Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design

Scott A. Green[a,b], Mark Billinghurst[b], XiaoQi Chen[a] & J. Geoffrey Chase[a]
[a]Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand
[b]Human Interface Technology Laboratory, New Zealand (HITLab NZ), Christchurch, New Zealand
scott.green@canterbury.ac.nz

Abstract: NASA's vision for space exploration stresses the cultivation of human-robotic systems. Similar systems are also envisaged for a variety of hazardous earthbound applications such as urban search and rescue. 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. However, little attention has been paid to joint human-robot teams. Making human-robot collaboration natural and efficient is crucial. In particular, grounding, situational awareness, a common frame of reference and spatial referencing are vital in effective communication and collaboration. Augmented Reality (AR), the overlaying of computer graphics onto the real worldview, can provide the necessary means for a human-robotic system to fulfill these requirements for effective collaboration. This article reviews the field of human-robot interaction and augmented reality, investigates the potential avenues for creating natural human-robot collaboration through spatial dialogue utilizing AR and proposes a holistic architectural design for human-robot collaboration.

Keywords: augmented reality, collaboration, communication, human-computer interaction, human-robot collaboration, human-robot interaction, robotics.

1. Introduction

NASA's vision for space exploration stresses the cultivation of human-robotic systems (NASA 2004). Fong and Nourbakhsh (Fong and Nourbakhsh 2005) point out that to reduce human workload, costs, fatigue driven error and risk, intelligent robotic systems will have to be part of mission design. They also observe that scant attention has been paid to joint human-robot teams, and making human-robot collaboration natural and efficient is crucial to future space exploration. Companies such as Honda (Honda 2007), Toyota (Toyota 2007) and Sony (Sony 2007) are also interested in developing consumer robots that interact with humans in the home and workplace. There is growing interest in the field of human-robot interaction (HRI) as can be determined by the inaugural conference for HRI (HRI2006 2006). The Cogniron project (COGNIRON 2007), MIT Media lab (Hoffmann and Breazeal 2004) and the Mitsubishi Electric Research Laboratories (Sidner and Lee 2005) recognize the need for human-robot collaboration as well, and are currently conducting research in this emerging area.
Clearly, there is a growing need for research on human-robot collaboration and models of communication between human and robotic systems. This article reviews the field of human-robot interaction with a focus on communication and collaboration. It also identifies promising areas for future research focusing on how Augmented Reality technology can support natural spatial dialogue and thus enhance human-robot collaboration.

First an overview of models of human-human collaboration and how they could be used to develop a model for human-robot collaboration is presented. Next, the current state of human-robot interaction is reviewed and how it fits into a model of human-robot collaboration is explored. Augmented Reality (AR) is then reviewed and how it could be used to enhance human-robot collaboration is discussed. Finally, a holistic architectural design for human-robot collaboration using AR 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, models of human-human communication are briefly reviewed. This review will provide a basis for the 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. It is clear that 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 provide 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 in face-to-face collaboration by providing visual feedback, regulating the flow of conversation, communicating emotions and relationships, and improving concentration by restriction of visual input (Kendon 1967), (Argyle 1967). In addition to gaze, humans use a wide range of non-verbal cues to assist in communication, 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 example pointing at something (Kendon 1983). In studying the behavior of human demonstration activities 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) show 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 and result 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 with those surroundings.

Clark and Brennan (Clark and Brennan 1991) provide a communication model to interpret collaboration. In their view, 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 feel confident that common ground is easily reached.

2.2. Human-Human Collaboration Model

This research employs a human-human collaboration model 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. Environment 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 the 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 and environmental cues. Thus, communication is reduced to one content heavy visual channel: text input. Predictably, this approach will have a number of effects on communication: less verbose communication, use of longer phrases, increased time to grounding, slower communication and few interruptions. Taking each of the three communication channels from this model in turn, characteristics of an effective human-robot collaboration system can be identified. The robot 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 can 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.

3. Human-Robot Interaction

The next several sections review current robot research and how the latest generation of robots supports these characteristics. Research into human-robot interaction,

the use of robots as tools, robots as guides and assistants, as well as the progress being made in the development of humanoid robots, are all examined. Finally, a variety of efforts to use robots in collaboration are examined and analyzed in the context of the human-human model presented.

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 benefit from human-robot interaction. For example, to increase the success rate of harvesting, a human-robot collaborative system was implemented for testing by (Bechar and Edan 2003). Results indicated that a human operator working with a robotic system with varying levels of autonomy resulted in improved harvesting of melons. Depending on the complexity of the harvesting environment, varying the level of autonomy of the robotic harvester increased positive detection rates in the amount of 4.5% – 7% from the human operator alone and 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 obtainable due to the dynamic operating conditions. Radiation surveys are completed initially 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 improve situational awareness and reduce 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, thus using the best sensors for the given situation, 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 multiple 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 utilized 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 to 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.

This section on Robots as Tools highlighted two important ingredients for an effective human-robot collaboration system. First, adjustable autonomy, enabling the system to vary the level of robotic system autonomy, increases productivity and is an essential component of an effective collaboration system. Second, situational awareness, or knowing what is happening in the robot's workspace, is also essential in a collaboration system. The human member of the team must know what is happening in the robot's work world to avoid collisions or damage to the robotics system.

3.2. Guide, Hosting 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 was designed with safety, reliability and social capabilities to enable it to be an effective member of the museum staff. Sage shows not only how speech capabilities affect communication, but also, that the form of speech and non-verbal communication influences how well communication takes place.

Fig. 1. Sage interacting with museum visitors through an LCD screen (Nourbakhsh, Bobenage et al. 1999).

The autonomous interactive robot Robovie 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.

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 was used as 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 a person can predict what may happen next by body language (gesture) or focus point of gaze.

Kuzuoka *et al.* (Kuzuoka, Yamazaki et al. 2004) designed a remote instruction robot, Gestureman, to investigate projectability properties. A remote operator, who was located in a separate room from a local user, controlled Gestureman. Through Gestureman's three cameras the remote operator 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. 2. This dual ecology led to 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 remote collaboration and communication.

