PING: A GROUP-TO-INDIVIDUAL DISTRIBUTED MEETING SYSTEM

Yong Rui, Eric Rudolph, Li-wei He, Rico Malvar, Michael Cohen and Ivan Tashev

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

ABSTRACT

Group-to-individual (G2I) distributed meeting is an important but understudied area. Because of the asymmetry between different parties in G2I meetings, it has two unique challenges: 1) the remote participant tends to be ignored by the local participants; and 2) the remote participant has inferior audio, video, and data experience than the local participants. To address these issues, in this paper we present PING, a system explicitly designed for G2I distributed meetings that combines recent advances in both hardware, e.g., microphone arrays, remote person stand-in devices, and software, e.g., audio-video processing, to improve users' G2I meeting experience. We report how PING addresses the above two challenges and its system design and implementation.

1. INTRODUCTION

With industry's globalization trend, more and more teams are becoming distributed. How to develop collaboration tools to make distributed teams more productive is thus important. Distributed meetings are among those tools and they have different types:

- Group-to-group (G2G): each group resides in its own meeting room.
- Individual-to-individual (I2I): each individual is in his/her own office.
- Group-to-individual (G2I): a group resides in a meeting room and an individual is in his/her own office.

G2G scenario is addressed by the traditional video conferencing systems, e.g., Polycom [5] and Tandberg [10]. I2I scenario is addressed by various emerging collaboration solutions, e.g., Web conferencing services [11], and audio-visual enhanced instant messaging (IM) systems [4]. For G2I, it received relatively less attention. It is worth pointing out that while G2G systems can be used for G2I, it is an unnecessarily expensive (e.g., $20,000+) and inconvenient solution. Similarly, while people could use I2I systems for G2I scenarios, the experience is far from satisfactory, e.g., the I2I camera has limited field of view – instead of covering the whole group, it can only cover a single person. Note also that there is a major difference between G2I and G2G/I2I, and this difference imposes extra challenges for G2I. While different parties in G2G or I2I are *symmetric*, parties in G2I are not. In G2I, the group of people who sit in the same meeting room are the *dominant* party and the remote individual is the *subordinate* party. This *asymmetry* results in two unique challenges in G2I:

1. Remote participants tend to be ignored by the local (in the room) participants. How to present remote participants to those in the meeting room so that they are all on equal footing?
2. Remote participants do not have the same audio, visual, and data experience as local participants. How to develop advanced software and hardware so that remote participants feel more like "being there" in the local meeting room?

To address these issues, we developed a G2I distributed meeting system called PING (Pervasive Information Networking for Groups), that supports real-time audio-visual communication and data collaboration. Through PING, we want to bring the recent advances in hardware and signal processing technologies together, explicitly model the unique features in the G2I scenario, and design a lightweight user interface (UI) layer to facilitate end users' experience. In addition, we want to significantly reduce the overall system cost by leveraging existing PC infrastructure, using inexpensive hardware, and moving the smarts to software. The rest of the paper is organized as follows. In Section 2 we describe the remote person stand-in device and how it can represent a remote participant in the local meeting room through advanced audio-visual hardware design. In Section 3, we present advanced audio, video and data processing algorithms, e.g., microphone array beam forming, live whiteboard capture, etc., so that remote participants feel more like "being there". In Section 4, we describe the PING architecture and implementation, and also how to seamless integrate local laptops into the overall meeting experience. We conclude the paper in Section 5.

2. REMOTE INDIVIDUAL'S PRESENCE

In G2I, one of the biggest complains that a remote person has is that he/she tends to be ignored by the local people, because he/she does not have a proper representation. There was previous research on how to represent remote people in a local room. For example, Hydra [7] uses a Sony Watchman monitor (8 cm diagonal), a Radio Shack camera, and a Sony Watchman speaker to represent and display a remote person. While Hydra influenced our design, given the hardware and software advances over the past decade, PING's solution provides much more sophisticated audio-visual experience for both the local people and the remote people in terms of the size of the live video, the field of view of the video, the quality of the microphone array, and its ability to localizing the sound source. In addition, the original Hydra system was designed for the I2I scenario [7], while PING focuses on the G2I scenario, which has *asymmetry* between different parties.

