Barriers to Collaboration: User-Centered Research and the Access Grid

Emilee J. Patrick
User Centered Research Group
Applications Research, Motorola Labs
epatrick@labs.mot.com

ABSTRACT
The Access Grid (AG) links geographically separated people so that they may participate simultaneously in different types of group activities. This technology shows clear potential for engendering conditions capable of supporting true distance collaboration; however, barriers exist that keep the AG from becoming a communication medium capable of rivaling face-to-face meetings. This paper reviews literature from social psychology, anthropology and computer-supported cooperative work, suggesting possible reasons for these barriers and supporting them with evidence from three separate observations of AG events. Proposed research for addressing these issues in the coming year is discussed.

Keywords
Collaboration, communication, nonverbal, awareness, CSCW, groupware, Access Grid Node, critical mass.

INTRODUCTION

'Early adopters' of a new communications technology are the first people to begin using a new medium for communication. They tend to do so because they find the role of 'innovator' to be personally rewarding, or they expect financial benefit to come their way as the technology becomes popular. Later adopters are generally persuaded or induced to participate by early adopters; however, people in this user group may find that they do not believe in the same vision or reap the same benefits as early adopters, and are therefore more likely to discontinue use of the technology. This is the root of the problem of "critical mass" [8]; new communications technology may fail before enough users exist for all to see the potential benefits of future universal access.

Adoption of the Access Grid (AG) as a medium for group communication is still firmly within its early stages. The capability for geographically separated groups to work together across wide distances currently exists, but in order for the full collaborative potential of the AG to be realized it is necessary to first understand from a psychological perspective how collaboration happens in a traditional face-to-face setting. This knowledge will then lead to further user-centered research into system requirements, ultimately transforming the AG into a communication medium unmatched by today's video- or audio- conferencing.

NONVERBAL COMMUNICATION

Research papers concerning nonverbal communication frequently cite early work of several social psychologists in reporting that anywhere from 60-90% of the meaning in human communication is conveyed through nonverbal channels. While no quantitative method with which to validate these figures exists, it is clear that the meaning of the spoken word can be profoundly altered by a change in pitch or an accompanying glance or gesture.

Previous work has established that there are three channels of human communication: linguistic, paralinguistic, and nonlinguistic [6]. *Linguistic* communication is the explicit meaning of spoken words, excluding all other cues. The term *paralinguistic* takes into account aspects of speech such as rate and tone of voice, which modify the meaning of spoken words. *Nonlinguistic* communication includes gesture, posture and body position, facial expression, and eye gaze. Even nonverbal behaviors such as the amount of personal space maintained between two people, or chronic lateness for a particular regular meeting communicate a message. Nonverbal information is generally transmitted unconsciously, and so courses are offered through infomercials and the Internet purporting to teach people how to spot a liar.

It is important to remember that nonverbal cues are learned, not innate, and cross-cultural differences exist. The gesture used in the United States to mean 'OK' (a circle made with the with thumb and index finger) stands for 'worthless' to the French, and to the Japanese means 'money'. In a book compiled by the Department of Anthropology at William and Mary College, a professor imparts the following advice to his students in preparation for venturing out into the field for the first time [10]:

"Much has been written about how important knowing the language is to successful cross-cultural interaction, but relatively few people understand that mastering the appropriate behavior takes precedence over mastering the language."

The most important nonverbal cues used in coordinating speaking turns are eye gaze and head turning [9]. A person in the role of the speaker tends to make short glances at regular intervals toward the face of their listener, while the listener directs his or her gaze toward the face of the speaker for the entire duration of the utterance. As the speaker nears the completion of their statement, their gaze rests on the listener, signaling that they are finished speaking [6]. Speakers use this gaze mechanism to direct their speech toward a particular person, and to indicate that they want that person to speak next.

AWARENESS

People working together in a shared physical workspace seem to take *workspace awareness* for granted. The past knowledge about people, objects, and activities in the shared space, and the sensitivity to spatial relationships in the immediate surroundings that is absorbed simply by being physically co-present enables conversation about the work, coordination of actions and plans, and anticipation of others' needs [7]. Not only is knowledge of the workspace from one's own perspective important, but awareness of what is in the visual field of other members of the workgroup is essential for building a shared frame of reference.