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 on a table and interacts through speech, hand gestures, gaze and mechanical interaction allowing it to obtain the necessary information about its surrounding and partner. CoRa's tasks include visual identification of objects presented by its human teacher, recognition of an object amongst many, 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, as shown in Fig. 3. After implementation of the humanoid figure, it was found that users felt more comfortable communicating with the robot and grounding was easier to achieve (Huttenrauch, Green et al. 2004). The results from the research on Cero highlight the importance of grounding in communication and the impact that gestures can have on grounding.
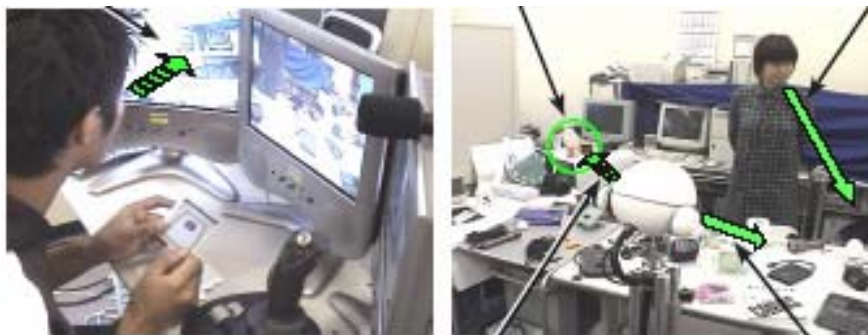

Fig 2. Gestureman: Remote user (left) with wider fov than robot, identifies object but does not project this intention to local participant (right) (Kuzuoka, Yamazaki et al. 2004).

Fig. 3. Cero robot with humanoid figure using gestures to enhance grounding (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. 4, uses vision and speech recognition to engage a human partner in a simple demonstration. Mel points to objects in the demo, tracks the gaze direction of the participant to ensure instructions are being followed, and looks at observers of the demonstration to acknowledge their presence. Mel actively participates in the conversation during the demonstration 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, gesture, gaze direction and speech are used to ensure two-way communication is taking place.

Fig. 4. Mel uses multimodal communication to interact with participants (Sidner and Lee 2005).

Lessons learned from this section for the design of an effective human-robot collaboration system include the need for effective natural speech. A multi-modal approach is necessary as communication is more than just speech alone. The communication behaviour of a robotic system is important as it should induce natural communication with human team members. And, lastly, grounding is a key element in communication, and thus collaboration.

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. Its anthropomorphic form allows it 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. 5. The co-worker interacts with Robonaut in a direct physical manner and is much like interacting with a human.

Fig. 5. Robonaut with coworker and remote human operator (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. 6. 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.
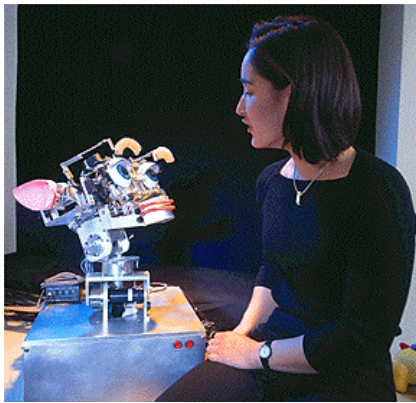
Fig. 6. Kismet displaying non-verbal communication cues (Breazeal, Edsinger et al. 2001).

Fig. 7. Leonardo activating middle button (left) and learning the name of the left button (right) (Breazeal, Brooks et al. 2003).

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. 7. 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 paper for effective communication with a stationary robot.

3.4. Summary

A few points of importance to human-robot collaboration should be noted. Varying the level of autonomy of human-robotic systems allows the strengths of both the robot and the human to be maximized. It 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.

Timing delays are an inherent part of a teleoperated system. It is important to design into the control system an effective means of coping with time delay. Force feedback in a remote controlled robot results in greater control, a more intuitive feel for the remote operator, less stress on the robotic system and better overall performance through tactile non-verbal feedback communication.

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. Research has shown that communication cues, such as the use of humour, emotion, and non-verbal cues, are essential to communication and effective collaboration.

4. Robots in Collaborative Tasks

Inagaki *et al.* (Inagaki, Sugie et al. 1995) propose 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 dialogue. 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 referencing. 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 dialogue. A multimodal interface was used, including 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 dialogue 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 the 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 benefit 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 dialogue, 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 the humans and robots 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 human agents are able to use spatial dialog to communicate and the autonomous agents use spatial reasoning to interpret 'left of' type elements from the spatial dialog. The ambiguities arising from such dialog are resolved through the use of modeling the situation in a simulator.

Research has shown that for robots to be effective partners they should interact meaningfully through mutual understanding. A human-robot collaborative system should take advantage of varying levels of autonomy and multimodal communication allowing the robotic system to work independently and ask its human counterpart for assistance when a problem is encountered. 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 dialogue.

5. 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 in a virtual environment the entire physical world is replaced by computer graphics, 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 is an ideal platform for human-robot collaboration because it provides the following important qualities:

- The ability to enhance reality.
- Seamless interaction between real and virtual environments.
- The ability to share remote views (ego-centric view).
- The ability to visualize the robot relative to the task space (exo-centric view).
- Spatial cues for local and remote collaboration.
- Support for transitional interfaces, moving smoothly from reality into virtuality.
- Support for a tangible interface metaphor.
- Tools for enhanced collaboration, especially for multiple people collaborating with a robot.

These attributes allow AR to support natural spatial dialogue by displaying the visual cues necessary for a human and robot to reach common ground and maintain situational awareness. The use of AR will support the use of spatial dialogue 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 worldview 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).

This section first provides examples of AR in human-human collaborative environments, and then the advantages of an AR system for human-robot collaboration are discussed. Mobile AR applications are then presented and an example of human-robot interaction using AR is discussed. The section concludes by relating the features of collaborative AR interfaces to the communication model for human-robot collaboration presented in section 2.

5.1. AR in Collaborative Applications

AR technology can be used to enhance face-to-face collaboration. For example, the Shared Space Project 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. 8. 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.

Fig. 8. Head Mounted Display (HMD) and virtual object registered 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. Attendees at the SIGGRAPH99 Emerging Technologies exhibit tested the Shared Space system by playing a game similar to Concentration. Around 3000 people tried the application and had no difficulties with playing together, displaying collaborative behavior seen in typical face-to-face interactions (Billinghurst, Poupyrev et al. 2000). The Shared Space interface 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. 9, 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 they turn 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.

Fig. 9: Using the MagicBook to move from Reality to Virtuality (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 VR, 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. 10 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.
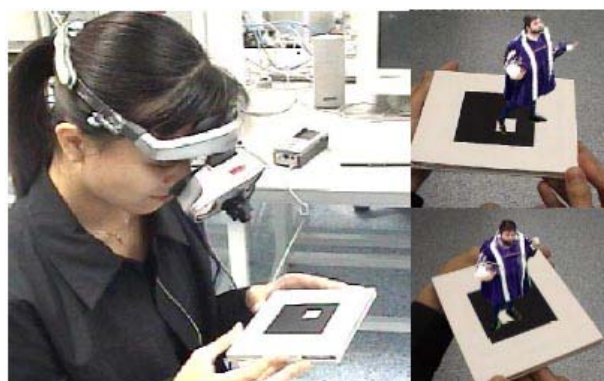

Fig. 10: Live 3D collaborator on fiducial marker (Prince, Cheok et al. 2002).