Our proposed solution in PING is the *remote person stand-in device.* As its name suggests, it serves as a remote person's representation in the local meeting room (Figure 1). The remote person appears in the monitor at about life size and at similar position and height as the local participants. As discussed in the introduction section, we want the remote and the local participants to be on equal footing in the meeting room. So, it is important that the remote person appears near life-size, and that he or she is

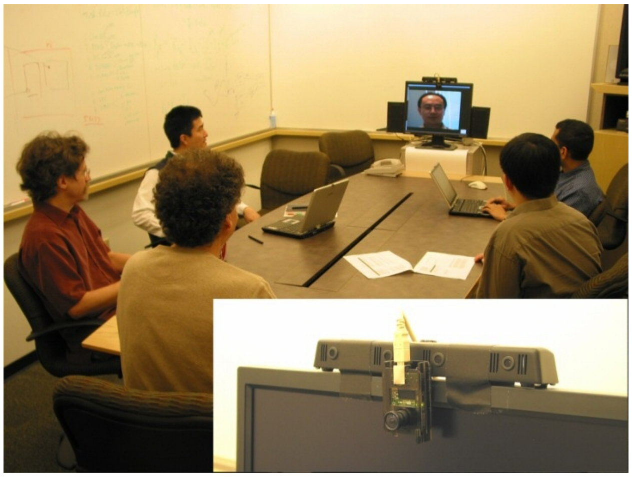


Figure 1. The meeting room; the PING stand-in device is at the end of the table, and consists of a flat panel monitor, a linear microphone array, a wide-angle camera, and two loudspeakers.

presented with a view of the meeting that is approximately the same as that one would get if sitting in the meeting itself. To achieve these goals, we selected a 20” LCD monitor. A wide-angle camera, a microphone array and two loudspeakers are placed as close to the monitor surface as possible to facilitate eye-to-eye contact, and to allow sound from the remote participant to appear to come from the monitor. As we will discuss later, the wide angle camera, when properly warped, provides a subjective experience much like sitting at the end of the table for the remote participant. Likewise, the microphone array provides the means to both focus the microphones “beam” and also to help detect the speaking person in the meeting room to automate a digital pan-tilt-zoom (PTZ) within the high resolution video.

The wide-angle camera in PING has a field of view of 110 degrees and is built on an off-the-shelf 2-mega-pixel OmniVision image sensor. It is worth pointing out that digital PTZ is significantly less intrusive than the mechanical PTZ used in most traditional video conferencing systems. The microphone array has a linear configuration with a length of 195 millimeters; it consists of four Panasonic WM-55A cardioid microphones. The microphone array’s geometry is symmetric and the distance between the central microphones is 55 millimeters. The microphone array is designed as an external USB device. The fact that all of the components of the system above can be built inexpensively demonstrates an important shift in the potential to assemble such a device in every meeting room.

3. REMOTE PERSON’S EXPERIENCE

As we discussed in the introduction section, a remote participant does not have the same audio, visual or data experience as local participants. To bridge that gap, PING contains several innovative signal processing algorithms that improve audio, video, and data collaboration.

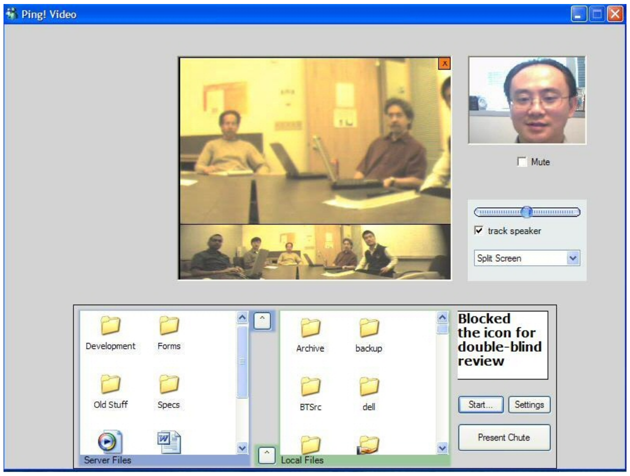
3.1. Microphone array audio processing

High-quality audio is the key to successful distributed meetings. Our audio processing engine consists of an acoustic echo canceller (AEC), a microphone array processor, a noise suppressor and an automatic gain control (AGC). The microphone array processor combines the signals from all microphones using our novel beamforming technology explicitly designed for meeting room audio capture [9]. It makes the microphone array act as a highly directional microphone, capturing sound from the current speaker, removing the ambient noises and reducing the room reverberation. The remaining stationary noise is further reduced by the noise suppressor. The overall noise reduction (from both beamforming and noise suppression) of the audio exceeds 26 dB over the output of an omni-directional microphone [9]. The system captures well human voices from up to 4 meters away, with an acceptable signal-to-noise ratio, typically better than 20 dB. The AGC further equalizes the levels of the output signal when speakers are at different distances. Not only our audio processing modules are effective, they are also light weight – they use less than 10% of the CPU of a modern PC [9]. While good audio capture is important for both offices and meeting rooms, because of the meeting room’s larger size, it is especially important in meeting rooms.

