: Yes
IRoom: Are you interested in foreign settings?
    User: Yes
IRoom: Are you interested in light reading
    User: No
IRoom: Are you interested in reading about people's joy and
happiness?
    User: No
IRoom: Are you interested in reading about people's pain
and suffering?
    User: No
IRoom: Are you interested in murder?
    User: No
IRoom: Are you interested in robbery?
    User: No
IRoom: Are you interested in James Bond movies?
    User: Yes
IRoom: Are you interested in violent conflicts?
    User: No
IRoom: Your book category is mystery.
IRoom: Are you interested in barbarian invasions?
    User: No
IRoom: Are you interested in knights?
    User: Yes
IRoom: Your time period of interest is the middle ages.
IRoom: I believe you will enjoy the following books
IRoom: The name of the Rose by Umberto Eco
IRoom: Six for gold by Mary Reed

```

Figure 11. Transcript of a run of the BookTeddyLightConversationAgent

Figure 12 shows the inference tree for the run of the book recommendation system shown in Figure 11.

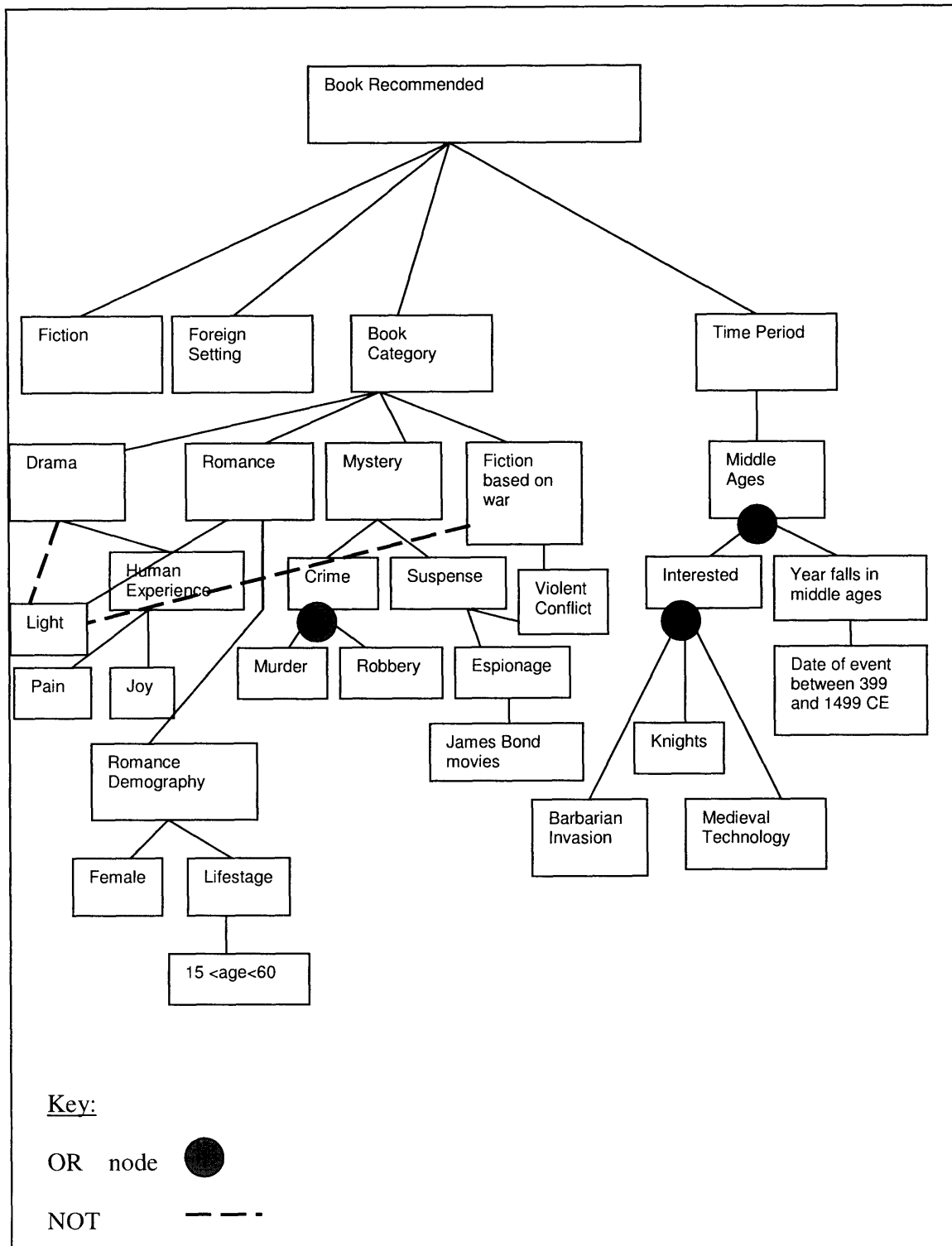


Figure 12. Inference tree for the BookTeddyLightConversationAgent

As evident from figure 11, the room does not follow a command and control architecture any more. It can hold a conversation with its users and use information from the conversation to infer facts and state those facts.

The historical book recommendation system serves the single purpose of illustrating the room's capability of holding discourse with its users. Therefore, the system is small and the inference tree is not too deep.

6 Evaluation

To evaluate the effectiveness of integrating a Robotic User Interface and expanding the speech system to enable the Intelligent Room to converse with its users, two different types of assessment techniques were used. A user study was carried out where typical users of intelligent spaces tested the IRoom's RUI and conversational abilities. Additionally, an expert heuristic evaluation was done by the author of this thesis to test any usability problems that may exist in the system.

6.1 User Study

The goal of the user study was to evaluate the effectiveness of the teddy bear and the room's conversational abilities in enhancing the interaction between the room and the user. Most subjects chosen for this study were expert computer users who were somewhat familiar with ubiquitous computing environments. Therefore, they were able to provide us with an opportunity to find out if the augmented IRoom meets the needs of the target audience.

6.1.1 Study Circumstances

Twenty nine subjects tested the Intelligent Room's new ability to converse, point at objects, and provide acknowledgement to user input. The study took place inside the Intelligent Room at the Stata Center at MIT. Two versions of the study were prepared; in one, subjects interacted with the room first in a *command and control* mode, and later in an *interactive* mode. The other group of users interacted with the IRoom first in the *interactive* mode, and later in the *command and control* mode. The subjects were randomly divided into two groups, and fifteen completed the *command and control first* version of the study whereas fourteen completed the *interactive first* version of the study.

6.1.2 Protocols for Study

Each subject was randomly assigned to either the *command and control first* or the *interactive first* version of the study. They then had to follow a set of written directions which asked them to say some sentences out loud in the room, and then answer questions about their interaction experience. The experimenter remained in the IRoom during the duration of the study and was available for questions or clarifications. Please see appendices B and C for the study directions provided to subjects. The investigator observed the user interacting

with the IRoom and took notes about the proportion of time they were looking at the bear while talking to the room; the experimenter also noted whether the users looked interested during the study and if they smiled at the bear. The entire study took approximately ten minutes per user.

6.1.3 Design of the Study

Since the study was designed to learn if the bear and the conversational abilities of the room has improved human-computer interaction in the room, subjects of the study were asked to carry out the same task inside the room twice, once in the *interactive* mode (when the room asks questions for clarification and provides acknowledgement via the bear's nods) and once in the *command and control* mode (where the user's command is followed without providing any acknowledgement or asking for clarifications). Specifically, the tasks that subjects had to complete were:

1. Saying hello to the teddy bear. In response, the room says "hello to you too" and the teddy bear waves its right arm. The purpose of this step is to familiarize the user with the bear and the room.
2. Switching on the lights of the room. The subject said out loud "switch on the lights". In the *command and control* mode, the room just turned the lights on. In the *interactive mode*, the room asks the user, using synthetic voice, "Do you want the lights turned on?" The user's answer to this

question is acknowledged by the bear's head nod, and lights are turned on only if the user says yes. The user then has to rate the usefulness of the room's confirmation question in a scale of 1 to 7 (1 is not useful, 4 is neutral, and 7 is very useful). Since we foresaw that some users may not find confirmation questions for a simple task such as switching on lights useful, we also asked users to rate the usefulness of confirming an important and irreversible step.

3. The user asks for a book recommendation from the system by saying out loud "recommend a book for me". In the *command and control* mode, the room speaks the name of a book that the user might enjoy, and displays the picture of the book in the front screen. In the *interactive* mode, the room asks questions about whether the user is interested in foreign settings, murder mysteries, drama, and so on, so that it can recommend a book based on this user's particular preferences. Also, every time the user answers a question, the bear nods its head as acknowledgement. Finally, the names of the books recommended are said out loud, and pictures of the books are displayed on a screen. The bear's right hand points towards the screen to attract the user's attention towards it. This step is designed to determine the usefulness of the robotic user interface's physical gestures. Users rate how useful the bear's head nod was as acknowledgement, and how useful the bear's hand motion was in attracting their attention towards the screen.

