4

Augmented Reality for Human-Robot Collaboration

Scott A. Green[1, 2], Mark Billinghurst[2], XiaoQi Chen[1] and J. Geoffrey Chase[1]
[1]Department of Mechanical Engineering, University of Canterbury
[2]Human Interface Technology Laboratory, New Zealand (HITLab NZ)
New Zealand

1. Introduction

Although robotics is well established as a research field, there has been relatively little work on human-robot collaboration. This type of collaboration is going to become an increasingly important issue as robots work ever more closely with humans. For example, in space exploration, recent research has pointed out that to reduce human workload, costs, fatigue driven error and risk, intelligent robotic systems will need to be a significant part of mission design (Fong and Nourbakhsh 2005). Fong and Nourbakhsh also observe that scant attention has been paid to joint human-robot teams, and that making human-robot collaboration natural and efficient is crucial to future space exploration. NASA's vision for space exploration stresses the cultivation of human-robotic systems (NASA 2004). In addition, companies such as Honda (Honda 2007), Toyota (Toyota 2007) and Sony (Sony 2007) are interested in developing consumer robots that interact with humans in the home and workplace. Finally, the Cogniron project (COGNIRON 2007), MIT Media lab (Hoffmann and Breazeal 2004) and the Mitsubishi Electric Research Laboratories (Sidner and Lee 2005), among others, are currently conducting research in human-robot interaction (HRI). HRI has become a research field in its own right, as shown by the 2006 inaugural conference for HRI with the theme Toward Human-Robot Collaboration (HRI2006 2006).

Research into human-human communication can be used as a starting point in developing a robust human-robot collaboration system. Previous research with humans has shown that grounding, situational awareness, a common frame of reference and spatial referencing are vital in effective communication. Clearly, there is a growing need for research on human-robot collaboration and models of communication between humans and robotic systems.

Augmented Reality (AR) is a technology for overlaying three-dimensional virtual graphics onto the users view of the real world. It also allows for real time interaction with these virtual graphics, enabling a user to reach into the augmented world and manipulate it directly. Augmented Reality could be used to overlay virtual imagery on a real robot and so display the internal state and intentions of the robot. Thus AR can bridge the divide between human and robotic systems and could enable effective human-robot collaboration.

In this chapter an overview of models of human-human collaboration is provided and how these models could be used to develop a model for human-robot collaboration is investigated. The field of human-robot interaction is reviewed and how it fits into a model

Source: Human-Robot Interaction, Book edited by Nilanjan Sarkar,
ISBN 978-3-902613-13-4, pp.522, September 2007, Itech Education and Publishing, Vienna, Austria

of human-robot collaboration is explored. Augmented Reality is introduced and then the effective use of AR for collaboration is discussed. The potential avenues for creating natural human-robot collaboration through spatial dialog utilizing AR are then investigated. Then the work that has been done in this area is discussed and a holistic architectural design for human-robot collaboration based on Augmented Reality is presented.

2. Communication and Collaboration

In this work, collaboration is defined as "working jointly with others or together especially in an intellectual endeavor". Nass *et al.* (Nass, Steuer et al. 1994) noted that social factors governing human-human interaction equally apply to human-computer interaction. Therefore, before research in human-robot collaboration is described, human-human communication is briefly reviewed and a model for human-human collaboration is presented. This model provides an understanding of the needs of an effective human-robot collaborative system.

2.1 Human-Human Collaboration

There is a vast body of research relating to human–human communication and collaboration. People use speech, gesture, gaze and non-verbal cues to communicate in the clearest possible fashion. In many cases, face-to-face collaboration is also enhanced by, or relies on, real objects or parts of the user's real environment. This section briefly reviews the roles conversational cues and real objects play in face-to-face human-human collaboration. This information is used to derive a set of guidelines for attributes that robots should have to effectively support human-robot collaboration.

A number of researchers have studied the influence of verbal and non-verbal cues on face-to-face communication. Gaze plays an important role by providing visual feedback, regulating the flow of conversation, communicating emotions and relationships, and improving concentration by restriction of visual input (Argyle 1967; Kendon 1967). In addition to gaze, humans use a wide range of non-verbal cues, such as nodding (Watanuki, Sakamoto et al. 1995), gesture (McNeill 1992), and posture (Cassell, Nakano et al. 2001). In many cases, non-verbal cues can only be understood by considering co-occurring speech, such as when using deictic gestures for pointing at something (Kendon 1983). In studying human behaviour it was observed that before conversational partners pointed to an object, they always looked in the direction of the object first (Sidner and Lee 2003). This result suggests that a robot needs to be able to recognize and produce non-verbal communication cues to be an effective collaborative partner.