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. 11. 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 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, as they would both view the workspace from the same perspective. The results of this research point out the importance of the location of task space, the need for a common reference frame and the ability to see the visual cues displayed by a collaborative partner.
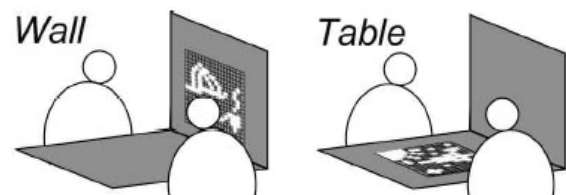

Fig. 11: Different location spaces for Kiyokawa *et al*. (Kiyokawa, Billinghurst et al. 2002) second experiment.

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 dialogue. 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 paper.

5.2. Mobile AR

Mobile AR is a good option for some forms of 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. 12. 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.
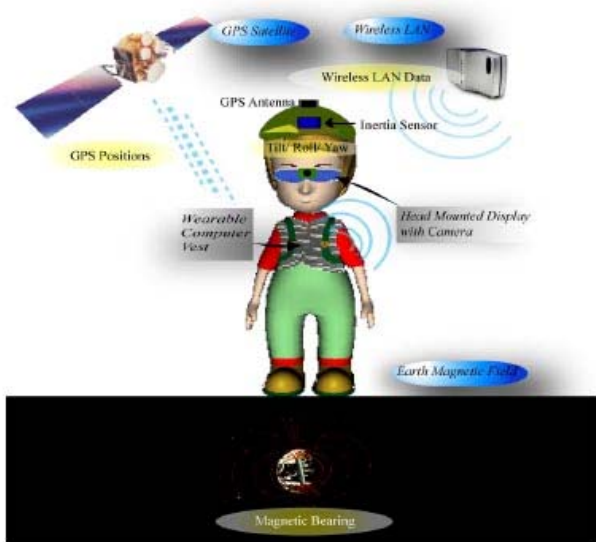

Fig. 12: Mobile AR setup interactive theatre experience (Cheok, Weihua et al. 2002).

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. 13. Multiple users can collaborate in three modes, follow mode, guide mode or meet mode. The meet mode will display the shortest path between the users and thus guide them to a meeting point.
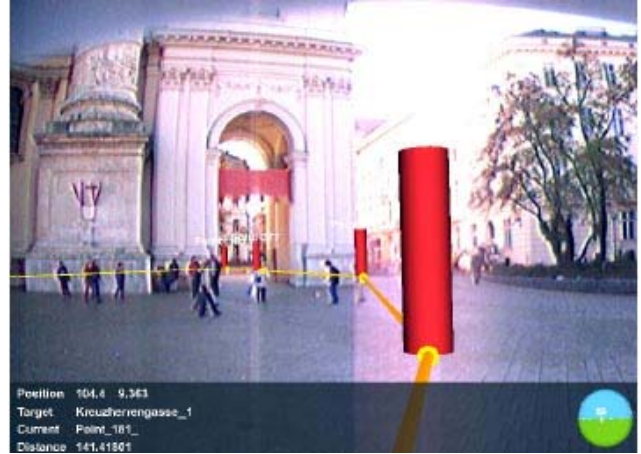

Fig. 13: Reitmayr and Schmalstieg navigation (Reitmayr and Schmalstieg 2004).

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. 14, 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.

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

Fig. 14: Human Pacman (Cheok, Fong et al. 2003).

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.

5.3. First Steps in Using AR in Human-Robot Collaboration
Milgram *et al* (Milgram, Zhai et al. 1993) highlighted the need for combining the attributes humans are good at with those that robots are good at to result in an optimized human-robot team. Humans are good at less accurate referencing, such as using 'here' and 'there', whereas robotic systems need highly accurate discrete information. Milgram *et al* pointed out the need for HRI systems that can transfer the interaction mechanisms that are considered natural for human communication to the precision required for machine information. Their approach was to use augmented overlays in a fixed work environment to enable the human 'director' to use spatial referencing to interactively plan and optimize a robotic manipulator arm.

Giesler *et al.* (Giesler, Steinhaus et al. 2004) are working on a system that allows a robot to interactively create a 3D model of an object on-the-fly. In this application, a laser scanner is used to read in an unknown 3D object. The information from the laser scan is overlaid through AR onto the video feed of the real world, as shown in Fig. 15. The user interactively creates a boundary box around
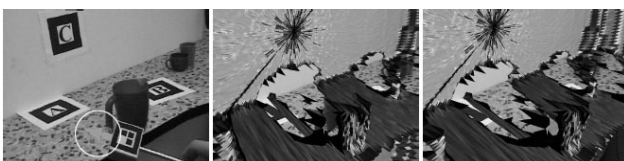
the appropriate portion of the laser scan by using voice commands and an AR magic wand. The wand uses the ARToolkit (ARToolKit 2007) and is made of fiducial markers for tracking. The wand is shown on the far left in Fig. 15. Using a combination of the laser scan and video image, a 3D model of a previously unknown object can be created.

In other work Giesler *et al.* (Giesler, Salb et al. 2004) are implementing an AR system that creates a path for a mobile robot to follow using voice commands and the same magic wand in their work above. Fiducial markers are placed on the floor and used to calibrate the tracking coordinate system. A path is created node by node, by pointing the wand at the floor and giving voice commands for the meaning of a particular node. Map nodes can be interactively moved or deleted. The robot moves from node to node using its autonomous collision detection capabilities. As goal nodes are reached, the node depicted in the AR system changes color to keep the user informed of the robots progress. The robot will retrace steps if an obstruction is encountered and create a new plan to arrive at the goal destination, as shown in Fig. 16.
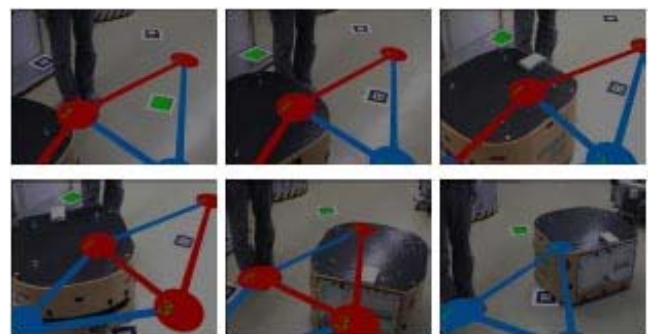

Fig. 15: Magic wand with fiducial tip and a scene with laser scan overlaid (Giesler, Steinhaus et al. 2004).

