The Effects of Multimodal Mobile Communications on Cooperative Team Interactions Executing Distributed Tasks

Gregory M. Burnett[1], Andres Calvo[2],Victor Finomore[1], and Gregory Funke[1]

[1] Air Force Research Laboratory, 711 Human Performance Wing, WPAFB, OH
[2] Ball Aersopace, Fairborn, OH
{gregory.burnett,andres.calvo.ctr,victor.finomore,
gregory.funke}@wpafb.af.mil

Abstract. Mobile devices are rapidly becoming an indispensible part of our everyday life. Integrated with various embedded sensors and the ability to support on-the-move processing, mobile devices are being investigated as potential tools to support cooperative team interactions and distributed real-time decision making in both military and civilian applications. A driving interest is how a mobile device equipped with multimodal communication capabilities can contribute to the effectiveness and efficiency of real-time, task outcome and performance. In this paper, we investigate the effects of a prototype multimodal collaborative Android application on distributed collaborating partners jointly working on a physical task. The mobile application's implementation supports real-time data dissemination of an active workspace's perspective between distributed operators. The prototype application was demonstrated in a scenario where teammates utilize different features of the software to collaboratively assemble a complex structure. Results indicated significant improvements in completion times when users visually shared their perspectives and were able to utilize image annotation versus relying on verbal descriptors.

Keywords: Multimodal interfaces, mobile computing, remote collaboration.

1 Introduction

Today's global and on-the-move workforce relies heavily on digital, mobile communication technology and its expanding interconnected networks to effectively accomplish task. Although modern workers are connected to vast amounts of information though the internet and domain specific databases, they can still encounter scenarios and situations that are outside their expertise. Often workers are required to complete the tasks on their own accord or must wait for the arrival of an expert for additional assistance, both unfavorable in time sensitive situations.

Before the advent of mobile devices such as smartphones and tablets, obtaining remote guidance was limited to information relayed orally between a worker and a remote helper, usually transmitted using a radio or telephone. However, as noted by

N. Streitz and C. Stephanidis (Eds.): DAPI/HCII 2013, LNCS 8028, pp. 358–367, 2013.

Wickens, Vidulich, and Sandry-Garza (1984), communication of spatial information is often more effective through a visual, rather than verbal, medium. Consequently, a collaborative technology which affords users the ability to represent and transmit spatial information pictorially, may positively impact performance. Transmitting spatial information in the same modality may result in reduced uncertainty and misunderstanding concerning the content of a message, thereby allowing users to more accurately and succinctly convey information about current and projected future states. Our objectives were to design and implement a prototype multimodal mobile application for remote collaboration. We evaluated the effects of shared video and audio communication on cooperative team performance towards the completion of an abstract building task.

Clark and Brennan (1991) discuss that in order to effectively collaborate, distributed pairs need to have an interactive dialogue to gain mutual understanding and form a common ground. This concept of common grounding, or clarity of instructional directives, can be achieved through various modalities. Gergle et al. (2004) report that "communicative information can be provided in the form of linguistic utterances, visual feedback, gestures, acoustic signals, or a host of other sources; all of which play an important role in successful communication" (p. 487). Utilizing multiple sources of information through multiple modalities is more effective than relying on a single source, such as verbal communication, to arrive at a common ground (Wickens & McCarley, 2008). For situations where the environment or situation can change abruptly, the ability to leverage several modalities and sources of information to maintain shared common grounding or situation awareness between the distributed parties is critical for a successful outcome.

In this paper, we discuss the implementation and evaluation of a prototype multimodal mobile application seeking to facilitate remote collaboration. In sections 2 and 3, we discuss related work and discuss the multimodal communication features chosen for inclusion into our mobile application. The purpose of these features is to support efficient communication grounding between collaborating remote partners working towards the completion of a physical task. In section 4, we report the evaluation methodology and the results from an interactive demonstration of the mobile application, where teams cooperatively built complex models from building blocks. Finally, a discussion and future work section highlight how this research can be used to provide design and potential deployment of real-time decision making capabilities supporting distributed collaboration.