Real objects and interactions with the real world can also play an important role in collaboration. Minneman and Harrison (Minneman and Harrison 1996) showed that real objects are more than just a source of information, they are also the constituents of collaborative activity, create reference frames for communication and alter the dynamics of interaction. In general, communication and shared cognition are more robust because of the introduction of shared objects. Real world objects can be used to provide multiple representations resulting in increased shared understanding (Clark and Wilkes-Gibbs 1986). A shared visual workspace enhances collaboration as it increases situational awareness (Fussell, Setlock et al. 2003). To support these ideas, a robot should be aware of its surroundings and the interaction of collaborative partners within those surroundings.

Clark and Brennan (Clark and Brennan 1991) provide a communication model to interpret collaboration. Conversation participants attempt to reach shared understanding or common ground. Common ground refers to the set of mutual knowledge, shared beliefs and assumptions that collaborators have. This process of establishing shared understanding, or "grounding", involves communication using a range of modalities including voice, gesture, facial expression and non-verbal body language. Thus, it is evident that for a human-robot team to communicate effectively, all participants will have to be able to easily reach common ground.

2.2 Human-Human Collaboration Model

This chapter investigates human-robot collaboration that is based on a human-human collaboration model, which itself is based on the following three components:

- The communication channels available.
- The communication cues provided by each of these channels.
- The affordances of the technology that affect the transmission of these cues.

There are essentially three types of communication channels available: audio, visual and environmental. Environmental channels consist of interactions with the surrounding world, while audio cues are those that can be heard and visual cues those that can be seen. Depending on the technology medium used communication cues may, or may not, be effectively transmitted between collaborators.

This model can be used to explain collaborative behavior and to predict the impact of technology on collaboration. For example, consider the case of two remote collaborators using text chat to collaborate. In this case, there are no audio or environmental cues. Thus, communication is reduced to one content heavy visual channel: text input. Predictably, this approach has a number of effects on communication: less verbose communication, use of longer phrases, increased time to reach grounding, slower communication and fewer interruptions.

Taking each of the three communication channels in turn, characteristics of an effective human-robot collaboration system can be identified. The robotic system should be able to communicate through speech, recognizing audio input and expressing itself through speech, highlighting a need for an internal model of the communication process. The visual channel should allow the robot to recognize and interpret human non-verbal communication cues and allow the robot to express some non-verbal cues that a human could naturally understand. Finally, through the environmental channel the robot should be able to recognize objects and their manipulation by the human, and be able itself to manipulate objects and understand spatial relationships.

2.3 Summary

This section discussed the general elements of collaboration and then covered how those aspects are seen in human-human collaboration. Human-robot collaboration will require the same fundamental elements, but with different context and avenues. This may well introduce limitations in some channels and increase fidelity in others. The next section, therefore, introduces the robot element and the ways in which robots are used, or might be used, for collaboration with humans.

3. Human-Robot Interaction

3.1 Robots As Tools

The simplest way robots can be used is as tools to aid in the completion of physical tasks. Although there are many examples of robots used in this manner, a few examples are given that highlight human-robot interaction and provide insight into collaboration. For example, to increase the success rate of melon harvesting, a human-robot collaborative system was implemented by (Bechar and Edan 2003). Results indicated that a human operator working with a robotic system with varying levels of autonomy resulted in significantly improved harvesting. Depending on the complexity of the harvesting environment, varying the level of autonomy of the robotic harvester increased positive detection rates by up to 7% from the human operator working alone and by as much as 20% compared to autonomous robot detection alone.

Robots are often used for hazardous tasks. For instance, the placement of radioactive waste in centralized intermediate storage is best completed by robots as opposed to humans (Tsoukalas and Bargiotas 1996). Robotic completion of this task in a totally autonomous fashion is desirable, but not yet achievable due to the dynamic operating conditions. Radiation surveys are initially completed through teleoperation, the learned task is then put into the robots repertoire so the next time the task is to be completed the robot will not need instruction. A dynamic control scheme is needed so that the operator can observe the robot as it completes its task, and when the robot needs help, the operator can intervene and assist with execution. In a similar manner, Ishikawa and Suzuki (Ishikawa and Suzuki 1997) developed a system to patrol a nuclear power plant. Under normal operation the robot is able to work autonomously, however in abnormal situations the human must intervene to make decisions on the robots behalf. In this manner, the system has the ability to cope with unexpected events.

Human-robot teams are used in Urban Search and Rescue (USAR). Robots are teleoperated and used mainly as tools to search for survivors. Studies completed on human-robot interaction for USAR reveal that the lack of situational awareness has a negative effect on performance (Murphy 2004; Yanco, Drury et al. 2004). The use of an overhead camera and automatic mapping techniques improved situational awareness and reduced the number of navigational errors (Scholtz 2002; Scholtz, Antonishek et al. 2005). USAR is conducted in uncontrolled, hazardous environments with adverse ambient conditions that affect the quality of sensor and video data. Studies show that varying the level of robot autonomy and combining data from multiple sensors increases the success rate of identifying survivors (Nourbakhsh, Sycara et al. 2005).

