Smart Visitor Guidance Robot with Embedded Touch Screen Panel

Nasharibha Lyngdoh[1], Kavitha. B. C [2]

[1]M.Tech, Embedded System, Department of Electronics and Communication Engineering, Hindustan University, Chennai, India

[2]Assistant Professor, Department of Electronics and Communication Engineering, Hindustan University, Chennai, India

Abstract: ***The study of this paper provides an in-depth analysis and practical solution to the problem of designing and implementing a human–robot team for simple conversational interactions. Awareness of time is particularly important because people have a low tolerance for long pauses in conversation. Robot receptionist is an automated robotic reception system that takes the role of a receptionist and guides the incoming visitors by queuing them to their respective destinations/officials by showing route maps with the help of touch screen color panel and multilingual voice generation. It is also capable of recording the employee signature and does real time signature verification to maintain a digital attendance book.***

Keywords: Human–robot interaction, Receptionist robot, speech recognition, behavior control.

1. Introduction

Autonomous robotics is a research field which has been in development since the middle of the 20th century and it is currently one of the main areas of interest within the field of robotics. Even though great breakthroughs have been achieved throughout the years, this area still has a long way to go, as much in terms of sensory, mechanical, and mobility capabilities as well as in the artificial intelligence and decision making domain, before it can achieve efficient and flexible behaviors comparable to the ones observed in animals and humans.

There are growing demands on robotics and automation in the service sector for many years due to man power shortage. Some need more cost effective systems, while others do more human friendly communication and interactive interface. The system will require more improved functions for physical, skill and intelligence support and assistance. Certainly industrial robots have helped people from the routine type of assembly and manufacturing for better production system. Thus robotics has made remarkable progress to industry.

The links between language and action system have recently received a lot of attention in psychology, neuroscience, and robotics. People's unpredictability is very much related with the different reactions they can express towards an unusual and unknown identity such as an automaton. For this reason, today's key for developing a successful people interacting agent might not be employing an extremely complex decision system that seeks to cover all possible situations, but rather to use a human being's almost unlimited self adaptive capacity to adjust itself to the robot platform. This can be achieved by providing, on the one hand, the means to help them feel comfortable with the whole situation and on the other hand, to guide them through the process of interaction by initially taking the initiative to start a communication, and then directing it to expected destination.

2. Related Work

Engagement and Collaboration in different manners are discussed in this section. Collaboration is a process by which two or more participants coordinate their actions toward achieving shared goals. Engagement is a fundamental process which governs all human interaction and has common features across a very wide range of interaction circumstances. Engagement and collaboration between a human and a robot has been previously investigated by Sidner et al [2]. They performed a study wherein a human interacted with a robot that was able to track faces (mutual facial gaze and directed gaze) while explaining a collaborative task. They found that the number of times the human looked back at the robot was significantly greater when the robot tracked the human's face and performed pointing gestures versus when it did not.

Bohus and Horvitz [7] investigated how to automatically learn engagement behaviors in multi-party situations in which a virtual agent is able to initiate, terminate, or suspend engagement with multiple humans in real time. They have developed predictive models from observations, and implemented the models in a system which uses machine learning to adapt to behaviors of participants, as well as to the environment. They have also shown experimental results which validate the approach with respect to learning to detect engagement intentions. Furthermore, they have created components able to sense engagement state, actions, and intentions, make engagement decisions, and render behaviors in a virtual agent. To evaluate these components, they conducted a preliminary observational study using an on-screen virtual agent which interacts with participants to play a trivia game. This system was able to track the engagement of multiple participants in an uncontrolled environment.

Mutlu et al. [4] are interested in the role that gaze play in forming participant roles during an interaction, specifically how a humanoid robot can use different gaze behaviors to establish the participant's role in the interaction. They have made use of three participant roles: addressee, bystander, and over hearer. An addressee is addressed by the speaker during the interaction, and is also a contributing member of the interaction. In an interaction where a bystander is present, the robot acknowledges the presence of a bystander; however, it does not directly speak to him. An over hearer is a participant that is not addressed, or acknowledged by the robot. This experiment showed that the gaze behaviors altered the subject's participation during the interaction with respect to their attention to the conversation, feelings of belonging to the group, and feelings toward the robot.