Gutwin and Greenberg (2001) list ten elements of workspace awareness, grouped into three categories [7]:

WHO
- **Presence:** is anybody there
- **Identity:** who is that
- **Authorship:** who's doing that

WHAT
- **Action:** what are they doing
- **Intention:** what's the end goal
- **Artifact:** what object are they working on

WHERE
- **Location:** where are they working
- **Gaze:** where are they looking
- **View:** where can they see
- **Reach:** where can they reach

Teammates gain this information through visible activity, nonverbal information, artifact use and placement, and explicit conversation (both formal and informal) that takes place in the workspace as a natural part of the work.

Workspace awareness is required for the success of two social processes essential for collaboration: grounding and perspective-taking. *Grounding* refers to the creation of a shared understanding between conversation participants about a specific item, idea or emotion. Paralinguistic and nonlinguistic feedback from listeners helps a both the speaker and the listener understand whether or not what was said has been understood.

Perspective-taking is a broader activity involving the creation of "a shared communicative context, in which each participant is aware of the point of view of the others" [6]. This process not only understanding the physical point of view of other people in the shared work space, but also their background and social role within the group.

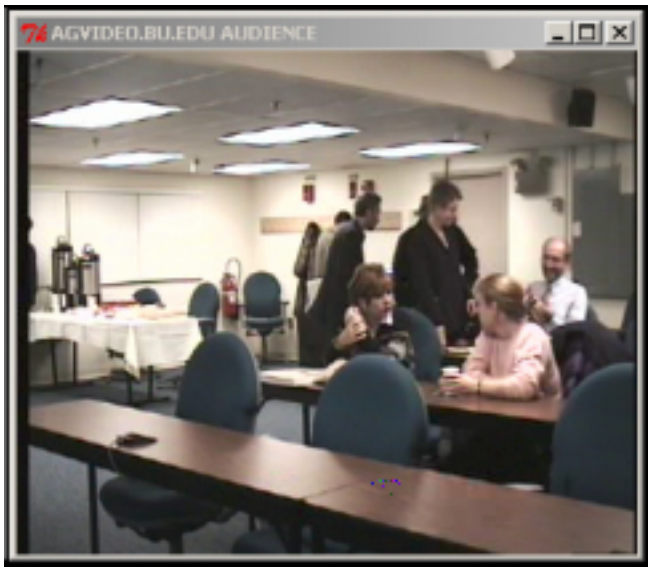

Figure 1: Attendees arriving for an AG Meeting

SOME AG OBSERVATIONS

The AG partially supports ways in which people gather and maintain workspace awareness and coordinate speaking turns, but there are many opportunities for improvement.

Eye Contact

Asymmetry in eye contact is the situation that occurs when a speaker directs their gaze toward the representation of the listener that appears on the display wall, instead of toward the video camera. The listener in a remote location watching the video feed coming from the speaker's location becomes slightly disoriented by the image of the speaker looking off to the side while addressing them.

A workaround used by AG operators entails placing the corresponding image of the listener as near as possible to the camera aimed towards the speaker, in order to simulate eye contact, which in turn facilitates smooth speaker transitions. This solution works well for interactions among a very small group of people, but does not scale well as the group becomes larger. In a situation where there is a large co-located group at one Access Grid node, a choice must be made regarding the level of awareness that will be conveyed to remote sites. A tradeoff exists between using a narrow camera angle to focus closely on a few selected participants, or zooming out to reveal more audience

members in less detail. In either situation, it is likely that coordinating speaking turns will require some effort on the part of the node operator to position cameras and video feeds in a useful and usable configuration.

Video Quality and Image Size

Video quality and image size have a great impact on the AG's potential for supporting collaboration. Low video quality and small video feed window size can result in the loss of important details for distinguishing facial expressions and other "back-channel" nonverbal responses necessary to the grounding process. This is especially problematic in the situation mentioned above, where an AG space is supporting a meeting where there are more than two or three people per camera. In a room with many people, facial expressions can be lost as the camera zooms out to capture everyone present (See Figure 1).

According to Ferraro's rules for personal space, recognition of others in a 'public' space such as a hallway at work is not mandatory because "subtle shades of meaning of voice, gesture, and facial expression are lost" when encountering people at distances of 12-25 feet (4-9 meters) [4]. This statement seems to fairly accurately describe the reaction of novice AG users to the people 'present' on the display wall. Images on the screen are treated as a passerby in the hallway might be – noticed, but without interaction.