Ohba *et al.* (Ohba, Kawabata et al. 1999) developed a system where multiple operators in different locations control the collision free coordination of several robots in a common work environment. Due to teleoperation time delay and the operators being unaware of each other's intentions, a predictive graphics display was used to avoid collisions. The predictive simulator enlarged the thickness of the robotic arm being controlled by other operators as a buffer to prevent collisions caused by time delay and the remote operators not being aware of each other's intentions. In further work, operator's commands were sent simultaneously to the robot and the graphics predictor to circumvent the time delay (Chong, Kotoku et al. 2001). The predictive simulator used these commands to provide virtual force feedback to the operators and avoid collisions that might otherwise have occurred had the time delay not been addressed. The predictive graphics display is an important means of communicating intentions and increasing situational awareness, thus reducing the number of collisions and damage to the system.

3.2 Guide, Host and Assistant Robots

Nourbakhsh *et al.* (Nourbakhsh, Bobenage et al. 1999) created and installed Sage, an autonomous mobile robot in the Dinosaur Hall at the Carnegie Museum of Natural History. Sage, shown in Fig. 1, interacts with museum visitors through an LCD screen and audio, and uses humor to creatively engage visitors. Sage also exhibits emotions and changes in mood to enhance communication. Sage is completely autonomous and when confronted with trouble will stop and ask for help. Sage shows not only how speech affects communication, but also how the form of speech and non-verbal communication influences how well communication takes place.

Figure 1. Musuem guide robot Sage (Nourbakhsh, Bobenage et al. 1999)

The autonomous interactive robot Robovie, shown in Fig 2, is a humanoid robot that communicates and interacts with humans as a partner and guide (Kanda, Ishiguro et al. 2002). Its use of gestures, speech and eye contact enables the robot to effectively communicate with humans. Results of experiments showed that robot communication behavior induced human communication responses that increased understanding. During interaction with Robovie participants spent more than half of the time focusing on the face of the robot, indicating the importance of gaze in human-robot communication.

Figure 2. Robovie interacting with school children (Kanda, Ishiguro et al. 2002.)

Robots used as guides in museums must interact with people and portray human-like behavior to be accepted. Kuzuoka *et al.* (Kuzuoka, Yamazaki et al. 2004) conducted studies in a science museum to see how humans project when they communicate. The term projection is the capacity to predict or anticipate the unfolding of events. The ability to project was found to be difficult through speech alone because speech does not allow a partner to anticipate what the next action may be in the way that body language (gesture) or gaze can.

Kuzuoka *et al.* (Kuzuoka, Yamazaki et al. 2004) designed a remote instruction robot, Gestureman, to investigate projection. A remote operator controlled Gestureman from a separate room. The operator, through Gestureman's three cameras, had a wider view of the local work space than a person normally would and so could see objects without the robot facing them, as shown in Fig. 3. This dual ecology led to the local human participants being misled as to what the robot was focusing on, and thus not being able to quickly locate what the remote user was trying to identify. The experiment highlighted the importance of gaze direction and situational awareness in effective collaboration.

Figure 3. The Gestureman experiment: Remote operator (left) with wider field of view than robot, identifies object but does not project this intention to local participant (right) (Kuzuoka, Yamazaki et al. 2004)

An assistant robot should exhibit a high degree of autonomy to obtain information about their human partner and surroundings. Iossifidis *et al.* (Iossifidis, Theis et al. 2003) developed CoRa (Cooperative Robot Assistant) that is modeled on the behaviors, senses, and anatomy of humans. CoRa is fixed to a table and interacts through speech, hand gestures, gaze and mechanical interaction, allowing it to obtain information about its surrounding and partner. CoRa's tasks include visual identification of objects presented by its human teacher, recognition of objects, grasping and handing over of objects and performing simple assembly tasks.

Cero (Huttenrauch, Green et al. 2004) is an assistant robot designed to help those with physical disabilities in an office environment. During the iterative development of Cero user studies showed that communicating through speech alone was not effective enough. Users commented that they could not distinguish where the front of the robot was nor could they determine if their commands to the robot were understood correctly. In essence, communication was not being effectively grounded. To overcome this difficulty, a humanoid figure was mounted on the front of the robot that could move its head and arms, see Fig. 4. With the humanoid figure users felt more comfortable communicating with the robot and grounding was easier to achieve (Huttenrauch, Green et al. 2004). These results highlight the importance of grounding in communication and also the impact that human-like gestures can have on the grounding process.