3. Existing System

An observational study of human engagement behavior is shown in which two participants sat across a table conversation with each other. During the periods of interest, coding is made where each person was looking at objects from the beginning and end of each person's speech. Connection between humans can either succeed or fail. Success occurs when the responder performs his intended response, and failure occurs when the responder does not respond.

3.1. Directed Gaze

In directed gaze, one of the initiator looks and optionally points at some object or group of objects following which the responder looks at the same objects. The initiator intends to bring the indicated object to the responder's attention, i.e., to make the object more salient in the interaction. This event is often synchronized with the initiator referring to the object in speech.

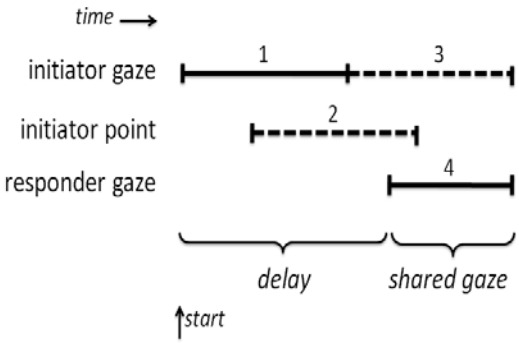

Figure 1: Time Line for Directed Gaze

By turning his gaze where directed, the responder intends to be cooperative and thereby signals his desire to continue the interaction. It is hard to accurately point at something without looking to see where it is located although one can easily imagine an exception to this situation, and other situations, where a person can point without looking. There are several different configurations of the hand in pointing, such as extended first finger, open hand (palm up or palm down), and a circular waving motion (typically over a group of objects). The initiator usually maintains the pointing at least until the responder starts looking at the indicated object [7]. However, the initiator may stop looking at the indicated object before the responder starts looking, especially when there is pointing. This is often because the initiator looks at the responder's face assumedly whether the responder has directed his gaze yet. Finally, there may be a period of shared gaze, i.e., a period when both the initiator and responder are looking at the same object. Shared gaze has been documented as an important component of human interaction.

3.2. Mutual Facial Gaze

Mutual facial gaze has a time line as shown in the figure and is similar to directed gaze, but simpler, since it does not involve pointing. The event starts when the initiator looks at the responder's face. After a delay, the responder looks at the initiator's face, which starts the period of mutual facial gaze. The delay can be zero, which occurs when both parties simultaneously look at each other. The intentions underlying mutual facial gaze are less clear than those for directed gaze. Both the initiator and responder in mutual facial gaze engage in this behavior because they intend to maintain the engagement process. Mutual facial gaze does however have other interaction functions. For example, it is typical to establish mutual facial gaze at the end of a speaking turn. Mutual facial gaze is often referred to as "making eye contact."

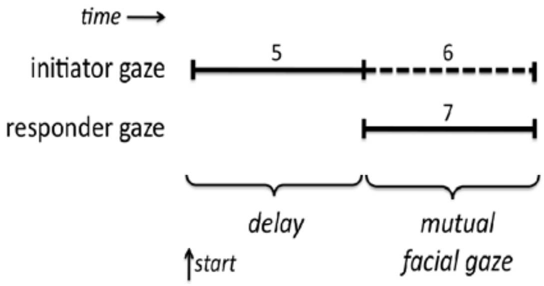

Figure 2: Time Line for Mutual Facial Gaze

3.3. Adjacency Pair

In linguistics, an adjacency pair consists of two utterances by two speakers, with minimal overlap or gap between them, such that the first utterance provokes the second utterance. A question-answer pair is a classic example of an adjacency pair. This concept includes both verbal (utterances) and non-verbal communication acts. So for example, a nod could be the answer to a question, instead of a spoken "yes." Adjacency pairs often overlap with the gestural connection events, such as directed gaze and mutual facial gaze. The simple time line for an adjacency pair is shown in the figure. First the initiator communicates what is called the first turn. Then there is a delay, which could be zero if the responder starts talking before the initiator finishes. Then the responder communicates what is called the second turn. In some conversational circumstances, this could also be followed by a third turn [1] in which the initiator repairs the responder's misunderstanding of the original communication.

