Multi User Multimodal Tabletop Interaction over Existing Single User Applications

Edward Tse[1,2], Saul Greenberg[1], Chia Shen[2]
[1]University of Calgary, [2]Mitsubishi Electric Research Laboratories
[1]2500 University Dr. N.W, Calgary, Alberta, Canada, T2N 1N4
[2]201 Broadway, Cambridge, Massachusetts, USA, 02139
[1](403) 210-9502, [2](617) 621-7500
[tsee, saul]@cpsc.ucalgary.ca, shen@merl.com

ABSTRACT
In this demonstration we illustrate the behavioural foundations of natural speech and gesture interactions on a digital table in practice through the application of these principles to four existing single user applications: Google Earth, Warcraft, Sims and Virtual Surgery. We demonstrate verbal alouds, rich hand gestures, multimodal input, interleaving actions and requests for assistance/validation. By making actions public through speech and gesture interaction, we can improve consequential communication and the group's common ground.

Categories and Subject Descriptors
H5.2 [**Information interfaces and presentation**]: User Interfaces. – Interaction Styles.

General Terms
Human Computer Interaction, Computer Supported Cooperative Work, Design, Human Factors

Keywords
Digital Tabletop Interaction, Multimodal Speech and Gesture Input, Behavioural Foundations

1. INTRODUCTION
Traditional desktop computers are unsatisfying for highly collaborative situations involving multiple co-located people exploring and problem-solving over rich spatial information. These situations include mission critical environments such as military command posts and air traffic control centers, in which paper media such as maps and flight strips are preferred even when digital counterparts are available [Cohen, 2002]. For example, Cohen et. al.'s ethnographic studies illustrate why paper maps on a tabletop were preferred over electronic displays by Brigadier Generals in military command and control situations [Cohen, 2002]. The 'single user' assumptions inherent in the electronic display's input device and its software limited commanders, as they were accustomed to using multiple fingers and two-handed gestures to mark (or pin) points and areas of interest with their fingers and hands, often in concert with speech [Cohen, 2002, McGee, 2001].

CSCW '06, November 4-8, 2006, Banff, Alberta, Canada.

Figure 1. Rich Multi User Digital Table Interaction

This work explores the recognition and use of people's natural explicit actions performed in real life table settings. These explicit actions (e.g., gaze, gesture and speech) are the interactions that make face to face collaborations so effective. Multimodal speech and gesture interaction over digital tables aims to provide the richness of natural interactions with the advantages of digital displays (e.g., real time updates, geospatial information of the entire planet, zooming and panning). Multiuser multimodal makes private actions (with a keyboard and mouse) public (with speech and gesture). This improved awareness of others' publicized actions results in a higher level of common ground between participants, and supports effective collaboration on a digital table.

2. BEHAVIOURAL FOUNDATIONS
Proponents of multimodal interfaces argue that the standard windows/icons/menu/pointing interaction style does not reflect how people work with highly visual interfaces in the everyday world [Cohen, 2002]. They state that the combination of gesture and speech is more efficient and natural. This video summarizes some of the many benefits gesture and speech input provides to individuals and groups. We illustrate the behavioural foundations with examples from our four demonstration systems: Google Earth, Warcraft III, The Sims, and Virtual Surgery.

Paper versus Digital Maps: Digital Maps on a table top provide many of the rich affordances of physical paper maps but also

Tse, E., Greenberg, S., Shen, C. Multi User Multimodal Tabletop Interaction over Existing Single User Applications. Adjunct Proc CSCW 2006, Demonstration and two page summary.

provide the ability to show real time updates, zoom and pan the map, fly to particular location using speech and access rich geospatial information from the Internet (as seen in Google Earth) [Tse, 2006]

Verbal Alouds: Alouds are high level spoken commands that are said for the benefit of the group rather than directed to any one individual person [Heath, 1991]. For example the verbal command "Build farm array" in Warcraft III provides others with awareness information about the actions recently performed.

Rich Hand Gestures: In traditional computing systems and gaming environments all input is assumed to originate from a keyboard, mouse or game controller. Interacting with rich gestural information produces meaningful awareness information to other participants around the table (e.g., using five fingers to pick up a digital television in The Sims or using a fist to operate a mallet in Virtual Surgery).

Speech vs Gesture: Speech is better suited for discrete commands while gesture is better suited for specifying a location on table. For example, in Warcraft III, players can perform a multimodal command using "Build a farm here [point]".

Interleaving Actions: Multiple people can interleave the group decision making process by interleaving their actions. For example, in Figure 1, one person can start a multimodal speech and gesture command using the "create tree [fist]" multimodal command. The other person can add trees by using his fist to stamp more trees, and can complete the command by saying "okay". Similarly, in Figure 2, one person selects a group of units while the other specifies where that unit should move.

Validation and Assistance: Since people are working closely together and monitoring the actions of others, people can recognize when others require assistance even when the other person has not explicitly requested it. For example, in virtual surgery one can observe when someone is reaching for an artifact such as a digital leg and can provide assistance before, during or after the leg has been moved.

Public Actions: Speech and gesture on a digital table makes public the interactions that would otherwise be difficult to observe with a keyboard and a mouse. These public actions produce consequential communication that others can use as cues for validation and assistance [Gutwin, 2004].

Common Ground: Shared understandings of context, environment and situations form the basis of a group's common ground [Clark, 1996]. A fundamental purpose behind all communications is the increase of common ground. This is achieved by obtaining closure on a group's joint actions. For example, in Figure 1, the "[fist] okay" phrase completes the "create tree [fist]" command, it also signifies an understanding of what command was said and consequently increases the group's common ground.

Figure 2. Two people interleaving actions over Warcraft III.

3. CONCLUSION

This video describes behavioural foundations of multimodal speech and gesture interaction on a digital table. If we desire effective collaboration over digital displays we need to support people's natural interactions that occur in the physical world. Multi user multimodal interaction is a first step approach to supporting the natural interactions of multiple people over large digital displays.

Please note: Papers and videos showing more detailed descriptions of the interactions mentioned in this paper can be found at http://grouplab.cpsc.ucalgary.ca/papers.

4. ACKNOWLEDGMENTS

We are grateful to our sponsors: Alberta Ingenuity, iCORE, and NSERC.

5. REFERENCES

[1] Clark, H. *Using language.* Cambridge Univ. Press, 1996.

[2] Cohen, P.R., Coulston, R. and Krout, K., Multimodal interaction during multiparty dialogues: Initial results. *Proc IEEE Int'l Conf. Multimodal Interfaces*, 2002, 448-452.

[3] McGee, D.R. and Cohen, P.R., Creating tangible interfaces by augmenting physical objects with multimodal language. *Proc ACM Conf. Intelligent User Interfaces*, 2001, 113-119.

[4] Heath, C.C. and Luff, P. Collaborative activity and technological design: Task coordination in London Underground control rooms. *Proc ECSCW*, 1991, 65-80

[5] Gutwin, C., and Greenberg, S. The importance of awareness for team cognition in distributed collaboration. In E. Salas, S. Fiore (Eds) *Team Cognition: Understanding the Factors that Drive Process and Performance,* APA Press, 2004, 177-201.

[6] Tse, E., Shen, C., Greenberg, S. and Forlines, C. (2006) Enabling Interaction with Single User Applications through Speech and Gestures on a Multi-User Tabletop. *Proceedings of AVI 2006*. To appear.

[7] Tse, E., Greenberg, S., Shen, C. and Forlines, C. (2006) Multimodal Multiplayer Tabletop Gaming**.** *Proceedings of the Workshop on Pervasive Games 2006*. To appear