Fig. 16. Robot follows AR path nodes, redirects when obstacle in way (Giesler, Salb et al. 2004).

Although Giesler *et al* (Giesler, Salb et al. 2004) did not mention a user evaluation, they did comment that the interface was intuitive to use. Results from their work show that AR is an excellent application to visualize planned trajectories and inform the user of the robots progress and intention. It was also mentioned that the ARToolkit (ARToolKit 2007) tracking module can be problematic, sometimes failing due to image noise and changes in lighting.

Bowen *et al* (Bowen, Maida et al. 2004) and Maida *et al* (Maida, Bowen et al. 2006) showed through user studies that the use of AR resulted in significant improvements in robotic control performance. Drury *et al* (Drury, Richer et al. 2006) showed through experiments that augmented real-time video with pre-loaded map terrain data resulted in a statistical difference in comprehension of 3D spatial relationships over using 2D video alone for operators of Unmanned Aerial Vehicles (UAVs). The results were better situational awareness of the activities of the UAV.

5.4. Summary

Augmented Reality is an ideal platform for human-robot collaboration as it provides the ability for a human to share a remote (ego-centric) view with a robot collaborative partner. In terms of the communication model used in this paper, AR will allow the human and robot to ground their mutual understanding and intentions through the visual channel affording a person the ability to see what a robot sees. AR supports the use of deictic gestures, pointing to a place in 3D space and referring to that point as "here", by allowing a 3D overlaid image to be referenced as "here".

AR also allows a human partner to have a worldview (exo-centric) of the collaborative workspace affording spatial understanding of the robots position relative to the surrounding environment. The exo-centric view will allow a human collaborator to know where he/she is in terms of the surrounding environment, as well as, in terms of the robot and other human and robot collaborators. The exo-centric view is vital when considering the field of view of an astronaut in a space suit. The helmet of a space suit does not swivel with neck motion so two astronauts working side by side are unable to see each other (Glassmire, O'Malley et al. 2004). AR can overcome this limitation by increasing the situational awareness of both the human and robot, even if the human is constrained inside a space suit.

Augmented reality supports collaboration between more than two people, thus providing tools for enhanced collaboration, especially for human-robot collaboration where more than one human may wish to collaborate with a robot. AR also supports transitional interfaces along the entire spectrum of Milgram's Reality-Virtuality continuum (Milgram and Kishino 1994), shown in Fig. 17. AR transitions seamlessly from the real world to an immersive data space, as demonstrated by the MagicBook application (Billinghurst, Kato et al. 2001). This seamless transition is yet another important aspect of AR that aids in the grounding process and increases situational awareness. In a study of the performance of human-robot interaction in urban search and rescue, Yanco *et al.* (Yanco, Drury et al. 2004) identified the need for situational awareness of the robot and its surroundings. AR technology can be used to display visual cues that can increase situational awareness and improve the grounding process, enabling the human to more effectively understand what the robot is doing and its internal state (Collett and MacDonald 2006), thus supporting natural spatial dialogue.
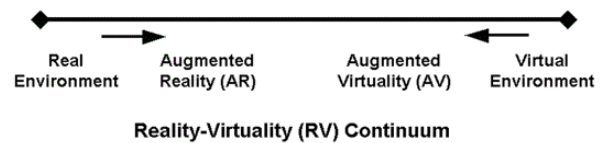
Fig. 17: Milgram's Reality-Virtuality Continuum (Milgram and Kishino 1994).

6. Research Directions in Human-Robot Collaboration

Given this review of the general state of human-robot collaboration, and the presentation and review of using AR to enhance this type of collaboration, the question is: what are promising future research directions? Two important concepts must be kept in mind when designing an effective human-robot collaboration system. One, the robotic system must be able to provide feedback as to its understanding of the situation and its actions (Scholtz 2002). Two, an effective human-robot system must provide mechanisms to enable the human and the robotic system to communicate effectively (Fong, Kunz et al. 2006). In this section, each of the three communication channels in the model presented is explored, and potential avenues to make the model of human-robot collaboration become a reality are discussed.

6.1. The Audio Channel

There are numerous systems readily available for automated speech recognition (ASR) and text to speech (TTS) synthesis. A robust dialogue management system will need to be developed that is capable of taking the appropriate human input from the ASR system and convert this input into appropriate robot commands. The dialogue management system will need to be able to take input from the robot control system, convert this information into suitable text strings for the TTS system to synthesize into understandable audible output for the human collaborators. The dialogue manager will thus need to support the ongoing discussion between the humans and the robotic system. The dialogue manager will need to enable a robot to express its intentions that will include the robot understanding the current situation and responding with alternative approaches to those proposed by the human collaborators or alerting the

human team members when a proposed plan is not feasible and provide reasoning for this determination. This type of clarification (Krujiff, Zender et al. 2006) will require the robotic system to understand the speech, interpret the speech in terms of its surroundings and goal, and express itself through speech. An internal model of the communication process will need to be developed.

The use of humour and emotion will enable the robotic agents to communicate in a more natural and effective manner, and therefore should be incorporated into the dialogue management system. An example of the effectiveness of this type of communication can be seen in Rea, a computer generated human-like real estate agent (Cassell, Bickmore et al. 1999). Rea is capable of multi-modal input and output using verbal and non-verbal communication cues to actively participate in a conversation. Audio can also be spatialized, in essence, placing sound in the virtual world from where it originates in the real world. Spatially locating sound will increase situational awareness and thus provide a means to communicate effectively and naturally.

6.2. The Environmental Channel

To collaborate, a robot will need to understand the use of objects by its human counterpart, such as using an object to point or making a gesture. AR can support this type of interaction by enabling the human to point to a 3D object that both the robot and human refer to, common ground, and use natural dialogue such as "go to *this* point", situational awareness. In a similar manner the robot would be able to express its intentions and beliefs by showing through the 3D overlays what its internal state, plans and understanding of the situation are. Thus using the shared AR environment as an effective spatial communication tool. Referencing a shared 3D environment will support the use of common and shared frames of references, thus affording the ability to effectively communicate in a truly spatial manner. As an example, if a robot did not fully understand a verbal command, it would be able to make use of the shared 3D environment to clearly portray to its collaborators what was not understood, what further information is needed and what the autonomous agent believes could be the correct action to take.