2 Related Work

Recent Computer Supported Cooperative Work (CSCW) research highlights several multimodal communication capabilities that show performance benefits when providing a shared perspective between Workers and remote Helpers. This is highlighted in a series of research efforts (e.g., Gergle et al., 2004; Fussell et al., 2000; Kirk et al., 2007; Kraut et al., 2002; Kuzuoka, 1992) that have leveraged streaming video with bi-directional audio between collaborating team members.

Kuzuoka's (1992) evaluation suggests that when Helpers are provided a shared perspective of the Worker's active focus, they have better situation awareness of the task and can provide improved guidance specific to the Worker's current needs. In addition, a shared perspective of the Worker's activities allows the Helper to monitor and assess the Worker's comprehension and accuracy (Kraut et al., 2002). Studies by Gergle (2005) and Fussell et al. (2000) suggest that the utility of sharing visual information positively affects the communication dialogue between Workers and Helper making the shared linguistic communication faster, less explicit, and more proactive then using audio alone during task completion.

Building upon sharing streaming video and bi-directional audio, the ability to share real-time markup annotations has been shown to improve cooperative performance on physical task completion (Kirk et al., 2007; Ou et al., 2003; Stevenson et al., 2008). Ou and colleagues' (2003) collaborative system DOVE (Drawing Over Video Environments), permitted remote Helpers to draw markup illustrations on shared video to assist in remote guidance. Their findings suggest that the utility of markup capability "significantly reduces performance time compared to camera alone." (p. 248). Kirk et al. (2007) and Stevenson et al. (2008) both describe systems that project remote gestures and markups on top of the Worker's workspace containing the physical task's objectives. Their research showed that task completion times were shorter and fewer mistakes were made when utilizing markup capabilities in conjunction with shared visual and auditory information.

A distinction that our current research makes from existing CSCW systems is the implementation of the various multimodal communication capabilities into the mobile domain. Specially, software development in the Android operating system supporting communication features executing on a mobile device thus enabling on-the-move, anytime, anywhere team collaboration. Based on previous work that demonstrated effective use of multimodal technology to foster distributed collaboration, we built and integrated custom software to enable, sharing video of the Worker's workspace, sharing full-duplex audio between users, and the support of markup annotation on captured still images in our prototype tool suite. The overarching goal of this research and development effort is to leverage multimodal mobile capabilities to establish and maintain shared awareness, provide precise guidance, and facilitate effective collaboration in a real-time, distributed task.

3 Implementation

We implemented the Worker's mobile application on a Samsung Galaxy Tablet running Android and the Helper's application on a personal computer running Windows 7. The mobile device's back-facing camera was used to capture a series of 800×600 images. Acquired images were then compressed into JPEG format, and transmitted to the remote Helper at an average rate of 30 frames per second. Our system also incorporates full-duplex audio communication. To transmit audio to the Worker, the Helper presses and holds a software button referred to as the push-to-talk button. Similarly, the Worker presses and holds a push-to-talk button to transmit audio. Initially, we considered

transmitting audio continuously by default, but we decided to incorporate the push-to-talk buttons to lower network consumption. The system transmits both audio and video using the UDP network protocol.

The mobile application's interface displays the Worker's video on the left half of the screen. The Helper can capture images from the Worker's video, annotate them, and send them back to the Worker, where they are displayed on the right half of the screen. The Helper annotates images using drawing tools similar to Microsoft Paint. Every time the Helper draws on the image, the Worker sees the updates in real time. Figure 1 illustrates the process of capturing and annotating an image.