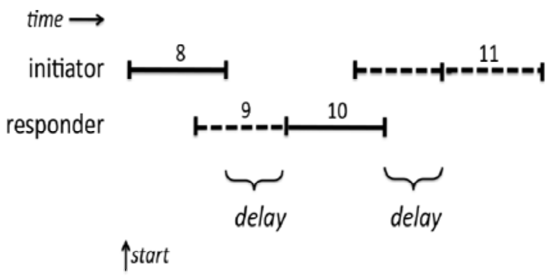

Figure 3: Time Line For Adjacency Pair

3.4. Backchannels

The figure shows the time line for backchannels. A backchannel is an event in which one party (the responder) directs a brief verbal or gestural communication [2] back to the initiator during the primary communication from the initiator to the responder. Examples of backchannels are nods and/or saying “uh, huh”. Backchannels are typically used to communicate the responder's comprehension of the initiator's communication and/or desire for the initiator to continue. Unlike the other three connection event types, the start of a backchannel event is defined as the start of the responder's behavior and this event has no concept of delay.

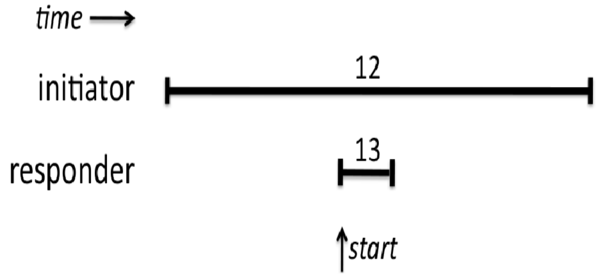

Figure 4: Time Line for Backchannels

3.5. Limitations in the Existing System

The major problems that arise mostly in educational institutions, public banks, conference halls are as the new visitors often don’t find the guidance to reach their destination. Receptionists are there but they need to be trained and must be provided wages. Human behaviors are unpredictable and they can go wrong in guiding a visitor with their punctuality and politeness.

4. Proposed System

In general in software development, the key to making a reusable component is careful attention to the setting in which it will be used and the division of labor between the component and the rest of the computational environment in which it is embedded.

4.1. Human-Robot Setting

The figure shows the setting of our current architecture and implementation, which mirrors the setting of the human engagement study, namely a human and a humanoid robot with a table of objects between them. Either the robot or the human can be the initiator (or responder) in the connection event time lines shown in the previous section. Like the engagement maintenance part of the human study, mobility is not part of this setting. Unlike the human study, we are not dealing here with manipulation of the objects or changes in stance (e.g., turning the body
to point to or manipulate objects on the side part of the table). Both the human and the robot can perform the following behaviors and observe them in the other:
1. Look at the other's face, objects on the table or “away",
2. Point at objects on the table,
3. Nod the head (up and down), and
4. Shake the head (side to side).

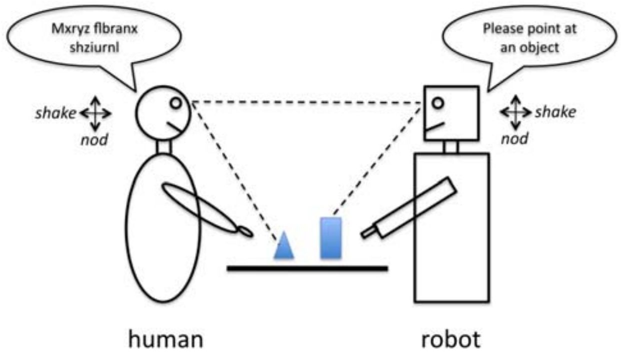

Figure 5: Human Robot Interaction Model

The project proposed here aims to design a smart visitor guidance robot and this system is used to overcome the existing drawbacks. This system consists of two units namely,
- Robotic Reception Unit.
- Visitor Alert Unit.