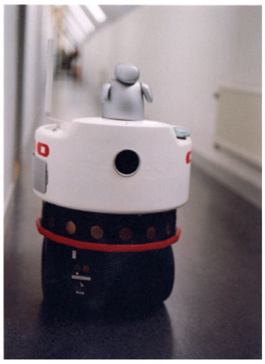

Figure 4. Cero robot with humanoid figure to enable grounding in communication (Huttenrauch, Green et al. 2004)

Sidner and Lee (Sidner and Lee 2005) show that a hosting robot must not only exhibit conversational gestures, but also must interpret these behaviors from their human partner to engage in collaborative communication. Their robot Mel, a penguin hosting robot shown in Fig. 5, uses vision and speech recognition to engage a human partner in a simple demonstration. Mel points to objects, tracks the gaze direction of the participant to ensure instructions are being followed and looks at observers to acknowledge their presence. Mel actively participates in the conversation and disengages from the conversation when appropriate. Mel is a good example of combining the channels from the communication model to effectively ground a conversation, more explicitly, the use of gesture, gaze direction and speech are used to ensure two-way communication is taking place.

Figure 5. Mel giving a demonstration to a human participant (Sidner and Lee 2005)

3.3 Humanoid Robots

Robonaut is a humanoid robot designed by NASA to be an assistant to astronauts during an extra vehicular activity (EVA) mission. It is anthropomorphic in form allowing an intuitive one to one mapping for remote teleoperation. Interaction with Robonaut occurs in the three roles outlined in the work on human-robot interaction by Scholtz (Scholtz 2003): 1) remote human operator, 2) a monitor and 3) a coworker. Robonaut is shown in Fig. 6. The coworker interacts with Robonaut in a direct physical manner and is much like interacting with a human.

Figure 6. Robonaut working with a human (left) and human teleoperating Robonaut (right) (Glassmire, O'Malley et al. 2004)

Experiments have shown that force feedback to the remote human operator results in lower peak forces being used by Robonaut (Glassmire, O'Malley et al. 2004). Force feedback in a teleoperator system improves performance of the operator in terms of reduced completion times, decreased peak forces and torque, as well as decreased cumulative forces. Thus, force feedback serves as a tactile form of non-verbal human-robot communication.

Research into humanoid robots has also concentrated on making robots appear human in their behavior and communication abilities. For example, Breazeal *et al.* (Breazeal, Edsinger et al. 2001) are working with Kismet, a robot that has been endowed with visual perception that is human-like in its physical implementation. Kismet is shown in Fig. 7. Eye movement and gaze direction play an important role in communication aiding the participants in reaching common ground. By following the example of human vision movement and meaning, Kismets' behavior will be understood and Kismet will be more easily accepted socially. Kismet is an example of a robot that can show the non-verbal cues typically present in human-human conversation.

Figure 7. Kismet showing facial expressions present in human communication (Breazeal, Edsinger et al. 2001)

Robots with human social abilities, rich social interaction and natural communication will be able to learn from human counterparts through cooperation and tutelage. Breazeal *et al.* (Breazeal, Brooks et al. 2003; Breazeal 2004) are working towards building socially intelligent cooperative humanoid robots that can work and learn in partnership with people. Robots will need to understand intentions, beliefs, desires and goals of humans to provide relevant assistance and collaboration. To collaborate, robots will also need to be able to infer and reason. The goal is to have robots learn as quickly and easily, as well as in the same manner, as a person. Their robot, Leonardo, is a humanoid designed to express and gesture to people, as well as learn to physically manipulate objects from natural human instruction, as shown in Fig. 8. The approach for Leonardo's learning is to communicate both verbally and non-verbally, use visual deictic references, and express sharing and understanding of ideas with its teacher. This approach is an example of employing the three communication channels in the model used in this chapter for effective communication.

Figure 8. Leonardo activating the middle button upon request (left) and learning the name of the left button (right) (Breazeal, Brooks et al. 2003.)

3.4 Robots in Collaborative Tasks

Inagaki *et al.* (Inagaki, Sugie et al. 1995) proposed that humans and robots can have a common goal and work cooperatively through perception, recognition and intention inference. One partner would be able to infer the intentions of the other from language and behavior during collaborative work. Morita *et al.* (Morita, Shibuya et al. 1998) demonstrated that the communication ability of a robot improves with physical and informational interaction synchronized with dialog. Their robot, Hadaly-2, expresses efficient physical and informational interaction, thus utilizing the environmental channel for collaboration, and is capable of carrying an object to a target position by reacting to visual and audio instruction.