When planning a node, care should be taken in advance to determine the number of people who will use the physical space at a given time. Sufficient hardware should be on hand to capture all participants in enough detail that gestures, posture and some facial expressions are transmitted.

Grounding

As mentioned previously, poor video quality can impede the grounding process between AG meeting participants in geographically distant locations. This process can also be made more difficult due to the fact that participants do not control the AG display themselves – they must depend upon node operators to arrange the display in such a way that all of the relevant video feeds are visible. In situations where speech from another location is audible but no corresponding video feed is visible on the display, operators rely on personal experience to guide them in searching for the correct video feed. One operator mentioned becoming familiar with locations that are frequently active on the AG through experience, and recognition of another operator's voice, as two methods used when attempting to locate video feed that is audible but not visible.

Node operators also employ a workaround for creating a shared understanding of what is in the visual field of participants in other locations. One video camera is dedicated to recording the display wall at each node. Figure 2 is an example from an AG event during the Supercomputing 2000 Conference, showing simultaneous video feeds from six different nodes. Flexibility in window size and placement allows differences from site to site that can make it very difficult to make sense out of the appearance of other displays.

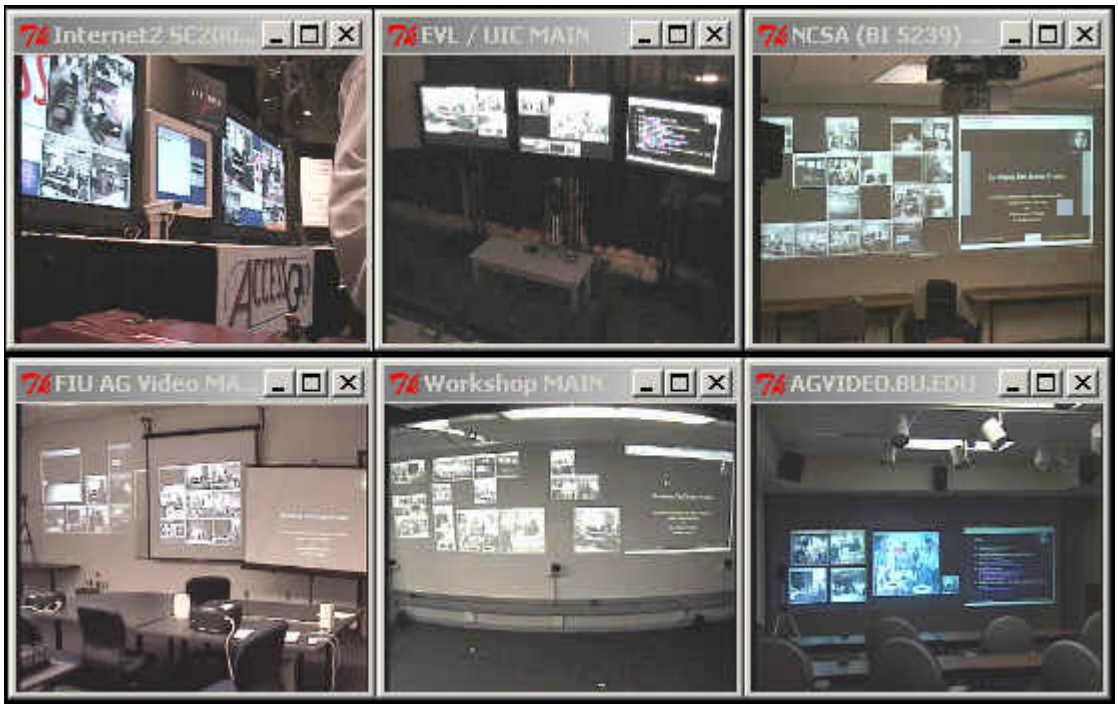

Figure 2: Six simultaneous screen captures of AG display camera feeds, during Supercomputing 2000 (11/7/00, 12:43 PM)

Perspective-taking

The lack of fundamental information about the identity of remote meeting participants is an impediment to successful perspective-taking on the AG. The window name appearing at the top of each video feed may not be clearly understood by a novice user, meaning that at times it is difficult to tell where the video originates. Also, there is no visible description or labeling of the identities of individual participants in each location. In a physical meeting, all participants would be visible at all times, and it is common meeting participants to introduce themselves to the group at the start of a meeting.