In this system the robotic reception unit acts as a master and visitor alert unit act as a slave.

5. Block Diagram

5.1. Visitor Alert Unit-Slave

In this unit the robot scans the space before it using sensors and if it finds a person moving it greets them with a welcome message in a clear MP3 quality voice. The robot will first separate the visitors into regular employee or a general visitor and greets them with a welcome message in a clear MP3 quality voice. It shows the greetings image and then renders an onscreen keypad with letters and prompts the user to enter his signature on the embedded touch screen panel display if he is an employee or just enter his name and the purpose of visit if the person is a general visitor. The robot also assigns a visitor ID number to each of them. The entered name along with the time will be saved in memory.

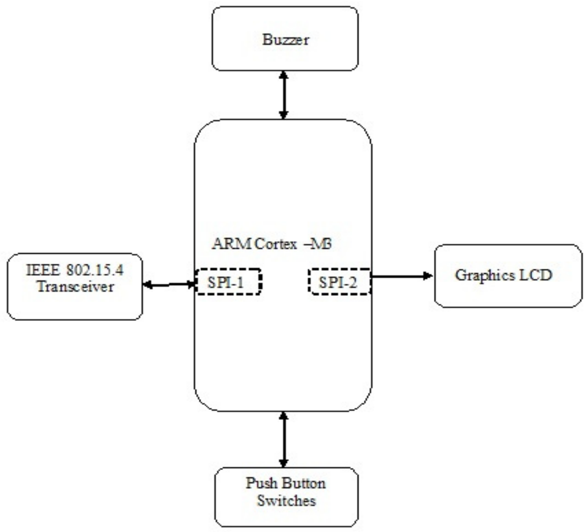

Figure 6: Visitor Alert Unit – Slave

5.2. Robotic Reception Unit- Master

After the slave unit is processed, the robot will present a list of names (ex: principal, chairman, etc..,) whom the visitor want to meet or a list of places (medical dispensary, seminar room etc..,) where the visitor want to go. It asks the user to touch their choice. Once the user chooses the right option on the screen, the system will acknowledge with voice and it shows a route map of that building for the user to reach his destination spot or the place where the respective official would be normally available. Before this process more importantly it sends the visitor request through a wireless network to other slave units fixed near the official persons that will display the visitor name and his purpose. The official can communicate with the robot (call or hold) using the push buttons on his device and the robot will announce the visitor ID audibly and manage them in the reception. Pressing a button on the robot will cause the system to show the signature employee names along with the corresponding time.

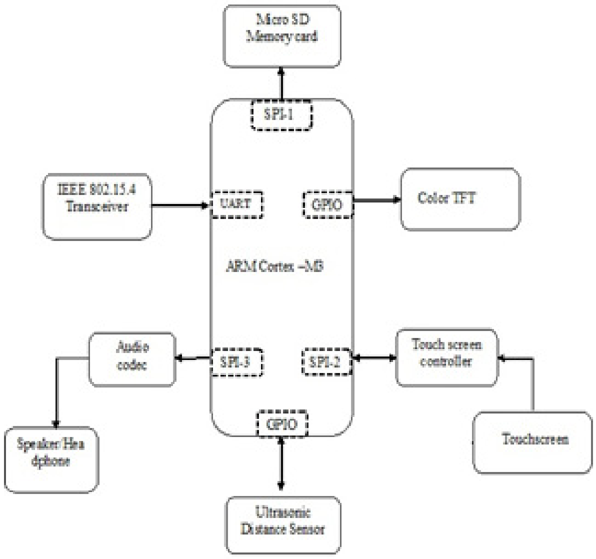

Figure 7: Robotic Reception Unit – Master

5.3. Project Advantages

- No need to pay salary since this is a one-time investment for buyers.
- Robotic receptionist will be always polite, punctual, more pleasant and careful with the customers and visitors.
- Young girls are often trained only to greet and guide the customer in shops. The reception robot can replace them.
- An aging population means, labor is going to be in severe shortage. This will create a strong demand for these automated robotic receptionists.
- The technology could be easily integrated into receptions of college, university, commercial buildings, libraries, medical hospitals, department stores, jewel shops, clothing stores etc.