Real physical objects can be used to interact with an AR application. For human-robot communication this translates into a more intuitive user interface, allowing the use of real world objects to communicate with a robot. The use of real world objects is especially important for mobile applications where the user will not be able to use typical computer interface devices, such as a mouse or keyboard.

6.3. The Visual Channel

In natural communication, speech is an important part of grounding a conversation. However, with the limited speech ability of robotic systems, visual cues also provide a means of grounding communication. AR, with its ability to provide ego- and exo-centric views and to seamlessly transition from reality to virtuality, can provide robotic systems with a robust manner in which to ground communication and allow human collaborative partners to understand the intention of the robotic system. AR can also transmit spatial awareness though the ability to provide rich spatial cues, ego- and exo-centric points of view, and also by seamlessly transitioning from the real world to an immersive VR world. An AR system could, therefore, be developed to allow for bi-directional transmission of gaze direction, gestures, facial expressions and body pose. The result would be an increased level of communication and more effective collaboration.

AR is an optimal method of displaying information for the user. Billinghurst *et al.* (Billinghurst, Bowskill et al. 1998) showed through user tests that spatial displays in a wearable computing environment were more intuitive and resulted in significantly increased performance. Fig. 18 shows spatial information displayed in a head stabilised and body stabilised fashion. Using AR to display information, such as robot state, progress and even intent, will result in increased understanding, grounding and, therefore, enhanced collaboration.
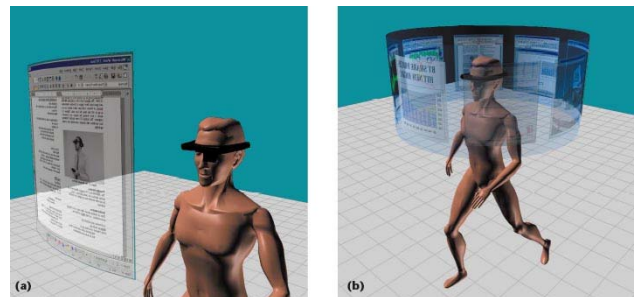

Fig. 18: Head stabilised (a) and body stabilised (b) AR information displays (Billinghurst, Bowskill et al. 1998).

6.4. General Research in AR

In order to develop natural human-robot collaboration, many aspects of AR should be explored, such as communication and data transfer. AR requires transmission of audio and video information, for mobile remote collaboration, an effective means of transmitting this information will be required. An effective AR system requires the means to continue operation in the case of an interruption in communication. Mobile computing should be researched to find an optimal configuration for the components of an AR system.

A data management system providing the right information at the right time will be needed. An AR system would benefit greatly from the ability to create new virtual content on the fly. The AR system should be usable in various spatial ranges. For example, a system should be developed that can be used for local collaboration with the robot, human and robot working

side by side, and at the same time the system should also support remote collaboration, human on earth or in space station and robot on planet surface or outside space station. The system should be able to support a combination of spatial configurations, i.e. local collaboration with the robot and at the same time allow collaboration from remote participants.

Tracking techniques have always been a challenge in AR. To support human-robot collaboration in various environments, robust tracking technologies will need to be researched. The AR system should be able to be used in virtually any environment and not be affected by changes in ambient conditions, such as lighting. Human-robot collaboration will occur in unprepared environments, therefore, research into using AR in unprepared environments is yet another area to be explored. AR shows promise to be an excellent platform for human-robot collaboration, much research still needs to be conducted to develop a viable AR human-robot collaboration system.

7. Architectural Design

Employing the lessons learned from this literature review, an architectural design has been developed for Human-Robot Collaboration (HRC). A multimodal approach is envisioned that combines speech and gesture through the use of AR that will allow humans to use natural speech and gestures to communicate with robotic systems. Through this architecture the robotic system will receive the discrete information it needs to operate while allowing human team members to communicate in a natural and effective manner by referencing objects, positions, and intentions through natural gesture and speech. The human and the robotic system will each maintain situational awareness by referencing the same shared 3D visual of the work world in the AR environment.

The architectural design is show in Fig. 19. The speech-processing module will recognize human speech and parse this speech into the appropriate dialog components. When a defined dialog goal is achieved through speech recognition, the required information will be sent to the Multimodal Communication Processor (MCP). The speech-processing module will also take information from the MCP and the robotic system and synthesize this speech for effective dialog with human team members. The speech processing will take place using the spoken dialog system Ariadne (Ariadne 2006). Ariadne was chosen for its capability for rapid dialog creation (Denecke 2002).

Gesture processing will enable a human to use deictic referencing and normal gestures to communicate effectively with a robotic system. To communicate effectively with a robotic system it is imperative that the system be able to translate the generic references humans use, such as pointing into 3D space and saying "go here",
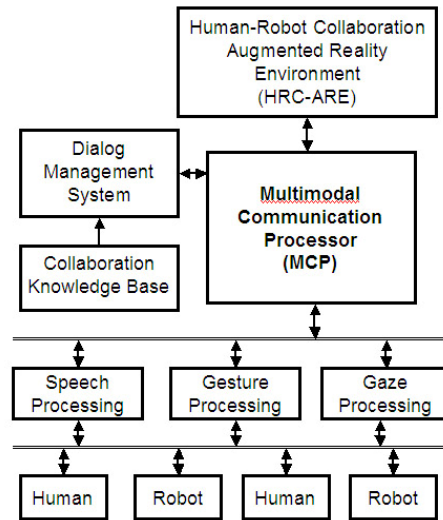

Fig. 19: Human Robot Collaboration System Architecture

into the discrete information a robotic system needs to operate. The gesture-processing module will recognize gestures used by a human and pass this information to the MCP. The MCP will combine the speech from the speech processing module, the gesture information from the gesture-processing module and use the Human-Robot Collaboration Augmented Reality Environment (HRC-ARE) to effectively enable the defining of ambiguous deictic references such as here, there, this and that. The disambiguation of the deictic references will be accomplished in the AR environment, as the AR environment is a 3D virtual replication of the robot's world allowing visual translation and definition of such deictic references. The human will be able to use a tangible paddle to reach into and interact with this 3D virtual world. This tangible interaction is a key feature of AR that makes it an ideal platform for HRC. The ARToolKit (ARToolKit 2007) will be used for the AR environment.