Natural human-robot collaboration requires the robotic system to understand spatial references. Tversky *et al.* (Tversky, Lee et al. 1999) observed that in human-human communication, speakers used the listeners perspective when the listener had a higher cognitive load than the speaker. Tenbrink *et al.* (Tenbrink, Fischer et al. 2002) presented a method to analyze spatial human-robot interaction, in which natural language instructions were given to a robot via keyboard entry. Results showed that the humans used the robot's perspective for spatial referencing.

To allow a robot to understand different reference systems, Roy *et al.* (Roy, Hsiao et al. 2004) created a system where their robot is capable of interpreting the environment from its perspective or from the perspective of its conversation partner. Using verbal communication, their robot Ripley was able to understand the difference between spatial references such as my left and your left. The results of Tenbrink *et al.* (Tenbrink, Fischer et al. 2002), Tversky *et al.* (Tversky, Lee et al. 1999) and Roy *et al.* (Roy, Hsiao et al. 2004) illustrate the importance of situational awareness and a common frame of reference in spatial communication.

Skubic *et al.* (Skubic, Perzanowski et al. 2002; Skubic, Perzanowski et al. 2004) also conducted a study on human-robotic spatial dialog. A multimodal interface was used, with input from speech, gestures, sensors and personal electronic devices. The robot was able to use dynamic levels of autonomy to reassess its spatial situation in the environment through the use of sensor readings and an evidence grid map. The result was natural human-robot spatial dialog enabling the robot to communicate obstacle locations relative to itself and receive verbal commands to move to or near an object it had detected.

Rani *et al.* (Rani, Sarkar et al. 2004) built a robot that senses the anxiety level of a human and responds appropriately. In dangerous situations, where the robot and human are working in collaboration, the robot will be able to detect the anxiety level of the human and take appropriate actions. To minimize bias or error the emotional state of the human is interpreted by the robot through physiological responses that are generally involuntary and are not dependent upon culture, gender or age.

To obtain natural human-robot collaboration, Horiguchi *et al.* (Horiguchi, Sawaragi et al. 2000) developed a teleoperation system where a human operator and an autonomous robot share their intent through a force feedback system. The human or robot can control the system while maintaining their independence by relaying their intent through the force feedback system. The use of force feedback resulted in reduced execution time and fewer stalls of a teleoperated mobile robot.

Fernandez *et al.* (Fernandez, Balaguer et al. 2001) also introduced an intention recognition system where a robot participating in the transportation of a rigid object detects a force signal measured in the arm gripper. The robot uses this force information, as non-verbal communication, to generate its motion planning to collaborate in the execution of the transportation task. Force feedback used for intention recognition is another way in which humans and robots can communicate non-verbally and work together.

Collaborative control was developed by Fong *et al.* (Fong, Thorpe et al. 2002a; Fong, Thorpe et al. 2002b; Fong, Thorpe et al. 2003) for mobile autonomous robots. The robots work autonomously until they run into a problem they can't solve. At this point, the robots ask the remote operator for assistance, allowing human-robot interaction and autonomy to vary as needed. Performance deteriorates as the number of robots working in collaboration with a single operator increases (Fong, Thorpe et al. 2003). Conversely, robot performance increases with the addition of human skills, perception and cognition, and benefits from human advice and expertise.

In the collaborative control structure used by Fong *et al.* (Fong, Thorpe et al. 2002a; Fong, Thorpe et al. 2002b; Fong, Thorpe et al. 2003) the human and robots engage in dialog, exchange information, ask questions and resolve differences. Thus, the robot has more freedom in execution and is more likely to find good solutions when it encounters problems. More succinctly, the human is a partner whom the robot can ask questions, obtain assistance from and in essence, collaborate with.

In more recent work, Fong *et al* (Fong, Kunz et al. 2006) note that for humans and robots to work together as peers, the system must provide mechanisms for these peers to communicate effectively. The Human-Robot Interaction Operating System (HRI/OS) introduced enables a team of humans and robots to work together on tasks that are well defined and narrow in scope. The agents are able to use dialog to communicate and the autonomous agents are able to use spatial reasoning to interpret 'left of' type dialog elements. The ambiguities arising from such dialog are resolved through the use of modeling the situation in a simulator.

3.5 Summary

From the research presented, a few points of importance to human-robot collaboration can be identified. Varying the level of autonomy of human-robotic systems allows the strengths of both the robot and the human to be maximized. It also allows the system to optimize the problem solving skills of a human and effectively balance that with the speed and physical dexterity of a robotic system. A robot should be able to learn tasks from its human counterpart and later complete these tasks autonomously with human intervention only when requested by the robot. Adjustable autonomy enables the robotic system to better cope with unexpected events, being able to ask its human team member for help when necessary.