6. Hardware Description

6. 1. ARM Cortex-M3

The central Cortex-M3 core is based on the Harvard architecture characterized by separate buses for instructions and data. The processor differs from the von Neumann architecture based ARM7 family of processors which use the same signal buses and memory for both instructions and data. Cortex-M3 processor can perform many operations in parallel, speeding application execution. The core pipeline has 3 stages: Instruction Fetch, Decode and Execute. The Cortex-M3 processor is a 32-bit processor, with a 32-bit wide data path, register bank and memory interface. There are 13 general-purpose registers, two stack pointers, a link register, a program counter and a number of special registers including a program status register.

6.2. LCD Controller/Driver

- Single chip LCD controller/driver
- 48 row, 84 column outputs
- Display data RAM 48x84 bits
- On-chip Generation
- External RES (reset) input pin
- Serial interface maximum 4.0 M bits/s
- CMOS compatible inputs
- Logic supply voltage range 2.7 to 3.3 V
- Low power consumption, suitable for battery operated systems
- Temperature range: -25 to +70 °C.

Figure 8: LCD Controller

6.3. IEEE 802.15.4

IEEE 802.15.4 deals with only PHY layer and portion of Data link layer. The higher-layer protocols are left to industry and the individual applications. The Zigbee Alliance is an association of companies involved with building higher-layer standards based on IEEE 802.15.4. This includes network, security, and application protocols. The services which network layer provides are more challenging to implement because of low power consumption requirement. In Data link layer IEEE 802 splits DLL into MAC and LLC sub layers. MAC provides data and management services to upper layers.

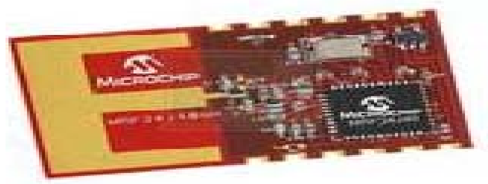

Figure 9: IEEE 802.15.4 Transceiver

6.4. Micro-SD Card

- Small form factor
- Non volatile Flash Memory
- Interfaced via SPI Protocol
- Maximum data rate is 25Mbps
- Supports capacity upto 2GB
- Allows FAT-32 formatting for easy file management
- Used as mass storage device in portable embedded system
- Developed by SD Card Association

6.5. Touch screen

- 4-wire touch screen interface
- Ratio metric conversion
- Single supply: 2.7v to 5v
- Up to 125khz conversion rate
- Serial interface
- Programmable 8- or 12-bit resolution
- 2 auxiliary analog inputs
- full power-down control

6.6. Audio codec

- Decodes and Encodes Ogg Vorbis with software plug-in
- Streaming support for MP3 and WAV
- Ear Speaker Spatial Processing
- Operates with a single 12-13MHz clock.
- Internal PLL clock multiplier
- Low-power operation
- High-quality on-chip stereo DAC with no phase error between channels
- Zero-cross detection for smooth volume change
- Stereo earphone driver capable of driving 30- load
- Quiet power-on and power-off

6.7. Software Tools Used

- Programming Language: Embedded C
- Development Tool: LPCXpresso IDE (Eclipse based)

6.8. Embedded Protocols Used

I2C, SPI, UART

6.9. Software Libraries Used

- Graphics Library for Color TFT LCD
- Graphics Library for Monochrome Graphics LCD
- Touchscreen Controller Driver via SPI protocol
- FatFs FAT-32 File System Library
- Micro-SD Card Driver Library via SPI
- Audio File Decoder Software
- IEEE 802.15.4 Wireless Network Software Stack via SPI
- Cortex-M3 Peripheral Device Driver Library

7. Conclusion

Engagement is the process that governs human interaction, and thus in order for robots and humans to collaborate in a natural way, robots must have a model of engagement. The robot must be able to both recognize the human's engagement behaviors, and properly generate its own engagement behaviors at the correct times. Holroyd has developed a model for engagement generation for use in a robotic system [1]. These models were implemented as open-source ROS nodes with the hope of benefiting the human-robot interaction research community. Holroyd, C.Rich evaluated the validity of models through a between-subjects three-arm experiment where humans interacted with humanoid robot to create puzzles [2].