Finally, as one node operator wisely said, "If they [the remote participants] can't see you, you're not there." It is not possible to build a shared understanding with an undetected AG meeting 'participant'. This lurking behavior could be unintentional; unfamiliarity with node operations could produce a situation where local cameras are disabled or positioned badly so as to effectively 'hide' an individual person. However, it should be a cause for some concern that it is potentially possible for someone to intentionally 'eavesdrop' on an AG meeting in the manner described above. Encryption may exclude remote sites without permission from entering a virtual venue, but once a site has connected meeting participants must rely on local security measures.

NEXT STEPS

Two parallel areas of research will be addressed by Motorola Labs in the coming year: 1) continued observational data collection to contribute to existing knowledge about how the AG is currently being used, and 2) experimental studies investigating methods for providing improved support for nonverbal cues and smooth transitions of speaking turns.

Takao (1999) suggests several methods for enhancing the transmission of nonverbal information in a desktop videoconferencing system while assisting with speaker transitions. He posits that using a video switching scheme controlled by a human session chair or automatically via audio detection might make it easier for participants trying to follow the conversation. However, the potential exists for information to be lost when speech overlaps (as it often does even in face to face conversations).

Because the AG supports a much larger display area than a single desktop monitor, it is not necessary to switch completely between video feeds to conserve screen real estate, and still provide a large enough image to transmit nonverbal cues. Initial work at Motorola will concentrate on prototyping and testing other methods for assisting smooth speaker transitions, including determining the impact of different levels of video quality on turn-taking and overlapping speech. Interface cues to make it easier for audience members to focus on the window containing the remote speaker will be tested.

Ongoing observational research will focus on recording AG meetings, and coding observed speech patterns in order to pinpoint existing barriers to collaboration.

REFERENCES

1. Barkhi, R., Jacob, V., Pipino, L. and Pirkul, H. The Effect of Communication Mode and Leadership Mode on Group Decisions. *Association for Information Systems 1996 Americas Conference*. Available at http://hsb.baylor.edu/ramsower/ais.ac.96/papers/barkhi.htm
2. Bavelas, J.B., Hutchinson, S., Kenwood, C., and Matheson, D.H. Using Face-to-face Dialogue as a Standard for Other Communication Systems. *Canadian Journal of Communication 22*, 1, 5-24.
3. Cheng, K-E. & Rana, A.R. Group Support Systems and the Speech Act Theory. *Association for Information Systems 1997 Americas Conference*. Available at http://hsb.baylor.edu/ramsower/ais.ac.97/papers/cheng.htm
4. Ferraro, G.P. Communicating across cultures: The nonverbal Dimension. In: *The Cultural Dimension of International Business*. 68-91. Prentice Hall: Englewood Cliffs, NJ, 1990.
5. Fulk, J. and Collins-Jarvis, L. Wired Meetings: Technological Mediation of Organizational Gatherings. In: Putnam, L. & Jablin, F., (eds.), *New Handbook of Organizational Communication*. Newbury Park: Sage, 2000. Available at http://www-rcf.usc.edu/~fulk/hbchap.htm
6. Fussell, S. R., & Benimoff, N. I. Social and cognitive processes in interpersonal communication: Implications for advanced telecommunications technologies. *Human Factors 37* (1995), 228-250.
7. Gutwin, C., and Greenberg, S. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work*, Kluwer Academic Press, 2001.
8. Markus, M.L. Towards a "Critical Mass" theory of Interactive Media. In: Fulk, J. and Steinfeld, C. (eds.), *Organizations and communication technology*. Sage Publications: London, 1990.
9. Sellen, A.J. Speech patterns in video-mediated conversations. *Proceedings of ACM CHI 92 Conference on Human factors in computing systems*, 49 – 59.
10. Serrie, H. Cross Cultural Interaction: some general considerations for innocents abroad. In: Sutlive, V.H., Altshuler, M., Zamora, M.D, & Kerns, V. (eds.), *Studies in Third World Societies: Anthropology and International Business*. Department of Anthropology: College of William and Mary, 1986.
11. Takao, S. The effects of narrow-band width multipoint videoconferencing on group decision making and turn distribution. *Proceedings of the 1999 international joint conference on Work activities coordination and collaboration* , 109-116.