3.2. Digital PTZ

The microphone array not only can use beamforming to enhance captured audio quality, it can also localize who is talking [9]. In [6] it is shown that when a meeting participant is remote, s/he wants to both see the whole meeting room to set the context (e.g., how many people are there and who they are), and also focus on the person who is currently talking. In the meeting room, we use a wide-angle camera as the video input device and developed a digital PTZ solution by cropping out the region of the video frame that contains the current speaker. Because the sensor on the video camera contains 2 mega pixels, even the cropped region contains high-fidelity imagery. By using this approach, we can show both the whole room in low resolution and the speaking person in high resolution (Figure 2). This approach simulates what would happen if the remote person were sitting in the local meeting room: we human have wide peripheral vision (almost 180 degrees) in low resolution to get the context and the fovea in our eyes only focuses on a narrow range with very high resolution to get the details.

3.3. Wide angle camera video de-warping



Ping! Video

☐ Mute

☑ track speaker

Split Screen

Blocked the icon for double-blind review

Figure 2. The view from the remote person; the upper-center portion is a live view from the meeting room: the person who is talking is shown at the top with high resolution, and the whole meeting room is shown at the bottom with low resolution. The upper-right portion shows the video of the local person him/herself. The bottom portion shows the UI for file sharing and transfer.

A unique feature for G2I is that there is a single camera but multiple people in the meeting room. People appear larger when they sit close to the camera and appear smaller when they are far away. This is especially true when the camera has a wide angle in order to capture the whole meeting room. Figure 3 shows a cylindrical projection of a meeting room which is equipped with a wide-angle camera. The big variance between people's head sizes gives the remote person inferior visual experience, e.g., cannot tell a far-away person's facial expression. To address this issue, in PING, we develop a spatially-varying-uniform scaling function **Error! Reference source not found.** to roughly equalize people's head sizes without causing undue distortion (Figure 4). The warped video gives the remote person a more natural visual experience of "being there".

3.4. Live whiteboard capture

Sections 3.1-3.3 discussed how PING can enhance remote person's audio-visual experience. In this section, we will describe how PING brings other meeting artifacts to remote participants. A physical whiteboard is an effective and easy-to-use tool for meetings, especially in scenarios such as brainstorming and project planning. However, in traditional audio and/or video conferences, remote participants cannot see the physical whiteboard content. To enable that, existing equipment requires instrumentation either in the pens (such as Mimio from Virtual Ink [3]) or in the whiteboard (such as SMARTBoard from SmartTech [8]). Unfortunately, these systems force users to change their normal behavior. For example, when a user writes on the whiteboard using instrumented Mimio marker, he may use his hand to erase/correct small content on the whiteboard. But even though the physical content is erased, the digital content is not – the user has to pick up an instrumented eraser to erase the digital content. We consider this as a burden to the user.

In PING, we allow the user to write freely on un-instrumented whiteboard surface using a regular dry-erase pen. To achieve this, our system uses an off-the-shelf 1.3-mega-pixel Aplux MU2 video camera, which captures images of the whiteboard at 7.5 frames/sec. From the input video sequence, our computer vision algorithm separates people in the foreground from the whiteboard background and extracts the pen strokes as they are deposited to the whiteboard [1]. Furthermore, the images are white-balanced and color-enhanced for greater compression rate and better viewing experience than the original video, as shown in Figure 5. Our system can decide to remove the foreground person, keep the foreground person or semi-transparently show both the foreground person and the whiteboard content [1]. The whole process is automatic and real-time, which significantly enhances the meeting experience for people both in the meeting room and remote.

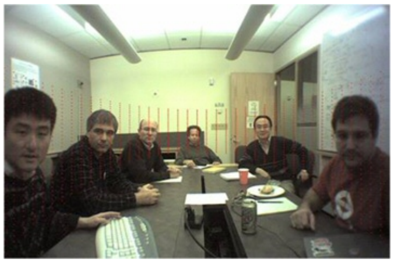


Figure 3. Stand-in device video before de-warping. The vertical lines show the warping mapping function.