8. Future Enhancement

The future enhancement includes that the Receptionist robot provides the perfect platform to support and test algorithms/solutions in several research areas of autonomous robotics (e.g., navigation, autonomous control, human-robot interaction). It would also be interesting if, upon encountering an unknown user, the Robot could autonomously capture a visual sample of the user's face (using the Robot's face detection capabilities to segment it), or a sound sample of the user's speech, and extract unique features from this sample data in order to add this person to the Receptionist's database. Sound and visual data could also be used to detect the user moods/expressions, providing the means for the Robot to adapt its own interaction approach accordingly.

References

[1] A. Holroyd. Generating engagement behaviors in human-robot interaction. Master's thesis, Worcester Polytechnic Institute, Worcester, Mass., USA, December 2010.

[2] A. Holroyd., C. Rich, C. Sidner, and B. Ponsler. Generating connection events for human robot collaboration. Submitted to ACM Conf. on Human-Robot Interaction, 2011.

[3] Boiney, L. Team "Decision Making in Time-Sensitive Environments" In Proceedings Command and Control Research and Technology McLean,VA, June 2005.

[4] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In Proc. ACM Conf. on Human-Robot Interaction, San Diego, Calif., USA, 2009.

[5] C.Rich, B. Ponsler, A. Holroyd, and C. Sidner. Recognizing engagement in human-robot interaction. In Proc. ACM Conf. on Human-Robot Interaction, Osaka, Japan, Mar. 2010.

[6] David T. Field,John A. Groeger "Temporal interval production and short-term memory" University of Surrey, Guildford, England Psychonomic Society, Inc. 2004.

[7] D. Bohus and E. Horvitz. Learning to predict engagement with a spoken dialog in openworld settings. In Proceedings of the SIGDIAL 2009 Conference, London, UK, September 2009. Association for Computational Linguistics, pp. 244-252.

[8] D. Bohus and E. Horvitz. Models for multiparty engagement in open-world dialog. In Proceedings of the SIGDIAL 2009 Conference, London, UK, September 2009. Association for Computational Linguistics, pp. 225-234.

[9] J. W. Crandall and M. A. Goodrich, "Principles of Adjustable Interactions." In proceedings of the 2002 AAAI Fall Symposium Human-Robot Interaction Workshop, North Falmouth, MA, USA, November 15-17, 2002.

[10] K. Hayashi, D. Sakamoto, T. Kanda, M. Shiomi, S. Koizumi, H. Ishiguro, T. Ogasawara, and N. Hagita, "Humanoid robots as a passive social medium—A field experiment at a train station," in Proc. 2nd ACM/IEEE Int. Conf. Hum.–Robot Interact., 2007, pp. 137–144.

Author Profile

Kavitha. B. C obtained her B.Tech degree Electronics and Communication Engineering from Cochin University of Science and Technology in the year 2000 and ME degree in Communication Systems from Hindustan College of Engineering, Anna university in 2007. Currently she is associated with Hindustan University as Assistant Professor. She has a total of 10 years teaching experience and her field of interest includes sensor networks and optical communication. She is also a lifetime member of IETE.

Nasharibha Lyngdoh obtained her B.Tech degree in Electronics and Communication Engineering from Dr MGR Educational and Research Institute, Chennai, Tamilnadu and she is currently pursuing M.Tech in Embedded Systems from Hindustan University, Chennai, Tamilnadu. Her field of interest includes Embedded Systems and Telecommunication.