4. Lastly, the users are asked whether they generally liked the presence of the bear, whether they thought the bear enhanced their interaction experience, whether they thought the room's conversational abilities enhanced their interaction experience, and whether they preferred interacting with the room in the command and control mode or in the interactive mode.

The purpose of the behavioral observation of the study by the investigator was to discover if the teddy bear was indeed acting as the focal point of conversation in the room, and if the users seemed interested in interacting with it. Thus, she answered the following questions after every study:

1. Did the subjects direct their talk towards teddy bear while talking to the room?
2. Did the subject focus on the teddy bear while the room was speaking?
3. Did the subject smile when the robotic teddy bear nodded, looked at the subject, or pointed?
4. Did the subject look interested during the interaction?

The study was designed to be a "Wizard of Oz" experiment; the experimenter was inputting the subjects' spoken utterances to the system to bypass the voice recognition step. This was necessary since the voice recognition system is user dependent, and needs to be trained before it can accurately understand a user.

6.1.4 Data Analysis

Since our sample size is small (twenty-nine users), the population variance is unknown and estimated by the sample variance. Also, since variability may be significant in small samples, and we assume that the data is normally distributed, we used one sided t-tests to test hypotheses on means of the rankings. All our numerical data is a ranking of usefulness on a scale of 1 to 7, where 1 is equivalent to not useful, 4 is equivalent to neutral, and 7 is equivalent to very useful. Thus, the null hypotheses for all our t-tests were:

H0: mean ranking ≤ 4 (user's are either neutral towards this feature, or do not find it useful)

The alternative hypotheses for all our tests were:

H1: mean ranking > 4 (user's find this feature useful)

The results of the t-tests were then used to determine whether the null hypotheses can be rejected at an alpha level of 0.05. [27]

For qualitative data with a limited set of possible values, we present the count for each of the possible values for the data. Thus, for a question with a yes or no answer, we report the number of users who chose yes versus the number that chose no.

6.1.5 Results

Twenty-five subjects out of twenty-nine found that the room's conversational abilities enhanced their interaction experience with the room. Twenty-two users out of twenty-nine generally liked the presence of the bear in the Intelligent Room, and twenty-one out of twenty-nine users said the bear played a part in enhancing their experience in interacting with the room.

The mean ranking for whether users enjoyed interacting with the bear was 5.9 out of 7, and a one-sided t-test at an alpha value of 0.05 showed that this ranking is statistically significantly higher than the neutral value of 4. Details of this t-test are shown in Appendix D. Thus, on average, users enjoyed interacting with the robotic teddy bear in the IRoom. One reason cited by many users is that they found the bear cute, friendly, and amusing. Users liked the innovative "bodily or physical communication" that a RUI can perform, such as nodding and pointing. Some even thought that the interaction with the robot helped alleviate loneliness. Even though most users found the bear enjoyable, some reported that they could not quite match the computer generated voice with the teddy bear as the voice was very low pitched, and was emanating from multiple speakers situated all around the room (surround sound). But, this problem can be easily fixed by changing the profile of the synthetic voice and using a single speaker situated near the bear.

The room asked for confirmation when requested by users to switch on lights in the interactive mode. The mean ranking for the usefulness of this confirmation is 4.034. This mean is not statistically significantly different from the neutral value of 4 according to a one sided t-test at an alpha level of 0.05. Therefore, it was not possible to reject the null hypothesis that the mean ranking is less than or equal to 4. Hence, on average, users did not find this confirmation useful. Users deemed confirmation for such a simple and easily reversible task unnecessary, and felt they would get irritated at the IRoom if it asked for too many confirmations.

The mean ranking for the usefulness of a confirmation question before carrying out an important an irreversible action is 6.6, which is statistically significantly different from the neutral value of 4 according to a one-sided t-test at an alpha level of 0.05. Users reported they want to be sure that their voice command was properly understood by the environment if they are doing something that cannot be easily undone. Thus, users want confirmations, but only for important tasks that cannot be undone, and not for something as simple and reversible as turning the light on.

The mean ranking for the usefulness of acknowledging a user's voice input via the bear's nodding was 4.6, which is statistically significantly different from the

neutral value of 4 at an alpha level of 0.05. Thus, on average, users did find the acknowledgement useful. Most users reported that the acknowledgement was a form of feedback, and that they prefer to receive feedback and system status over not knowing what the system is doing. Also, users felt the bear's nods added to the interactive experience as they felt like they were talking to another "person" who understands what they are saying. However, some users reported noticing a perceptible delay between the bear's nods and the user's voice command and felt that the experience would be much smoother if this delay could be reduced.