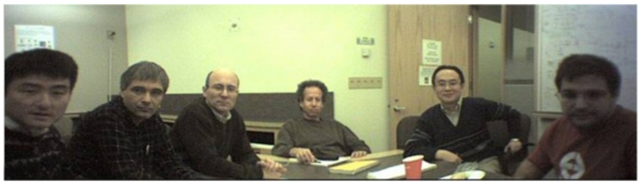


Figure 4. Same video, after warping.

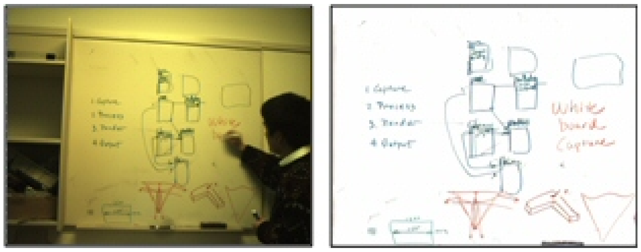
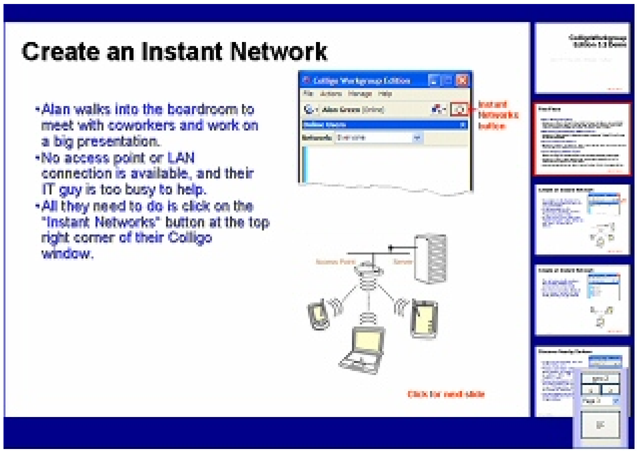


Figure 5. Live whiteboard capture. The image on the left is one of the raw image sequence captured by the whiteboard camera; the image on the right is the processed and enhanced image.

3.5. Data collaboration

Effective data collaboration is as important as audio-visual communication in meetings. For a team of workers on a joint project, they normally have a public team data folder. The bottom-left portion of Figure 2 shows this folder (called Server Files). The bottom-middle portion of the UI is the folder of files on one's local PC or laptop (called Local Files). Server files are accessible to all the participants in the meeting, while local files are private to the PC owner. Files can be easily transferred between public and private spaces by a simple drag-and-drop operation.

File transfer/copy only addresses part of the story. Sometimes a meeting participant does not want others to have a copy of the file, because it is either still in progress or confidential. However, he/she may want to allow others to *view* the document for the current discussion. This is a common practice that PING wants to support. Currently, PING supports file viewing for PowerPoint,



Create an Instant Network

- Alan walks into the boardroom to meet with coworkers and work on a big presentation.
- No access point or LAN connection is available, and their IT guy is too busy to help.
- All they need to do is click on the "Instant Networks" button at the top right corner of their Colligo window.

Figure 6. The PING viewer shows PowerPoint presentations. A participant is allowed to be out-of-sync with the presenter. On the right are the thumbnail slides showing at what point the presenter is. The participant can click on buttons at the lower-right corner to sync back to the presenter or to skip forward or back.

Word, and other printable file formats, as well as images. When a file is dragged into the “Present Chute” button**,** shown in bottom-right portion of Figure 2, the “image” version of the file is sent to all the PING clients connected to the meeting. The client then launches a “PING viewer” (Figure 6) to show the file. The owner of the file has control on “Prev”, “Next”, etc. operations. However, s/he could also allow others to be out of sync, e.g., skipping ahead. Each PING viewer has a “Sync” button (bottom-right portion of Figure 6). When clicked, it will sync back to the owner’s view of the document. Each user can also make annotations on the viewer, such as a text comment or a scribble. They will be seen by all other PING clients connected to the meeting. In summary, PING provides flexible choices, e.g., copy vs. present, in-sync vs. out-of-sync viewing, to enhance users’ data collaboration experience.