For robots to be effective partners they should interact meaningfully through mutual understanding. Situational awareness and common frames of reference are vital to effective communication and collaboration. Communication cues should be used to help identify the focus of attention, greatly improving performance in collaborative work. Grounding, an essential ingredient of the collaboration model, can be achieved through meaningful interaction and the exchange of dialog.

A robot will be better understood and accepted if its communication behaviour emulates that of humans. The use of humour and emotion can increase the effectiveness of a robot to communicate, just as in humans. A robot should reach a common understanding in communication by employing the same conversational gestures used by humans, such as gaze direction, pointing, hand and face gestures. During human-human conversation, actions are interpreted to help identify and resolve misunderstandings. Robots should also interpret behaviour so their communication comes across as more natural to their human conversation partner. Communication cues, such as the use of humour, emotion, and non-verbal cues, are essential to communication and thus, effective collaboration.

4. Augmented Reality for Human-Robot Collaboration

Augmented Reality (AR) is a technology that facilitates the overlay of computer graphics onto the real world. AR differs from virtual reality (VR) in that it uses graphics to enhance the physical world rather than replacing it entirely, as in a virtual environment. AR enhances rather replaces reality. Azuma *et al.* (Azuma, Baillot et al. 2001) note that AR computer interfaces have three key characteristics:

- They combine real and virtual objects.

- The virtual objects appear registered on the real world.
- The virtual objects can be interacted with in real time.

AR also supports transitional user interfaces along the entire spectrum of Milgram's Reality-Virtuality continuum (Milgram and Kishino 1994), see Fig. 9.

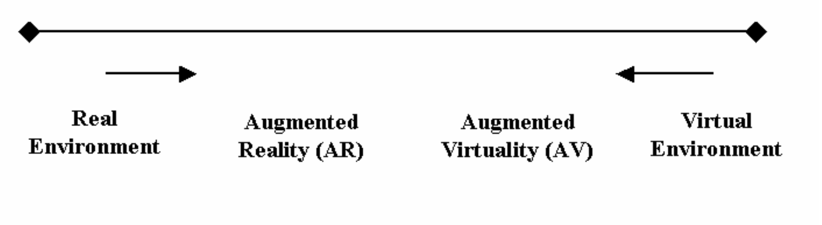

Figure 9. Milgram's reality-virtuality continuum (Milgram and Kishino 1994)

AR provides a 3D world that both the human and robotic system can operate within. This use of a common 3D world enables both the human and robotic system to utilize the same common reference frames. The use of AR will support the use of spatial dialog and deictic gestures, allows for adjustable autonomy by supporting multiple human users, and will allow the robot to visually communicate to its human collaborators its internal state through graphic overlays on the real world view of the human. The use of AR enables a user to experience a tangible user interface, where physical objects are manipulated to affect changes in the shared 3D scene (Billinghurst, Grasset et al. 2005), thus allowing a human to reach into the 3D world of the robotic system and manipulate it in a way the robotic system can understand.

This section first provides examples of AR in human-human collaborative environments, and then discusses the advantages of an AR system for human-robot collaboration. Mobile AR applications are then presented and examples of using AR in human-robot collaboration are discussed. The section concludes by relating the features of collaborative AR interfaces to the communication model for human-robot collaboration presented in section two.

4.1 AR in Collaborative Applications

AR technology can be used to enhance face-to-face collaboration. For example, the Shared Space application effectively combined AR with physical and spatial user interfaces in a face-to-face collaborative environment (Billinghurst, Poupyrev et al. 2000). In this interface users wore a head mounted display (HMD) with a camera mounted on it. The output from the camera was fed into a computer and then back into the HMD so the user saw the real world through the video image, as depicted in Fig. 10.

This set-up is commonly called a video-see-through AR interface. A number of marked cards were placed in the real world with square fiducial patterns on them and a unique symbol in the middle of the pattern. Computer vision techniques were used to identify the unique symbol, calculate the camera position and orientation, and display 3D virtual images aligned with the position of the markers (ARToolKit 2007). Manipulation of the physical markers was used for interaction with the virtual content. The Shared Space application provided the users with rich spatial cues allowing them to interact freely in space with AR content.

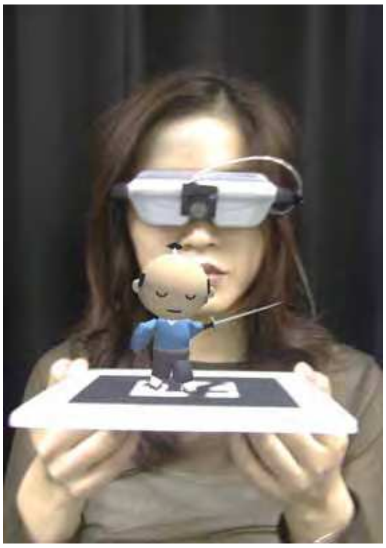

Figure 10. AR with head mounted display and 3D graphic placed on fiducial marker (Billinghurst, Poupyrev et al. 2000)

Through the ability of the ARToolkit software (ARToolKit 2007) to robustly track the physical markers, users were able to interact and exchange markers, thus effectively collaborating in a 3D AR environment. When two corresponding markers were brought together, it would result in an animation being played. For example, when a marker with an AR depiction of a witch was put together with a marker with a broom, the witch would jump on the broom and fly around.