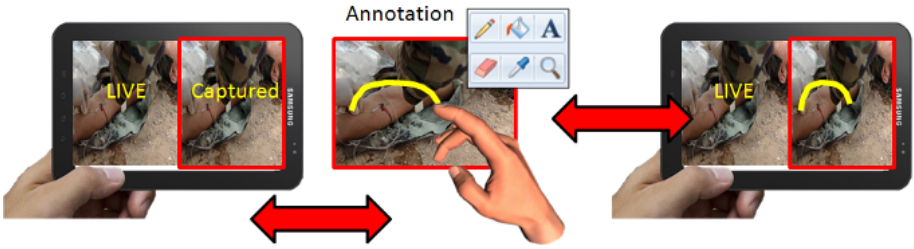

Fig. 1. The Helper captures an image from the Worker's video and annotates it to instruct the user on how to perform a task

4 Evaluation

To evaluate the effectiveness of the prototype mobile collaborative system two person teams had to work together to construct a multi-level, abstract structure with building blocks. The two people were separated from each other and had to utilize different features of the mobile collaborative system to build the structure. The Helper had a representation of the completed structure which they had to communicate to the Worker who physically assembled the blocks based on the Helper's guidance. This task was selected because of the high degree of communication and cooperation required between Worker and Helper to complete the task successfully. This type of task requires detailed collaboration for block identification, orientation alignment, and location placement. The mobile features investigated were Audio, Video with Markup, Video with Audio, and Video with Markup and Audio.

4.1 Participants

Volunteers for this study included 32 participants (17 men and 15 women) ranging in age from 23-30 (M=25) years. The participants teamed up in pairs of two, consisting of a Worker and a Helper, collaborating using various modalities to complete the building task. All participants had normal hearing and normal or corrected-to-normal vision

4.2 Experiment Design

A within-subject design, balanced using a Latin-square procedure was employed with the four levels of modality interface (Audio, Video with Markup, Video with Audio, and Video with Markup and Audio). All participants took part in a training session to familiarize themselves with the task and devices. The teams trained by collaboratively communicating with each other to construct one practice model per experimental condition. Teams were given the option for more practice trials; however, none of them felt the need for more. The four experimental conditions and building model configurations were randomized for each team.

4.3 Apparatus

Sixteen building block guides were used in the experiment. Each guide consisted of 46 pieces and had three levels. The model pieces illustrated in the guides were randomly selected from a total of 108 pieces that consisted of eight colors (orange, black, blue, red, yellow, brown, dark green, and lime green) and six sizes (1×2, 1×3, 1×4, 2×2, 2×3, and 2×4 studs). The teams worked cooperatively to identify and place blocks onto a green board that measured 10 inches by 10 inches. Building blocks were located in a pile next to the green board. Worker used a Samsung Galaxy Tablet running our developmental Android application to interact with the Helper through a Wi-Fi connection. The Galaxy Tablet was mounted on a stand above the green board to allow the participant to freely use their hands, as seen in Figure 2.

Fig. 2. Worker's Mobile Device Apparatus

The Helper was situated in front of a workstation, which was isolated from the experimental area. The Helper's workstation allowed them to communicate via voice and/or annotate images (depending on the trial condition) from the Worker's tablet to assist them in their task. The Helper's annotations consisted of free form shapes that were filled with selectable colors, as shown in Figure 3.

4.4 Procedure

The team, consisting of a Worker and a Helper, collaborated using various communication modalities to complete the building task. The modality interfaces

investigated

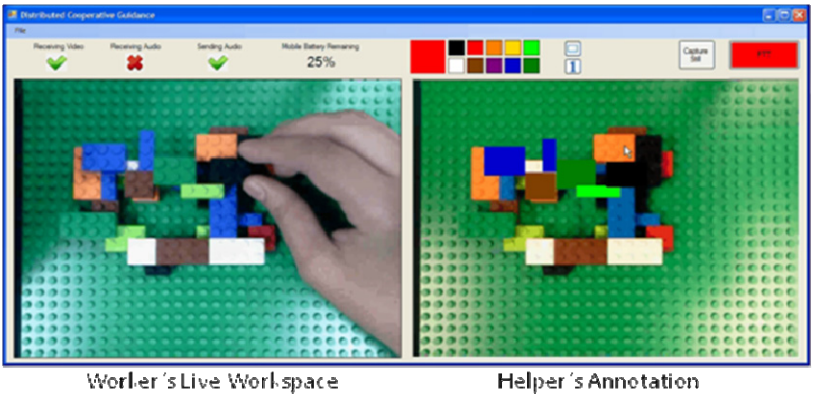

Fig. 3. Helper's Workstation

were Audio, Video with Markup, Video with Audio, and Video with Markup and Audio.