All twenty-nine users saw the pictures of the recommended books displayed on the screen in the *command and control* mode (where the bear did not point its right hand towards the screen), and twenty of them said that the bear's hand being pointed towards the screen during the *interactive* mode was not helpful. The reason behind this is two-folds. Firstly, the image viewer occupies almost the entire screen, which is equivalent to half the front wall of the IRoom. When something of such a significant size appears, users look at it anyway, without needing a bear to point at it for them. Also, when the bear did point, most of the users did not notice the bear's movement, let alone find it useful, since the bear is much smaller than the size of the image viewer. However, many of the users commented that if the robot were bigger than the image viewer, attracting the user's attention using the robot's hand gesture would be useful. Additionally, the

bear's arm does not have any fingers; therefore, many users did not realize that the movement of its arms was meant to point towards the screen. Figure 13 shows a close-up picture of the robot's arm pointing towards the screen to illustrate the difficulty that users faced in understanding what the bear was doing. As is apparent from the picture, it is somewhat difficult for a user to comprehend that the bear's right arm is pointing at the screen.



Figure 13. Robotic User Interface pointing at screen showing picture of recommended book

Thus, it seems that if we had a different, bigger robot, with fingers for pointing, this feature would add to the overall interaction experience in the room. But with the current RUI, users did not find this feature helpful.

Lastly, fourteen out of the twenty-nine users said they would prefer to use the room in the *interactive* mode, while eleven reported the intention of using the room in the *command and control* mode. Three users suggested that they would like a mix of both modes depending on their needs.

The user study has shown that the Robotic User Interface does in fact enhance user experience in interacting with the Intelligent Room. Users find its nodding head motions for acknowledging voice input helpful as a form of feedback; they like its presence in the IRoom since it acts as a focal point in the room, and is cute, and therefore fun to interact with. The results also show that the room's conversational abilities enhance user interaction with the room, and that users would like confirmation before the room performs an important action. Figure 14 summarizes these results.

Feature	User Test Results
Robotic User Interface as interaction tool	<p>22/29 users generally liked the presence of the RUI in the room</p> <p>21/29 users found the RUI enhanced their interaction experience with the IRoom</p> <p>Mean ranking for "enjoyed interacting with robotic teddy bear" is 5.9 (statistically significantly higher than 4)</p>
Conversational abilities of IRoom	25/29 users found that the room's conversational abilities enhanced their interaction experience
Usefulness of confirmation question for switching on light	Users did not find this feature useful. Mean usefulness ranking was 4.034, which is not statistically significantly different from the neutral ranking of 4
Usefulness of confirmation question for an important and irreversible action	Users found this feature useful. Mean usefulness ranking was 6.6, which is statistically significantly higher than 4
Usefulness of acknowledging user's voice input using the RUI's head nod	Users found this feature useful. Mean usefulness ranking was 4.5, which is statistically significantly greater than 4.
Usefulness of the bear pointing towards the screen showing pictures of recommended books	20/29 users found this feature unhelpful.
Interactive vs. Command and Control	<p>14/29 users would rather use the interactive mode</p> <p>11/29 users would rather use the command and control mode</p> <p>3/29 users would like to use a mix of both modes</p>

Figure 14. Summary of user study results

The investigator observed that, on average, while issuing voice commands to the room, users were focusing their attention on the robot 62% of the time. However, there was great variation in this data (standard deviation 33%). During the time when the room was speaking, users' eyes focused on the robot around 54% of the time (standard deviation 35%). These observations suggest that whether users find the RUI a focal point of conversation in the room or not varies greatly from individual to individual. Users who did not focus on the robot during

conversation of the room said they found the room's speech and the RUI disjointed, probably because the speech was coming from all directions, and not from the teddy bear.

The investigator also noticed that all twenty-nine users looked very interested during interaction with the room, and that twenty-six out of twenty-nine users smiled at the robot at some point during the interaction. These observations imply that users did enjoy interacting with the robot.

6.2 Expert Evaluation

A heuristic evaluation of the RUI was carried out by the author of this thesis to find any usability problems. The RUI and the conversational abilities of the room were rated using Nielsen's ten heuristic guidelines [61]. The results are as follows:

1. Visibility of system status

The system provides visibility of system status through both verbal feedback, such as spoken confirmation of user request, and physical feedback in the form of the RUI's head nods.

2. Match between system and the real world

The language spoken by the system is essentially user language, and not technical words that are difficult for the user to understand. For example, to switch on the lights in the IRoom, the user simply needs to say “switch on the lights”. Similarly, to ask for a confirmation, the system uses natural questions like “do you want the lights switched on?” Thus the RUI system matches the real world.

3. User control and freedom

Users have control over whether they want to use the RUI or not. If they do not wish to interact with the RUI or converse with the room, they simply need to say out loud “stop the bear”.

4. Consistency and standards

The system consistently uses natural language that users would normally use to interact with other humans. The commands are also easy for users to learn as phrasing is consistent. Some typical imperatives include “start the bear”, “stop the bear”, “show me the door”, “show me the projector”.