User studies have found that people have no difficulties using the system to play together, displaying collaborative behavior seen in typical face-to-face interactions (Billinghurst, Poupyrev et al. 2000). The Shared Space application supports natural face-to-face communication by allowing multiple users to see each other's facial expressions, gestures and body language, demonstrating that a 3D collaborative environment enhanced with AR content can seamlessly enhance face-to-face communication and allow users to naturally work together.

Another example of the ability of AR to enhance collaboration is the MagicBook, shown in Fig. 11, which allows for a continuous seamless transition from the physical world to augmented and/or virtual reality (Billinghurst, Kato et al. 2001). The MagicBook utilizes a real book that can be read normally, or one can use a hand held display (HHD) to view AR content popping out of the real book pages. The placement of the augmented scene is achieved by the ARToolkit (ARToolKit 2007) computer vision library. When the user is interested in a particular AR scene they can fly into the scene and experience it as an immersive virtual environment by simply flicking a switch on the handheld display. Once immersed in the virtual scene, when a user turns their body in the real world, the virtual viewpoint changes accordingly. The user can also fly around in the virtual scene by pushing a pressure pad in the direction they wish to fly. When the user switches to the immersed virtual world an inertial tracker is used to place the virtual objects in the correct location.

Figure 11. MagicBook with normal view (left), exo-centric view AR (middle), and immersed ego-centric view (right) (Billinghurst, Kato et al. 2001)

The MagicBook also supports multiple simultaneous users who each see the virtual content from their own viewpoint. When the users are immersed in the virtual environment they can experience the scene from either an ego-centric or exo-centric point of view (Billinghurst, Kato et al. 2001). The MagicBook provides an effective environment for collaboration by allowing users to see each other when viewing the AR application, maintaining important visual cues needed for effective collaboration. When immersed in the VR environment, users are represented as virtual avatars and can be seen by other users in the AR or VR scene, thereby maintaining awareness of all users, and thus still providing an environment supportive of effective collaboration.

Prince *et al.* (Prince, Cheok et al. 2002) introduced a 3D live augmented reality conferencing system. Through the use of multiple cameras and an algorithm determining shape from silhouette, they were able to superimpose a live 3D image of a remote collaborator onto a fiducial marker, creating the sense that the live remote collaborator was in the workspace of the local user. Fig. 12 shows the live collaborator displayed on a fiducial marker. The shape from silhouette algorithm works by each of 15 cameras identifying a pixel as belonging to the foreground or background, isolation of the foreground information produces a 3D image that can be viewed from any angle by the local user.

Communication behaviors affect performance in collaborative work. Kiyokawa *et al.* (Kiyokawa, Billinghurst et al. 2002) experimented with how diminished visual cues of co-located users in an AR collaborative task influenced task performance. Performance was best when collaborative partners were able to see each other in real time. The worst case occurred in an immersive virtual reality environment where the participants could only see virtual images of their partners.

In a second experiment Kiyokawa *et al.* (Kiyokawa, Billinghurst et al. 2002) modified the location of the task space, as shown in Fig. 13. Participants expressed more natural communication when the task space was between them; however, the orientation of the task space was significant. The task space between the participants meant that one person had a reversed view from the other. Results showed that participants preferred the task space to be on a wall to one side of them, where they could both view the workspace from the same perspective. This research highlights the importance of the task space location, the need for a common reference frame and the ability to see the visual cues displayed by a collaborative partner.

Figure 12. Remote collaborator as seen on AR fiducial marker (Prince, Cheok et al. 2002)

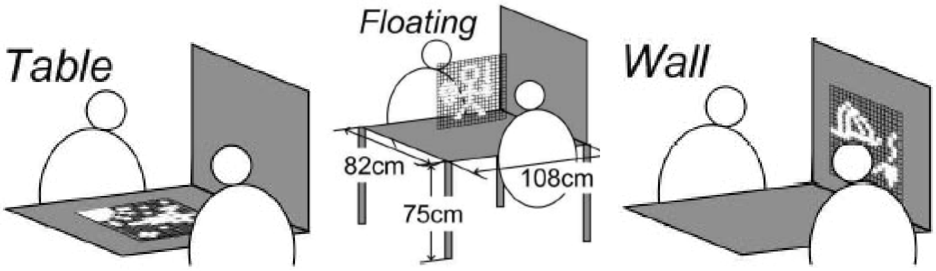

Figure 13. Different locations of task space in Kiyokawa *et al* second experiment (Kiyokawa, Billinghurst et al. 2002