In the Audio condition, the Helper had to verbally describe the color, size, orientation, and placement of the building blocks to the Worker from the active build guide, shown in Figure 4 (a). The Helper's instructional dialogue describing the block and placement was not restricted in any manner, and it was left up to the teams to generate their unique shared common language used in the building process. The Video with Markup condition consisted of the Helper capturing a still picture of the Worker's live perspective from the mobile device's integrated camera. The still image could then be annotated in real-time on the Helper's workstation. The annotation process required the Helper to select the color used in the annotation, followed by clicking and holding the left mouse button down while dragging until the desired shape was created. Upon releasing the left mouse button, the markup annotation was fused with the still image and transmitted to the Worker, as shown in Figure 4 (b). The Helper could undo their annotation by selecting the right mouse button. The undo process could be applied five times to clear past annotations. If five corrections were not sufficient, the Helper could recapture a still image and apply fresh annotations. The Video with Audio condition consisted of the Helper monitoring the Worker's perspective while supplying verbal guidance to describe and place building blocks properly in the model. The Video with Markup and Audio condition combined the Audio and Video conditions so that the Helper and Worker were able to talk to each other as well as send annotated images. In each condition, team members were asked to complete the task as fast as possible without making any errors. Immediately following each condition, both the Helper and the Worker independently completed the NASA-Task Load Index (NASA-TLX, Hart & Staveland, 1988), a validated measure of perceived mental workload.

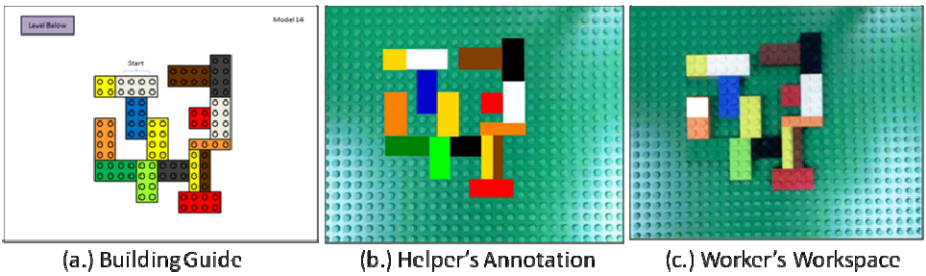

Fig. 4. Reference Guide, Helper's guidance to Worker, and Worker's execution of guidance

5 Results

Accuracy was measured by accurately placing the specific building block in the correct location as determined by the building guide. All teams in all four experimental conditions achieved accuracy of the building task of at least 98.3 %. Thus team performance was measured through completion time. Mean completion times for the four experimental conditions are presented in Figure 5.

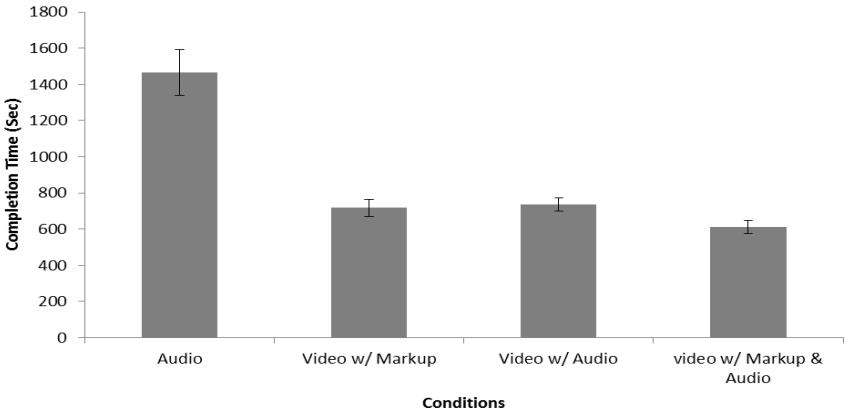

Fig. 5. Mean completion times for each of the four experimental conditions. Error bars are standard errors.