5. Error prevention

Each time the user makes an important and irreversible request, the RUI confirms it before acting. Thus, the system prevents errors.

6. Recognition rather than recall

The user needs to know what utterances the room understands. If they do not remember these, they need to refer to the grammar file. Thus, the system does not completely meet this guideline. However, this is a minor problem since the system is designed to understand sentences that a user would normally say. For example, to switch on a light, the user simply needs to say “switch the light on”; she does not need to issue a complicated command that she needs to remember.

7. Flexibility and efficiency of use

As of now, the system behaves the same with both novice and expert users. Since expert users may not want confirmations or acknowledgements, they may find the system inefficient and slow.

8. Aesthetic and minimalist design

Dialogues between the room and the user only contain relevant information. Therefore, the system is designed minimally.

9. Help users recognize, diagnose, and recover from errors

Currently, if the user issues a voice command that is incorrect, the system does not respond since the incorrect voice command is likely not in the

grammar file. Therefore, the system lets the user realize that she might have made a mistake by not carrying out the incorrectly requested task.

10. Help and documentation

If the RUI cannot perform a function requested by the user, such as “show me the door”, it is unable to articulate the exact error that occurred. However, there is help available on our research group’s webpage about why such errors might occur and how to repair them. Since the IRoom is a complex system, the users usually contact system administrators when the room does not behave properly. As help from the system administrators is available, this issue is a minor problem.

7 Lessons Learnt

During the course of this project, the most important lessons we learnt were the importance of creating a truly seamless multi-modal system and choosing an appropriate RUI for communication with users. We also learnt that it is very difficult to manage a complex base of code and write extensible code that can be scaled for use in large systems.

We learnt that for the RUI to be effective in a multi-modal system, seamless integration of the different components of the system is essential. For example, the RUI used in this project is a cute robotic teddy bear. Therefore, when it is used with a low pitched adult male voice, users cannot associate the teddy bear with the voice. Such a discrepancy between what users expect the robot's voice to be like and what they actually hear makes it apparent to the users that the speech system and the RUI are two separate projects that were later integrated. Thus, it is extremely important that the different components of a multi-modal system are assimilated in such a way that they are represented to the users as a single system, not two separate ones being forced to work together. The match between the voice and the robot was difficult to establish also because the room has a surround sound system; therefore, the room's speech was played over six speakers situated all around the IRoom, which made it even more difficult to

associate the voice with the robot. A single speaker located next to the robot would have reduced this association problem. Another example illustrating the importance of seamless integration is that some users noticed a perceptible delay between the time they issued a voice command and the time when the robot acknowledged it by nodding its head. Thus, it is crucial to analyze each part of a multi-modal system after the components are integrated to ensure that they all work together well to provide the desired end-effect to users.

We also learnt that if a RUI is to be used for physical communication with users, its gestures have to be specific. For example, when the robot in this project pointed towards the screen to draw the user's attention to it, most users did not understand the purpose of its right hand movement. The absence of fingers in the right hand was the main reason why users did not understand that the robot was pointing towards the screen. Therefore, it is very important that an appropriate RUI is used for such physical communication. Our robot was too small in size and lacked appropriate hand structure to actually point towards a specific object. If an RUI needs to point, it should be designed keeping these constraints in mind.

Lastly, we found out that it is extremely difficult to add a component to an already large and complex system. It is even more difficult to add a component in such a way that it is extensible and scalable. For example, currently, the room is capable

of conversing with users only when the `TeddyLightConversationAgent` or the `BookTeddyLightConversationAgent` are running.

8 Further Work

Even though the Robotic User Interface has added significantly to the interaction experience between the room and the user, many improvements are still desirable. For example, the direction and location of the objects in the room (for example, the door, the projectors and lamps) are hard coded. Therefore, if the robotic bear is moved to another location in the IRoom, or if it is moved to face another direction in the same location, it will no longer be able to point at the right direction when asked to show an object in the room.

The bear can be augmented with vision capabilities, perhaps using the current visual sensors present in the room, so that it can determine its own orientation, and then modify the ".mot" motion files accordingly before playing them so that it can point towards the correct direction.

Additionally, more sentences can be added to the grammar so that the robot can respond to a variety of requests from the user. This will also require handles to other agents in the room so that those agents can be used using voice command and confirmations.

Gesture grammars can be incorporated into the system so that the robot can