These results show that AR can enhance face-to-face collaboration in several ways. First, collaboration is enhanced through AR by allowing the use of physical tangible objects for ubiquitous computer interaction. Thus making the collaborative environment natural and effective by allowing participants to use objects for interaction that they would normally use in a collaborative effort. AR provides rich spatial cues permitting users to interact freely in space, supporting the use of natural spatial dialog. Collaboration is also enhanced by the use of AR since facial expressions, gestures and body language are effectively transmitted.

In an AR environment multiple users can view the same virtual content from their own perspective, either from an ego- or exo-centric viewpoint. AR also allows users to see each other while viewing the virtual content enhancing spatial awareness and the workspace in an AR environment can be positioned to enhance collaboration. For human-robot collaboration, AR will increase situational awareness by transmitting necessary spatial cues through the three channels of the communication model presented in this chapter.

4.2 Mobile AR

For true human-robot collaboration it is optimal for the human to not be constrained to a desktop environment. A human collaborator should be able to move around in the environment the robotic system is operating in. Thus, mobilily is an important ingredient for human-robot collaboration. For example, if an astronaut is going to collaborate with an autonomous robot on a planet surface, a mobile AR system could be used that operates inside the astronauts suit and projects virtual imagery on the suit visor. This approach would allow the astronaut to roam freely on the planet surface, while still maintaining close collaboration with the autonomous robot.

Wearable computers provide a good platform for mobile AR. Studies from Billinghurst *et al.* (Billinghurst, Weghorst et al. 1997) showed that test subjects preferred working in an environment where they could see each other and the real world. When participants used wearable computers they performed best and communicated almost as if communicating in a face-to-face setting (Billinghurst, Weghorst et al. 1997). Wearable computing provides a seamless transition between the real and virtual worlds in a mobile environment.

Cheok *et al.* (Cheok, Weihua et al. 2002) utilized shape from silhouette live 3D imagery (Prince, Cheok et al. 2002) and wearable computers to create an interactive theatre experience, as depicted in Fig. 14. Participants collaborate in both an indoor and outdoor setting. Users seamlessly transition between the real world, augmented and virtual reality, allowing multiple users to collaborate and experience the theatre interactively with each other and 3D images of live actors.

Reitmayr and Schmalstieg (Reitmayr and Schmalstieg 2004) implemented a mobile AR tour guide system that allows multiple tourists to collaborate while they explore a part of the city of Vienna. Their system directs the user to a target location and displays location specific information that can be selected to provide detailed information. When a desired location is selected, the system computes the shortest path, and displays this path to the user as cylinders connected by arrows, as shown in Fig. 15.

The Human Pacman game (Cheok, Fong et al. 2003) is an outdoor mobile AR application that supports collaboration. The system allows for mobile AR users to play together, as well as get help from stationary observers. Human Pacman, see Fig. 16, supports the use of tangible and virtual objects as interfaces for the AR game, as well as allowing real world physical interaction between players. Players are able to seamlessly transition between a first person augmented reality world and an immersive virtual world. The use of AR allows the virtual Pacman world to be superimposed over the real world setting. AR enhances collaboration between players by allowing them to exchange virtual content as they are moving through the AR outdoor world.

Figure 14. Mobile AR setup (left) and interactive theatre experience (right) (Cheok, Weihua et al. 2002)

Figure 15. Mobile AR tour guide system (left) and AR display of path to follow(right) (Reitmayr and Schmalsteig 2004)

Figure 16. AR Human Pacman Game (Cheok, Fong 2003)

To date there has been little work on the use of mobile AR interfaces for human-robot collaboration; however, several lessons can be learnt from other wearable AR systems. The majority of mobile AR applications are used in an outdoor setting, where the augmented objects are developed and their global location recorded before the application is used. Two important issues arise in mobile AR; data management, and the correct registration of the outdoor augmented objects. With respect to data management, it is important to develop a system where enough information is stored on the wearable computer for the immediate needs of the user, but also allows access to new information needed as the user moves around (Julier, Baillot et al. 2002). Data management should also allow for the user to view as much information as required, but at the same time not overload the user with so much information that it hinders performance. Current AR systems typically use GPS tracking for registration of augmented information for general location coordinates, then use inertial trackers, magnetic trackers or optical fiducial markers for more precise AR tracking. Another important item to design into a mobile AR system is the ability to continue operation in case communication with the remote server or tracking system is temporarily lost.