The gaze-processing module will track the users gaze through the use of a head mounted display. This gaze tracking will enable each human team member to view the HRC-ARE from his or her own perspective. This personal viewing of the work world will result in increased situational awareness as each team member will view the work environment form their own perspective and will be able to change their perspective simply by moving around the 3D virtual environment as they would a real world object, or they could move the 3D virtual world around and maintain their position by moving the real world fiducial marker that the 3D world is "attached" to. Not only will human team members be able to maintain their perspective of the robotic system's work environment, but they will also be able to smoothly switch to the robot's view of the work environment. This ability to smoothly switch between an exo-centric (God's

eye) view of the work environment to an ego-centric (robotic system's) view of the work environment is yet another feature of AR that makes it ideal for HRC and enables the human to quickly and effectively reach common ground and maintain situational awareness with the robotic system.

The Dialog Management System (DMS) will be aware of the communication that needs to take place for the human and robot to collaboratively complete a task. The MCP will take information from the speech, gesture and gaze processing modules along with information generated from the HRC-ARE and supply it to the DMS. The DMS will be responsible for combining this information and comparing it to the information stored in the Collaboration Knowledge Base (CKB). The CKB will contain information pertaining to what is needed to complete the desired tasks that the human-robot team wishes to complete. The DMS will then respond through the MCP to either human team members or the robotic system, whichever is appropriate, facilitating dialog and tracking when a command or request is complete.

The MCP will be responsible for receiving information from the other modules in the system and sending information to the appropriate modules. The MCP will thus be responsible for combining multimodal input, registering this input into something the system can understand and then sending the required information to other system modules for action. The result of this system design is that a human will be able to use natural speech and gestures to interact with a robotic system.

8. Conclusion

This paper began by showing a need for human-robot collaboration. Human-human communication was discussed; a model for human-human collaboration created and this model was used as a reference model for human-robot collaboration. The state of human-robot interaction was reviewed and how this interaction fit into the model of human-robot collaboration was explored. Finally, Augmented Reality technology was reviewed and how AR could be used to enhance human-robot collaboration was explored.

The model developed for human communication is based on three components, the communication channels available, the communication cues provided by each of these channels, and the technology that affects the transmission of these cues. There are three channels for communication: visual, audio and environmental. Depending on the transmission medium used, communication cues may not be effectively transmitted. Applying this model to human-robot collaboration, the characteristics of an effective human-robot collaborative system can be analyzed. An effective system should strive to allow communication cues to be transferred in all three channels. Therefore, the robot must be able to

understand and exhibit audio, visual and environmental communication cues.

Effective human-robot collaborative systems should make use of varying levels of autonomy. As a result, the system would better capitalize on the strengths of both the human and robot. More explicitly, the system would capitalize on the problem solving skills of a human and the speed and dexterity of a robot. Thus, a robot would be able to work autonomously, while retaining the ability to request assistance when guidance is needed or warranted.

In terms of communication, a robot will be better understood and accepted if its communication behaviour more explicitly emulates that of a human. Common understanding should be reached by using the same conversational gestures used by humans, those of gaze, pointing, and hand and face gestures. Robots should also be able to interpret and display such behaviours so that their communication appears natural to their human conversational partner.

Finally, Augmented Reality has many benefits that will help create a more ideal environment for human-robot collaboration and advance the capability of the communication channels discussed. AR technology allows the human to share an ego-centric view with a robot, thus enabling the human and robot to ground their communication and intentions. AR also allows for an exo-centric view of the collaborative workspace affording spatial awareness. Multiple collaborators can be supported by an AR system; so multiple humans could collaborate with multiple robotic systems. Human-robot collaborative systems can, therefore, significantly benefit from AR technology because it conveys visual cues that enhance communication and grounding, enabling the human to have a better understanding of what the robot is doing and its intentions. A multimodal approach in developing a human-robot collaborative system would be the most effective, combining speech (spatial dialog), gesture and a shared reference of the work environment, through the use of AR. As a result, the collaboration will be more natural and more effective.

9. Acknowledgements

10. References

Argyle, M. (1967). The Psychology of Interpersonal Behavior. London, Penguin Books.

Ariadne (2006). http://www.opendialog.org/.

ARToolKit (2007). http://www.hitl.washington.edu/artoolkit/.

Azuma, R., Y. Baillot, et al. (2001). "Recent advances in augmented reality." IEEE Computer Graphics and Applications **21**(6): 34-47.

Bechar, A. and Y. Edan (2003). "Human-robot collaboration for improved target recognition of agricultural robots." Industrial Robot **30**(5): 432-436.

Billinghurst, M., J. Bowskill, et al. (1998). "Spatial information displays on a wearable computer." IEEE Computer Graphics and Applications **18**(6): 24-31.

Billinghurst, M., R. Grasset, et al. (2005). "Designing Augmented Reality Interfaces." Computer Graphics SIGGRAPH Quarterly, 39(1), 17-22 Feb.

Billinghurst, M., H. Kato, et al. (2001). "The MagicBook: A transitional AR interface." Computers and Graphics (Pergamon) **25**(5): 745-753.

Billinghurst, M., I. Poupyrev, et al. (2000). Mixing realities in Shared Space: An augmented reality interface for collaborative computing. 2000 IEEE International Conference on Multimedia and Expo (ICME 2000), Jul 30-Aug 2, New York, NY.

Billinghurst, M., S. Weghorst, et al. (1997). Wearable computers for three dimensional CSCW. Proceedings of the 1997 1st International Symposium on Wearable Computers, Oct 13-14, Cambridge, MA, USA, IEEE Comp Soc, Los Alamitos, CA, USA.

Bowen, C., J. Maida, et al. (2004). "Utilization of the Space Vision System as an Augmented Reality System for Mission Operations." Proceedings of AIAA Habitation Conference, Houston TX.

Breazeal, C. (2004). "Social interactions in HRI: The robot view." IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews Human-Robot Interactions **34**(2): 181-186.

Breazeal, C., A. Brooks, et al. (2003). "Humanoid Robots as Cooperative Partners for People." MIT Media Lab, Robotic Life Group, Submitted for review to International Journal of Humanoid Robots December 15.

Breazeal, C., A. Edsinger, et al. (2001). "Active vision for sociable robots." IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans **31**(5): 443-453.

Cassell, J., T. Bickmore, et al. (1999). "Embodiment in conversational interfaces: Rea." Conference on Human Factors in Computing Systems - Proceedings Proceedings of the CHI 99 Conference: CHI is the Limit - Human Factors in Computing Systems, May 15-May 20: 520-527.

Cassell, J., Y. Nakano, et al. (2001). Non-Verbal Cues for Discourse Structure. Association for Computational Linguistics Annual Conference (ACL).