determine which gesture to make under which circumstances by looking up a table that matches situations with robotic gestures appropriate for those situations. Such a change would make the system more scalable. Currently, which “.mot” file needs to be played when is hard-coded in the software. Therefore, if we want the robot to perform a different gesture, the code needs to be modified. If, a gesture grammar is incorporated in the system, such changes to the code will no longer be necessary since the desired effect can be achieved by modifying the gesture grammar file.

Now that the room can hold conversation with users, in addition to the historical knowledge base system, any number of other knowledge based systems can be added to the room to assist users in variety of tasks.

Lastly, we have not explored the robotic teddy bear’s potential as a haptic input device in this project. One scenario where this might be helpful is as follows:

Bob asks the room to switch on the right projector. The system makes an error in voice recognition and thinks Bob wants the left projector switched on. The robot points towards the left projector and asks Bob if this is the projector he wants switched on. Instead of repeating his request in words, Bob moves the robot’s right arm until it points in the direction of the right projector and says “No, this is the projector I want turned on.” The room then turns on the correct projector.

Thus, further work that uses the robot also as an input device has the potential to further enrich interaction between users and the IRoom.

9 Conclusion

In this project, human-computer interaction with the Intelligent Room, a ubiquitous computing environment, was augmented by using the novel approach of integrating an Robotic User Interface with the room. This addition has significantly enhanced the interaction between the IRoom and its users, as evidenced from the user study results.

The room can now physically communicate with the user by using robotic gestures in response to the user's questions and commands. For example, the robot's hand gestures allow the room to point at different objects, and to greet the user by waving. Different objects in the room can be indicated by the movement of the robot's arms and nodding motions are used to acknowledge comprehension.

Dialogue with users was established by improving the natural language system of the room to further enrich interaction. The resulting mixed initiative architecture allows the room to ask users for clarification.

To demonstrate the enhanced capabilities of the room, we created an application case study that recommends historical fictions to users, depending on their

preferences. Both physical and speech communication is used in this system. The IRoom can determine user preferences by asking questions, and then use the robot to acknowledge the user's answers. Finally, the robot points towards a screen with pictures of the books recommended.

A user study established that the presence of the Robotic User Interface is generally appreciated by users, and that the robot, along with the conversational abilities of the room, has made interacting with the room a much more rewarding and fun experience.

Ubiquitous and human-centric computing environments are an inevitable part of the future. The work done in this project is hopefully a small step towards that intelligent future, where humans will be able to interact with their computers naturally and intuitively.

References

1. M. Coen, B. Phillips, N. Warshawsky, L. Weisman, S. Peters, and P. Finin. "Meeting the Computational Needs of Intelligent Environments: The Metaglu System". 1st International Workshop on Managing Interactions in Smart Environments (MANSE'99). pp. 201-212, Ireland: Springer-Verlag, 1999
2. B. Phillips. "Metaglu: A Programming Language for Multi-Agent Systems", MIT M.Eng Thesis, 1999
3. N. Warshawsky. "Extending the Metaglu Multi-Agent System". MIT M.Eng Thesis, 1999
4. D. Sekiguchi, M. Inami, and S Tachi. RobotPHONE: RUI for Interpersonal Communication. [Online Paper] Available at < <http://www.star.t.u-tokyo.ac.jp/projects/RobotPHONE/CHI2001-RobotPHONE.pdf>>
5. M. Weiser. The computer for the twenty-first century. Scientific American, pp. 94–104, Sep 1991
6. M. Weiser, and J. S. Brown. The coming age of calm technology. Technical report, Xerox PARC, 1996
7. M. Weiser. "The world is not a desktop". Interactions, pp. 7-8, January 1994
8. M. Weiser. "Hot Topics: Ubiquitous Computing". IEEE Computer, October 1993
9. M. Weiser. "Some Computer Science Problems in Ubiquitous Computing," Communications of the ACM, July 1993
10. W. Shakespeare. Hamlet. Act 3, Scene 4. Available at <<http://www-tech.mit.edu/Shakespeare/hamlet/full.html> >
11. S. Peters, G. Look, K. Quigley, H. Shrobe, and K. Gajos. "Hyperglue: Designing High-Level Agent Communication for Distributed Applications". MIT CSAIL TechReport, 2006

12. M. Coen. Sodabot: A software agent environment and construction system. Technical report, MIT AI Lab, June 1994
13. M. Coen. Building brains for rooms: Designing distributed software agents. In *Proceedings of the Ninth Conference on Innovative Applications of Artificial Intelligence. (IAAI97)*, 1997
14. M. Coen, H. Kautz, B. Selman, S. Ketchpel, and C. Ramming. An experiment in the design of software agents. In *Proceedings of the Twelfth National Conference on Artificial Intelligence, AAAI-94*, 1994
15. M. Torrance. Advances in human-computer interaction: The intelligent room. Working Notes of the CHI 95 Research Symposium, Aizu, Japan, may 1995
16. AIRE Homepage. [Online webpage] Available at <<http://aire.csail.mit.edu/>>
17. Project Oxygen. [Online webpage] Available at <<http://www.oxygen.lcs.mit.edu/index.html>>
18. T. Hammond, K. Gajos, R. Davis, and H. Shrobe. System for Capturing and Indexing Software Design Meetings. In International Workshop on Agents in Design (WAID'02), Aug 2002
19. R. Brooks. The intelligent room project. In Proceedings Cognitive Technology Conference (CT'97), Aizu, Japan, 1997
20. The Conversational Humanoid. [Online webpage] Available at <<http://www.media.mit.edu/gnl/projects/humanoid/>>
21. J. Cassell, T. Bickmore, M. Billingham, L. Campbell, K. Chang, H. Vilhjálmsón, and H. Yan. "Embodiment in Conversational Interfaces: Rea." Proceedings of the CHI'99 Conference, Pittsburgh, PA, pp. 520-527, 1999
22. J. Cassell, and H. Vilhjálmsón. "Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous." *Autonomous Agents and Multi-Agent Systems* 2(1): 45-64, 1999