4. SYSTEM IMPLEMENTATION

PING is built on top of Microsoft Windows. The architecture of PING consists of four layers. At the very bottom are the Windows SDKs, including those for the Microsoft Real Time Communication (RTC), Windows SharePoint Services (WSS), and other Microsoft Office SDKs. Above those are the PING plug-ins that handle media capture: audio, video, live whiteboard, and Microsoft Office documents, e.g., PowerPoint and Word. PING Core is the middleware that handles basic send/receive audio/video services. The top layer includes the PING client application and the PING UI (Figure 2). A PING client application can be launched from a meeting room PC (which connects to the remote person stand-in device), a remote participant’s PC, or laptops brought to the meeting room. They have different features to enhance the overall meeting experience.

4.1. PING client on meeting room PC

The PING client running on meeting room PC is a special version, as it also serves as the meeting server software control module. It manages all the regular PING clients and file send/receive requests. It is also responsible for generating unique URL ID for different meetings such that only authorized people can participate in the meeting and have access to the meeting files. Specifically, it performs the following operations:

- Meeting initiation.
- Meeting security (who is allowed to join the meeting and/or transfer/present documents).
- Keeping track of meeting “state” events, e.g., who is presenting, what is being presented, and mouse cursor movements.
- Receiving and re-broadcasting “live” messages such as annotations, page flips, live whiteboard data, etc.
- Receiving, storing, and re-broadcasting presentation pages to the non-presenters.
- Keeping track of meeting settings, such as the public folder location, what was presented, by who and when, and keeping a log file of such information.

4.2. PING client on laptops

In the meeting room, some people may bring laptops while others do not. PING supports both cases but gives those who bring laptops some extra choices. Unlike the remote PC, the laptops in the meeting room do not need to send or receive audio and video. By factoring out the audio-visual UI, PING client can be run as a standalone application in the laptops in the meeting rooms. The stand-in device has dual functionalities, and can be shared between showing live video of remote people and showing content, e.g., PowerPoint file, live whiteboard content, etc. The PING client running on the meeting room PC has a special single-click button to allow switching between modes. When people bring their laptops to the meeting room, the stand-in flat panel can be dedicated to showing remote people and the PING client running on the laptops can be used to show file content. Even if the stand-in flat-panel monitor shows file content, the laptops can still be used to skip ahead in the slide deck (see Section 3.5 and Figure 6). This mechanism better integrates the local laptops into the overall meeting experience.

5. CONCLUSIONS

G2I distributed meeting is an important but under-served area. Through PING, we have brought the recent advances in hardware and signal processing technologies together, explicitly modeled the unique features in the G2I scenario, and designed a lightweight user interface (UI) layer to facilitate end users’ experience. We summarize the challenges and solutions as follows:

- Problem: remote people tend to be ignored. Solution: stand-in device acting as a representative for remote people.
- Problem: remote people do not have the same audio experience as local people. Solution: microphone array, noise suppression, and AGC for high-fidelity audio capture.
- Problem: remote people do not have the same visual experience as local people. Solution: wide-angle camera combined with video de-warping and digital PTZ to provide both the peripheral view and foveated view.
- Problem: if someone writes on whiteboard, remote people cannot see. Solution: live whiteboard capture and transmission.
- Problem: in the meeting room, some bring laptops and others do not. Solution: support for both sets of people. Even no one brings a laptop, the experience is still good. If someone brings a laptop, s/he will enjoy additional features (e.g., skipping ahead on slides).

6. REFERENCES

[1] L. He, Z. Liu and Z. Zhang, Why take notes? Use the whiteboard capture system, *Proc. ICASSP*, 2003

[2] Z. Liu, and M. Cohen, Head-size equalization for better visual perception of video conferencing, *Proc. IEEE ICME*, July, 2005, Amsterdam, The Netherlands.

[3] Mimio, http://www.mimio.com

[4] MSN Messenger, http://messenger.msn.com/

[5] PolyCom, http://www.polycom.com

[6] Y. Rui, A. Gupta and J.J. Cadiz, Viewing meetings captured by an omni-directional camera, *Proc. of ACM CHI 2001* , Seattle, WA, March, 2001.

[7] A. Sellen, Speech patterns in video-mediated conversations, *Proc. of ACM CHI*, 1992, pp 49-59.

[8] Smart Technologies, Inc, http://www.smarttech.com

[9] I. Tashev, H. S. Malvar. "A new beamformer design algorithm for microphone arrays". *Proc. of ICASSP*, Philadelphia, PA, USA, March 2005

[10] Tandberg, http://www.ivci.com/videoconferencing_tandberg.html

[11] WebEx, http://www.webex.com