Cheok, A. D., S. W. Fong, et al. (2003). "Human Pacman: A Mobile Entertainment System with Ubiquitous Computing and Tangible Interaction over a Wide Outdoor Area." Mobile HCI: 209-223.

Cheok, A. D., W. Weihua, et al. (2002). Interactive theatre experience in embodied + wearable mixed reality space. Proceedings. International Symposium on Mixed and Augmented Reality, ISMAR.

Chong, N. Y., T. Kotoku, et al. (2001). Exploring interactive simulator in collaborative multi-site teleoperation. 10th IEEE International Workshop on Robot and Human Communication, Sep 18-21, Bordeaux-Paris, Institute of Electrical and Electronics Engineers Inc.

Clark, H. H. and S. E. Brennan (1991). Grounding in Communication. Perspectives on Socially Shared Cognition. L. Resnick, Levine J., Teasley, S. Washington D.C., American Psychological Association**:** 127 - 149.

Clark, H. H. and D. Wilkes-Gibbs (1986). "Referring as a collaborative process." Cognition **22**(1): 1-39.

COGNIRON (2007). http://www.cogniron.org/InShort.php.

Collett, T. H. J. and B. A. MacDonald (2006). "Developer Oriented Visualisation of a Robot Program." Proceedings 2006 ACM Conference on Human-Robot Interaction, March 2-4: 49-56.

Denecke, M. (2002). "Rapid Prototyping for Spoken Dialogue Sysems." Proceedings of 19th International Conference on Computational Linguistics **1**: 1-7.

Drury, J., J. Richer, et al. (2006). "Comparing Situation Awareness for Two Unmanned Aerial Vehicle Human Interface Approaches." Proceedings IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR). Gainsburg, MD, USA August.

Fernandez, V., C. Balaguer, et al. (2001). Active human-mobile manipulator cooperation through intention recognition. 2001 IEEE International Conference on Robotics and Automation, May 21-26, Seoul, Institute of Electrical and Electronics Engineers Inc.

Fong, T., C. Kunz, et al. (2006). "The Human-Robot Interaction Operating System." Proceedings of 2006 ACM Conference on Human-Robot Interaction, March 2-4: 41-48.

Fong, T. and I. R. Nourbakhsh (2005). "Interaction challenges in human-robot space exploration." Interactions **12**(2): 42-45.

Fong, T., C. Thorpe, et al. (2002a). "Robot As Partner: Vehicle Teleoperation With Collaborative Control." Multi-Robot Systems: From Swarms to Intelligent Automata, 01 June.

Fong, T., C. Thorpe, et al. (2002b). Robot, asker of questions. IROS 2002, Sep 30, Lausanne, Switzerland, Elsevier Science B.V.

Fong, T., C. Thorpe, et al. (2003). "Multi-robot remote driving with collaborative control." IEEE Transactions on Industrial Electronics **50**(4): 699-704.

Fussell, S. R., L. D. Setlock, et al. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. The CHI 2003 New Horizons Conference Proceedings: Conference on Human Factors in Computing Systems, Apr 5-10,

Ft. Lauderdale, FL, United States, Association for Computing Machinery.

Giesler, B., T. Salb, et al. (2004). Using augmented reality to interact with an autonomous mobile platform. Proceedings- 2004 IEEE International Conference on Robotics and Automation, Apr 26-May 1, New Orleans, LA, United States, Institute of Electrical and Electronics Engineers Inc., Piscataway, United States.

Giesler, B., P. Steinhaus, et al. (2004). Sharing skills: Using augmented reality for human-Robot collaboration. Stereoscopic Displays and Virtual Reality Systems XI, Jan 19-21, San Jose, CA, United States, International Society for Optical Engineering, Bellingham, WA 98227-0010, United States.

Glassmire, J., M. O'Malley, et al. (2004). Cooperative manipulation between humans and teleoperated agents. Proceedings - 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, HAPTICS 2004, Mar 27-28, Chicago, IL, United States, IEEE Computer Society, Los Alamitos;Massey University, Palmerston, United States;New Zealand.

Hoffmann, G. and C. Breazeal (2004). Robots that Work in Collaboration with People. AAAI Fall Symposium on the Intersection of Cognitive Science and Robotics, Washington, D.C.

Honda (2007). "http://world.honda.com/ASIMO/."

Horiguchi, Y., T. Sawaragi, et al. (2000). Naturalistic human-robot collaboration based upon mixed-initiative interactions in teleoperating environment. 2000 IEEE International Conference on Systems, Man and Cybernetics, Oct 8-Oct 11, Nashville, TN, USA, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ, USA.

HRI2006 (2006). http://www.hri2006.org/.

Huttenrauch, H., A. Green, et al. (2004). "Involving users in the design of a mobile office robot." IEEE Transactions on Systems, Man and Cybernetics, Part C 34(2): 113-124.

Inagaki, Y., H. Sugie, et al. (1995). Behavior-based intention inference for intelligent robots cooperating with human. Proceedings of the 1995 IEEE International Conference on Fuzzy Systems. Part 3 (of 5), Mar 20-24, Yokohama, Jpn, IEEE, Piscataway, NJ, USA.

Iossifidis, I., C. Theis, et al. (2003). Anthropomorphism as a pervasive design concept for a robotic assistant. 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 27-31, Las Vegas, NV, United States, Institute of Electrical and Electronics Engineers Inc.

Ishikawa, N. and K. Suzuki (1997). Development of a human and robot collaborative system for inspecting patrol of nuclear power plants. Proceedings of the 1997 6th IEEE International Workshop on Robot and Human Communication, RO-MAN'97, Sep 29-Oct 1, Sendai, Jpn, IEEE, Piscataway, NJ, USA.

Julier, S., Y. Baillot, et al. (2002). "Information filtering for mobile augmented reality." IEEE Computer Graphics and Applications 22(5): 12-15.

Kanda, T., H. Ishiguro, et al. (2002). Development and evaluation of an interactive humanoid robot "Robovie". 2002 IEEE International Conference on Robotics and Automation, May 11-15, Washington, DC, United States, Institute of Electrical and Electronics Engineers Inc.

Kendon, A. (1967). "Some Functions of Gaze Direction in Social Interaction." Acta Psychologica 32: 1-25.

Kendon, A. (1983). "Gesture and Speech: How They Interact." Nonverbal Interaction. J. Wiemann, R. Harrison (Eds). Beverly Hills, Sage Publications: 13-46.

Kiyokawa, K., M. Billinghurst, et al. (2002). Communication behaviors of co-located users in collaborative AR interfaces. International Symposium on Mixed and Augmented Reality, ISMAR.