23. C. Breazeal, and B. Scassellati. "Infant-like Social Interactions between a Robot and a Human Caretaker" To appear in *Special issue of Adaptive Behavior on Simulation Models of Social Agents*. [Online Paper] Available at <http://www.ai.mit.edu/projects/sociable/publications.html>
24. C. Breazeal, and B. Scassellati. "How to build robots that make friends and influence people" To appear in *IROS99*, Kyonjiu, Korea. [Online Paper] Available at <http://www.ai.mit.edu/projects/sociable/publications.html>
25. L. Morency, and T. Darrel . "From Conversational Tooltips to Grounded Discourse: Head Pose Tracking in Interactive Dialog Systems" *International Conference on Multimodal Interfaces*, pp. 32-37, College State, PA, October 2004
26. C. L. Sidner*, C. D. Kidd, C. Lee, and N. Lesh, "Where to Look: A Study of Human-Robot Engagement". In *Proceedings of Intelligent User Interfaces*, Portugal, 2004
27. A. C. Tamhane, D. D. Dunlop. *Statistics and Data Analysis from Elementary to Intermediate*. Prentice Hall, NJ. pp 249-254. 2000
28. C. S. Horstmann and G. Cornell. *Core Java*. Sun Microsystems Press. Prentice Hall. 2004
29. B. Johanson, and A. Fox. Extending tuplespaces for coordination in interactive workspaces. *J. Syst. Softw.*, 69(3):243–266, 2004
30. M. V. Kleek. *Intelligent Environments for Informal Public Spaces: the Ki/o Kiosk Platform*. Master's thesis, MIT, Feb 2003
31. *Oxygen: Pervasive, human-centered computing*. MIT Laboratory of Computer Science, Cambridge, Massachusetts, USA, May 2002
32. D. Sutherland. *RMI and java distributed computing*. Technical report, Sun Microsystems, Inc., Nov 1997
33. B. Kottahachchi. *ACCESS: Access Controls for Cooperatively Enabled Smart Spaces*. MIT M. Eng Thesis, Feb 2005
34. A. Kulkarni. *A Reactive Behavioral System for the Intelligent Room*. MIT M. Eng, Thesis, 2002

- 35.K. Koch. Situating a Natural Language System in a Multimodal Environment. MIT M.Eng Thesis, 2001
- 36.K. Gajos. A Knowledge-Based Resource Management System for the Intelligent Room. MIT M.Eng. Thesis, 2000
- 37.G. Look and S. Peters. Plan-Driven Ubiquitous Computing. Student Oxygen Workshop, 2003
- 38.M. Kleek. Supporting Informal Computer-Human Collaboration: The Kiosk Platform. Student Oxygen Workshop, 2002
- 39.H. Shrobe, K. Quigley, S. Peters, R. Kochman, M. Kleek, G. Look and K. Gajos. Building Software Infrastructure for Human-Centric Pervasive Computing. MIT Artificial Intelligence Laboratory Research Abstracts, September 2002
- 40.G. Look and H. Shrobe. Planlet: Supporting Plan-Aware Ubiquitous Computing Applications. MIT Artificial Intelligence Laboratory Research Abstracts, September 2002
- 41.A. Oh, R. Tuchinda, and Lin Wu. MeetingManager: A Collaborative Tool in the Intelligent Room. Student Oxygen Workshop, July 2001
- 42.A. Chang. A Help System for the Intelligent Room. Student Oxygen Workshop, 2001
- 43.K. Gajos and L. Weisman. Rascal -- A Resource Manager For E21. Proceedings of Student Oxygen Workshop, 2001
- 44.S. Peters. Infrastructure for Multi-User, Multi-Spatial Collaborative Environments. Student Oxygen Workshop, 2001
- 45.N. Hanssens, A. Kulkarni, R. Tuchinda, and T. Horton. Building Agent-Based Intelligent Workspaces. ABA Conference Proceedings, June 2002
- 46.A. Kulkarni. Design Principles of a Reactive Behavioral System for the Intelligent Room. Bitstream: The MIT Journal of EECS Student Research. 2002
47. S. Peters. Semantic Networks for Knowledge Representation in an IE. Student Oxygen Workshop. 2002

48. K. Gajos and A. Kulkarni. FIRE: An Information Retrieval Interface For Intelligent Environments. Proceedings of International Workshop on Information Presentation and Natural Multimodal Dialogue IPNMD December, 2001
49. K. Gajos, L. Weisman, and H. Shrobe. Design Principles for Resource Management Systems for Intelligent Spaces. Proceedings of the 2nd International Workshop on Self-Adaptive Software, 2001
50. M. Coen, B. Phillips, N. Warshawsky, L. Weisman, S. Peters, and P. Finin. Meeting the Computational Needs of Intelligent Environments: The Metaglu System. 1st International Workshop on Managing Interactions in Smart Environments, pp. 201—212, 1999
51. M. Coen, L. Weisman, K. Thomas, and M. Groh. A Context Sensitive Natural Language Modality for an Intelligent Room. 1st International Workshop on Managing Interactions in Smart Environments, pp. 68-79, 1999
52. M. Coen. The Future of Human-Computer Interaction, or How I learned to stop worrying and love my Intelligent Room. IEEE Intelligent Systems, March, 1999
53. M. Coen. Design Principles for Intelligent Environments. Fifteenth National Conference on Artificial Intelligence, pp. 547-554, 1998
54. L. Arnstein, R. Grimm, C. Hung, J. Kang, A. LaMarca, G. Look, S. Sigurdsson, J. Su, G. Borriello. Systems Support for Ubiquitous Computing: A Case Study of Two Implementations of Labscape. Proceedings of the First International Conference on Pervasive Computing. In Proceedings of the First International Conference on Pervasive Computing. August, 2002
55. R. Davis. Production Rules as a representation for a knowledge-based consultation Program. Artificial Intelligence Journal, 8:15-45, 1977
56. R. Duda, J. Gaschnig, and P. Hart. Model design in the prospector consultant system for mineral exploration, Expert Systems in the Microelectronic Age Michie (ed.), pp. 154—167, Edinburgh, University Press, 1979
57. A. Campbell, V. Hollister, R. Duda, and P. Hart. Recognition of a hidden mineral deposit by an artificial Intelligence program, Science. Sept, 1982

58. A. Oh, H. Fox, M. V. Kleeck, A. Adler, K. Gajos, L. Morency, T. Darrell, "Evaluating Look-to-Talk: A Gaze-Aware Interface in a Collaborative Environment", Proceedings of CHI 2002, Minneapolis, MN, April, 2002
59. E. Weinstein, K. Steele, A. Agarwal, and J. Glass, "LOUD: A 1020-Node Modular Microphone Array and Beamformer for Intelligent Computing Spaces", MIT/LCS Technical Memo MIT-LCS-TM-642, April, 2004
60. The Kismet Webpage. [Online webpage] Available at <http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html>
61. J. Nielsen. Finding Usability Problems through Heuristic Evaluation. In Proceedings of *Human Factors in Computing Systems CHI'92*, pp 373-80. New York: ACM Press, 1992

Appendix A: BookTeddyLightConversation.gram