Data from Figure 5 was tested for statistical significance by means of a 4 (condition) within- subjects analysis of variance (ANOVA). A significant main effect was found for completion time across the four experimental conditions, F (3, 42) = 34.2, $p < .01$. Post hoc tests indicated that teams completed the building task significantly faster in the Video with Markup and Audio (M = 625.0 s) condition as compared to Video with Markup (M = 735.1 s) and Video with Audio (M = 739.6 s) which were not significantly different from each other, but were both faster than Audio alone (M = 1490.3 s).

Participants' mean perceived mental workload scores for each experimental condition for the Helper and the Worker are displayed in Figure 6.

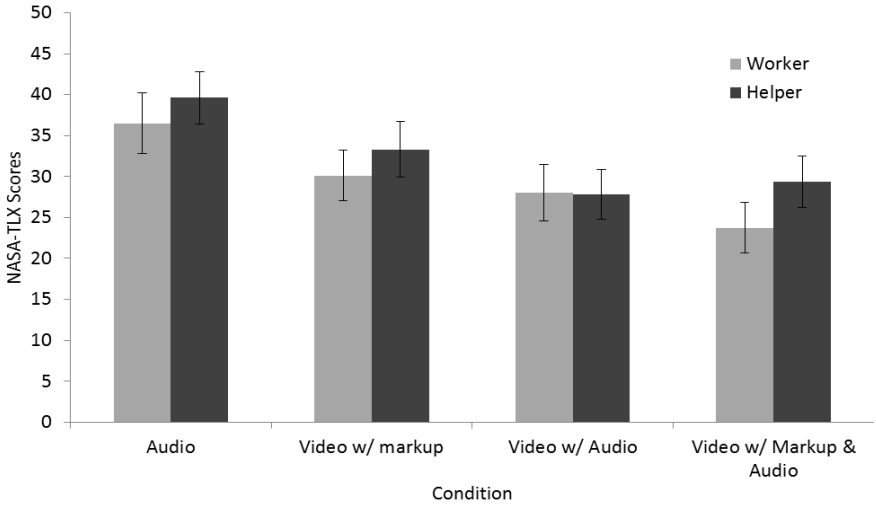

Fig. 6. Mean TLX for each of the four experimental conditions

A 2 (role) x 4 (condition) mixed ANOVA was completed on the NASA – TLX data in Figure 6. A statistically significant main effect was found for conditions, F (3, 177) = 14.39, $p < .01$. Post hoc tests indicated that participants rated the Audio (M = 39.92) as the most mentally demanding condition. Video with Markup and Audio (M = 27.83) and Video with Audio (M = 29.31) were not significantly different then each other but were less demanding then Video with Markup (M = 33.23). No other source of variance was found to be statistically significant, $p > .05$.

6 Discussion

This study evaluated the effectiveness of distributed teams working together to build an abstract structure out of building blocks with the use of a prototype multimodal mobile collaborative tool suite. The developed software allowed distributed teammates to verbally communicate, share video imagery, and send annotated picture messages to foster team collaboration. Our results indicated that the use of multimodal communications on a mobile device improved team performance when collaborating on their task. While all teams successfully completed the task with a high degree of accuracy there were significant differences in the complete times based on the functions available to the team. Teams performed the task quickest in the video with markup and audio condition and slowest in the audio only conditions. Both the Worker and the Helper rated the audio condition as the most mentally demanding condition.

The audio condition serves as a baseline condition to compare the additional features of the mobile prototype tool suite against since a majority of real-time coordination

between distributed teammates is currently accomplished this way. This study replicated many of the earlier studies (Gergle et al., 2004; Fussell et al., 2000; Kirk et al., 2007; Kraut et al., 2002; Kuzuoka, 1992) that showed improvement in task performance when a common ground was established by use of shared perspective as well as the transmission of annotation of images to convey directives (Ou et al., 2003; Stevenson et al., 2008). The integration of voice communication with the ability of the Helper to view the Worker's environment and freely annotate and transmit images was found to be the most effective condition to complete the task quickly and accurately.