Krujiff, G.-J. M., H. Zender, et al. (2006). "Clarification Dialogues in Human-Augmented Mapping." Proceedings of 2006 ACM Conference on Human-Robot Interaction, March 2-4: 282-289.

Kuzuoka, H., K. Yamazaki, et al. (2004). Dual ecologies of robot as communication media: Thoughts on coordinating orientations and projectability. 2004 Conference on Human Factors in Computing Systems - Proceedings, CHI 2004, Apr 24-29, Vienna, Austria, Association for Computing Machinery, New York, NY 10036-5701, United States.

Maida, J., C. Bowen, et al. (2006). "Enhanced Lighting Techniques and Augmented Reality to Improve Human Task Performance." NASA Tech Paper TP-2006-213724 July.

McNeill, D. (1992). Hand and Mind: What Gestures Reveal about Thought. Chicago, The University of Chicago Press.

Milgram, P. and F. Kishino (1994). "Taxonomy of mixed reality visual displays." IEICE Transactions on Information and Systems E77-D(12): 1321-1329.

Milgram, P., S. Zhai, et al. (1993). Applications of Augmented Reality for Human-Robot Communication. In Proceedings of IROS 93: International Conference on Intelligent Robots and Systems, Yokohama, Japan.

Minneman, S. and S. Harrison (1996). A Bike in Hand: A Study of 3D Objects in Design. Analyzing Design Activity. N. Cross, H. Christiaans and K. Dorst. Chichester, J. Wiley.

Morita, T., K. Shibuya, et al. (1998). Design and control of mobile manipulation system for human symbiotic humanoid: Hadaly-2. Proceedings of the 1998 IEEE International Conference on Robotics and Automation. Part 2 (of 4), May 16-20, Leuven, Belgium, IEEE, Piscataway, NJ, USA.

Murphy, R. R. (2004). "Human-robot interaction in rescue robotics." Systems, Man and Cybernetics, Part C, IEEE Transactions on **34**(2): 138-153.

NASA (2004). The Vision for Space Exploration: National Aeronautics and Space Administration, http://www.nasa.gov/pdf/55583main_vision_space_exploration2.pdf.

Nass, C., J. Steuer, et al. (1994). Computers are social actors. Proceedings of the CHI'94 Conference on Human Factors in Computing Systems, Apr 24-28, Boston, MA, USA, Publ by ACM, New York, NY, USA.

Nourbakhsh, I. R., J. Bobenage, et al. (1999). "Affective mobile robot educator with a full-time job." Artificial Intelligence **114**(1-2): 95-124.

Nourbakhsh, I. R., K. Sycara, et al. (2005). "Human-robot teaming for Search and Rescue." IEEE Pervasive Computing **4**(1): 72-77.

Ohba, K., S. Kawabata, et al. (1999). Remote collaboration through time delay in multiple teleoperation. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'99): Human and Environment Friendly Robots with High Intelligence and Emotional Quotients', Oct 17-Oct 21, Kyongju, South Korea, IEEE, Piscataway, NJ, USA.

Prince, S., A. D. Cheok, et al. (2002). 3-D live: Real time interaction for mixed reality. The eight Conference on Computer Supported Cooperative Work (CSCW 2002), Nov 16-20, New Orleans, LA, United States, Association for Computing Machinery.

Rani, P., N. Sarkar, et al. (2004). "Anxiety detecting robotic system - Towards implicit human-robot collaboration." Robotica **22**(1): 85-95.

Reitmayr, G. and D. Schmalstieg (2004). "Collaborative Augmented Reality for Outdoor Navigation and Information Browsing." Proc. Symposium Location Based Services and TeleCartography 2004 Geowissenschaftliche Mitteilungen Nr. 66.

Roy, D., K.-Y. Hsiao, et al. (2004). "Mental imagery for a conversational robot." Systems, Man and Cybernetics, Part B, IEEE Transactions on **34**(3): 1374-1383.

Scholtz, J. (2002). "Human Robot Interactions: Creating Synergistic Cyber Forces." In A. Schultz and L. Parker, eds., Multi-robot Systems: From Swarms to Intelligent Automata, Kluwer.

Scholtz, J. (2003). Theory and evaluation of human robot interactions. Proceedings of the 36th Annual Hawaii International Conference on System Sciences.

Scholtz, J., B. Antonishek, et al. (2005). A Comparison of Situation Awareness Techniques for Human-Robot Interaction in Urban Search and Rescue. CHI 2005 | alt.chi, April 2- 7, Portland, Oregon, USA.

Sidner, C. L. and C. Lee (2003). Engagement rules for human-robot collaborative interactions. System Security and Assurance, Oct 5-8, Washington, DC, United States, Institute of Electrical and Electronics Engineers Inc.

Sidner, C. L. and C. Lee (2005). "Robots as laboratory hosts." Interactions **12**(2): 24-26.

Skubic, M., D. Perzanowski, et al. (2004). "Spatial language for human-robot dialogs." Systems, Man and Cybernetics, Part C, IEEE Transactions on **34**(2): 154-167.

Skubic, M., D. Perzanowski, et al. (2002). Using spatial language in a human-robot dialog. 2002 IEEE International Conference on Robotics and Automation, May 11-15, Washington, DC, United States, Institute of Electrical and Electronics Engineers Inc.

Sony (2007). "http://www.sony.net/SonyInfo/QRIO/story/index_nf.html."

Tenbrink, T., K. Fischer, et al. (2002). "Spatial Strategies in Human-Robot Communication." Korrekturabzug Kuenstliche Intelligenz, Heft 4/02, pp 19-23, ISSN 0933-1875, arendtap Verla, Bemen.

Toyota (2007). "http://www.toyota.co.jp/en/special/robot/."

Tsoukalas, L. H. and D. T. Bargiotas (1996). Modeling instructible robots for waste disposal applications. Proceedings of the 1996 IEEE International Joint Symposia on Intelligence and Systems, Nov 4-5, Rockville, MD, USA, IEEE, Los Alamitos, CA, USA.

Tversky, B., P. Lee, et al. (1999). "Why do Speakers Mix Perspectives?" Spatial Cognition Computing **1**: 399-412.

Watanuki, K., K. Sakamoto, et al. (1995). "Multimodal interaction in human communication." IEICE Transactions on Information and Systems **E78-D**(6): 609-615.

Yanco, H. A., J. L. Drury, et al. (2004). "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition." Human-Computer Interaction Human-Robot Interaction **19**(1-2): 117-149.