```

grammar edu.mit.aire.applications.games.toy.TeddyLightConversation;

// contains turn on/off, brighten, etc...
import <edu.mit.aire.metagluе.speech.lib.verbs.deviceActionVerb>;

// The locations should be provided externally and not hard-coded in
// the grammar file.

<devices> = foo | boo;

<all> = ([all | both] the (lights | lamps)) {all};

<thing> = (<devices> | <all>) {thing};

// This is for "Turn the light by the window off"
<change> = turn | switch;
<on> = on {verb-turnOn};
<off> = off {verb-turnOff};
<state> = <on> | <off>;

// "Turn on the lamp by the widow"
public <sentenceOne> = <deviceActionVerb> [the] <thing> [(some | a bit |
a little bit) more];

// "Turn the lamp by the window on"
public <sentenceTwo> = <change> [the] <thing> <state>;

//Confirm if light needs to be switched on
public <yeslight> = (yes | yeah | yep) {yeslight};
public <nolight> = (no | nah | nope) {nolight};

//Reads poetry if user is bored
public <bored> = (I am bored | entertain me) {bored};

//Recommends books
public <bookrecommendation> = (I want a book) {bookrecommendation};

//Interacts with users
public <hello> = (hello | hi) { hello };
public <teddy> = (teddy | bear ) { teddy };

//Locations of objects in the IRoom
public <whereisdoor> = (show me the door) { whereisdoor };
public <whereisrightlight> = (show me the right light | show me the
right lamp | where is the right light | where is the right lamp)
{whereisrightlight};

public <whereisleftlight> = (show me the left light | show me the left
lamp | where is the left light | where is the left lamp)
{whereisleftlight};

```

```
public <whereisrightprojector> = (show me the right projector | where is  
the right projector ) {whereisrightprojector};
```