This study extended the aforementioned CSCW studies in that the coordination between distributed teammates was done on a mobile device. This is a critical addition to the field of collaborative technologies in that it allows these tools to be more accessible to the general public who normally utilize mobile technologies. This prototype multimodal mobile tool affords users the capability to seek remote guidance outside of one's knowledge base in real-time and uninhibited by location, as long as there is connectivity for the mobile device.

7 Future Work

The ability to share a visual perspective between collaborating partners has been shown to enhance cooperative performance. A limitation to the existing visual dissemination capability is that the Helper only receives visual information on what the Worker is currently focusing the mobile device's camera on and is constrained to the camera's field of view. This "soda straw" perspective can reduce the Helper's overall situation awareness and requires them to rely on the Worker to modify and/or expand awareness through camera movements or panning. Therefore, an extension to the visual capturing feature that would improve the Helper's ability to collaborate could be a virtual immersion in the Worker's scenario. This can be achieve through computer vision techniques that stitch a series of individual snap shots to form a 3D perspective similar to Google's Sphere, as depicted in Figure 7. The new perspective of the Worker's workspace can give the Helper the freedom to pan, zoom, etc. to obtain the necessary vantage view angle to provide better communication and guidance.

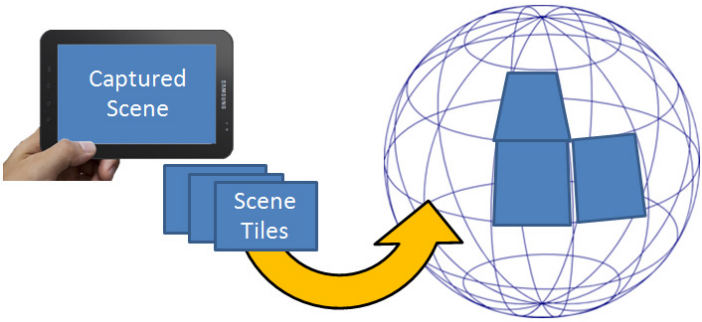

Fig. 7. Immersive 3D scene generated from a series of Worker's still images

References

1. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnick, L.B., Levine, R.M., Teasley, S.D. (eds.) Perspectives on Socially Shared Cognition, pp. 127–149 (1991)
2. Fussel, S.R., Kraut, R.E., Siegel, J.: Coordination of communication: effects of shared visual context on collaborative work. In: Proceedings of CSCW 2000, pp. 21–30. ACM Press, NY (2000)
3. Gergle, D., Kraut, R.E., Fussell, S.R.: Action as language in a shared visual space. In: Proceedings of CSCW 2004, vol. 6(3), pp. 487–496. ACM Press, NY (2004)
4. Gergle, D.: The value of shared visual space for collaborative physical task. In: Proceedings of CHI 2005, pp. 1116–1117 (2005)
5. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North-Holland, Amsterdam (1988)
6. Kirk, D., Rodden, T., Fraser, D.S.: Turn it this way: ground collaborative action with remote gestures. In: Proceedings of Computer Human Interaction (CHI): Distributed Interaction, San Jose, CA (2007)
7. Kraut, R.E., Darren, G., Fussell, S.R.: The use of visual information in shared visual co-presence. In: Proceedings of CSCW 2002, pp. 31–40 (2002)
8. Kuzuoka, H.: Spatial workspace collaboration: a sharedview video support system for remote collaboration capability. In: Proceedings of CHI 1992, pp. 533–540 (1992)
9. Ou, J., Fussell, S.R., Chen, X., Setlock, L.D., Yang, J.: Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In: International Conference on Multimodal Interfaces, pp. 242–249. ACM Press, Vancouver (2003)
10. Stevenson, D., Li, J., Smith, J., Hutchins, M.: A collaborative guidance case study. In: AUIC 2008, Wollongong, NSW, Australia, pp. 33–42 (2008)
11. Wickens, C.D., Vidulich, M., Sandry-Garza, D.: Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. Human Factors 26, 533–543 (1984)