```
public <whereislefttprojector> = (show me the left projector | where is  
the left projector ) {whereisleftprojector};
```


Appendix B: Robotic User Interface Study - Directions for Participants. “Command and Control First” Version.

1. Please say “hello”
2. *Command and Control mode:*
Please say “switch on the light”
Please say “switch off the light”
3. *Interactive mode:*
Please say “switch on the light” again and answer any questions asked by the room

4. Please answer the following questions:

- i. Did you enjoy interacting with the bear when you said hello and the bear replied back and waved its hand? Please circle one of the numbers below.

1 = I hate it. 4 = neutral. 7 = it was very enjoyable

1 2 3 4 5 6 7

Why?

- ii. Did you find the questions asked by the room to confirm your intentions about switching on the light useful? Please circle one of the numbers below.

1 = not useful. 4 = neutral. 7 = very useful

1 2 3 4 5 6 7

Why?

Would you find it useful, if instead of asking confirmation for switching on the lights, the room asked for confirmation before doing an important and irreversible action?

1 = not useful. 4 = neutral. 7 = very useful

1 2 3 4 5 6 7

iii. Would you rather interact with the room in Command and Control mode or Interactive mode? Please circle one.

Command and Control

Interactive

5. Please say “recommend a book for me”
6. Please answer the following question:
- i. Did you look at the screen showing the picture of the recommended book? Circle one.
- YES NO
7. Please say “recommend a book for me” again and answer any questions asked by the room.
8. Please answer the following questions:
- i. Did you find the bear pointing its hand towards the screen useful?
- YES MAYBE NO
- ii. Did you find the bear’s acknowledgement to your answers by nodding its head useful? Please circle one of the numbers below.
- 1 = not useful. 4 = neutral. 7 = very useful
- 1 2 3 4 5 6 7
- Why?

-
-
- iii. Did you generally like the presence of the bear in the Intelligent Room?
- YES MAYBE NO
- iv. Did the presence of the bear enhance your experience in interacting with the room?

YES MAYBE NO

- v. Did conversing with the room while it deduced your book recommendation enhance your experience in interacting with the room?

YES MAYBE NO

Appendix C: Robotic User Interface Study - Directions for Participants. “Interactive First” Version.

1. Please say “hello”
2. *Interactive mode:*
Please say “switch on the light” and answer any questions asked by the room
Please say “switch off the light”
3. *Command and Control mode:*
Please say “switch on the light” again

4. Please answer the following questions:

- i. Did you enjoy interacting with the bear when you said hello and the bear replied back and waved its hand? Please circle one of the numbers below.

1 = I hate it. 4 = neutral. 7 = it was very enjoyable

1 2 3 4 5 6 7

Why?

- ii. Did you find the questions asked by the room to confirm your intentions about switching on the light useful? Please circle one of the numbers below.

1 = not useful. 4 = neutral. 7 = very useful

1 2 3 4 5 6 7

Why?

Would you find it useful, if instead of asking confirmation for switching on the lights, the room asked for confirmation before doing an important and irreversible action?

1 = not useful. 4 = neutral. 7 = very useful

1 2 3 4 5 6 7

- iii. Would you rather interact with the room in Command and Control mode or Interactive mode? Please circle one.

Command and Control

Interactive

5. Please say “recommend a book for me” and answer any questions asked by the room.

6. Please answer the following questions:

i. Did you find the bear pointing its hand towards the screen useful?

YES MAYBE NO

ii. Did you find the bear’s acknowledgement to your answers by nodding its head useful? Please circle one of the numbers below.

1 = not useful. 4 = neutral. 7 = very useful

1 2 3 4 5 6 7

Why?

7. Please say “recommend a book for me” again.

8. Please answer the following question:

i. Did you look at the screen showing the picture of the recommended book? Circle one.

YES NO

ii. Did you generally like the presence of the bear in the Intelligent Room?

YES MAYBE NO

iii. Did the presence of the bear enhance your experience in interacting with the room?

YES MAYBE NO

- iv. Did conversing with the room while it deduced your book recommendation enhance your experience in interacting with the room?

YES MAYBE NO

Appendix D: Data Analysis for User Study

Question Analyzed: Did you enjoy interacting with the bear when you said hello and the bear replied back and waved its hand? Please circle one of the numbers below.

1 = I hate it. 4 = neutral. 7 = it was very enjoyable

1 2 3 4 5 6 7

Table 1.

H0: Mean ranking ≤ 4	
Users either did not enjoy interacting with the bear, or were neutral towards it	
H1: Mean ranking > 4	
Users enjoyed interacting with the bear	
Sample Mean (m)	5.897
Sample Standard Deviation (s.d)	1.012
Sample Variance	1.025
Number of samples (n)	29
$t = (m - 4) / (s.d / \sqrt{n})$	10.09
T for Degree of freedom = 28 Alpha = 0.05	1.701
Conclusion	$t > t$ for $df = 28$, $\alpha = 0.08$
	H0 cannot be rejected
	Thus, sample mean is statistically significantly greater than 4
	Users enjoyed interacting with the bear

Question Analyzed: Did you find the questions asked by the room to confirm your intentions about switching on the light useful? Please circle one of the numbers below.

1 = not useful. 4 = neutral. 7 = very useful

1 2 3 4 5 6 7

Table 2.

H0: Mean ranking ≤ 4	
Users did not find the confirmation question for switching lights on useful, or are neutral towards it	
H1: Mean ranking > 4	
Users found the confirmation question for switching lights on useful	
Sample Mean (m)	4.034
Sample Standard Deviation (s.d)	1.955
Sample Variance	3.820
Number of samples (n)	29
$t = (m - 4) / (s.d / \sqrt{n})$	0.095
T for Degree of freedom = 28 Alpha = 0.05	1.701
Conclusion	$t < t$ for $df = 28$, $\alpha = 0.08$
	H0 cannot be rejected
	Thus, sample mean is not statistically significantly different from 4
	Users did not find the confirmation question for switching lights on useful

Question Analyzed: Would you find it useful, if instead of asking confirmation for switching on the lights, the room asked for confirmation before doing an important and irreversible action?

1= I hate it. 4 = neutral. 7 = it was very enjoyable

1 2 3 4 5 6 7

Table 3.

H0: Mean ranking ≤ 4	
Users either do not think confirmation questions for important and irreversible actions taken by the room are useful, or are neutral towards it	
H1: Mean ranking > 4	
Users think confirmation questions for important and irreversible actions taken by the room are useful	
Sample Mean (m)	6.571
Sample Standard Deviation (s.d)	0.836
Sample Variance	0.698
Number of samples (n)	28
$t = (m - 4) / (s.d / \sqrt{n})$	16.28
T for	1.703
Degree of freedom = 27	
Alpha = 0.05	
Conclusion	$t > t$ for $df = 27$, $\alpha = 0.08$
	H0 can be rejected
	Thus, sample mean is statistically significantly greater than 4
	Users think confirmation questions for important and irreversible actions are useful

Question Analyzed: Did you enjoy interacting with the bear when you said hello and the bear replied back and waved its hand? Please circle one of the numbers below.

1 = I hate it. 4 = neutral. 7 = it was very enjoyable

1 2 3 4 5 6 7

Table 4.

H0: Mean ranking ≤ 4 Users either did not find acknowledgement via the robot's head nod movement useful or were neutral towards it	
H1: Mean ranking > 4 Users found acknowledgement to their voice input via the robot's head movement useful	
Sample Mean (m)	4.586
Sample Standard Deviation (s.d)	1.722
Sample Variance	2.966
Number of samples (n)	29
$t = (m - 4) / (s.d / \sqrt{n})$	1.833
T for Degree of freedom = 28 Alpha = 0.05	1.701
Conclusion	$t > t$ for $df = 28$, $\alpha = 0.08$ H0 can be rejected Thus, sample mean is statistically significantly greater than 4 Users found the acknowledgement to their voice input via the robotic user interface's head nod movement useful