Tse, E., Greenberg, S. and Shen, C. (2006)
Motivating Multimodal Interaction Around Digital Tabletops.
Video Proceedings of ACM CSCW'06 Conference on Computer Supported Cooperative Work, November, ACM Press.
Video and two-page summary. Duration 3:25

Motivating Multimodal Interaction Around Digital Tabletops

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]@cs.ucalgary.ca, shen@merl.com

ABSTRACT

In this video we provide motivation for exploring natural speech and gesture interactions on a digital table through the implementation of speech and gesture wrappers around existing single user applications. We briefly compare paper vs digital maps and demonstrate verbal alouds, rich hand gestures, speech for commands, gestures for specifying locations, interleaving actions, validation and assistance.

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 keyboard and mouse desktop computer interaction is 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]**.**

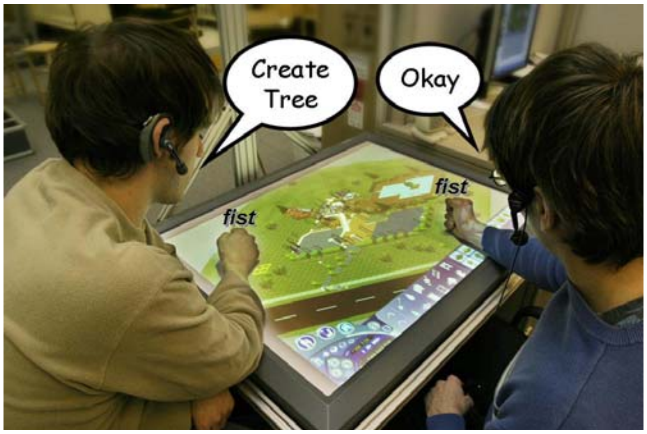

Figure 1. Rich Multi User Digital Table Interaction

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.

Paper versus Digital Maps: Digital Maps on a table top provide many of the rich affordances of physical paper maps but also provide the ability to show real time updates, zoom and pan the map and access rich geospatial information from the Internet [Tse, 2006]

CSCW'06, November 11, 2006, Banff, Alberta, Canada.

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]. Alouds allow people around a table to double check the actions of others to ensure best outcomes.

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 provides a richness normally only found in manipulations of tangible objects such as a gun in an arcade. Rich hand gestures also produces awareness information that is meaningful to other participants.

Speech for Commands, Gesture for Locations: Proponents of multimodal interfaces argue that speech is better suited for issuing commands (e.g., fly to Boston) that would otherwise be difficult to describe in a gesture language whereas gesture is better suited for deictic actions such as pointing to a location on the table [Cohen, 2000]. This means that designers of tabletop systems can leverage the strengths of each modality by designing appropriate interactions for both speech and gestures (e.g., create pool table [point]).

Interleaving Actions: In many of our examples, we show how multiple people can closely turn take multimodal commands. 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.

Interleaving actions distributes the decision making process across all of the co-located participants. This allows participants to double check the actions of others and provides the opportunity for each participant around the table to feel like they are a part of the decision making process.

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. This rich shared common ground supports effective collaborative experiences and outcomes on a digital table [Clark, 96].

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.

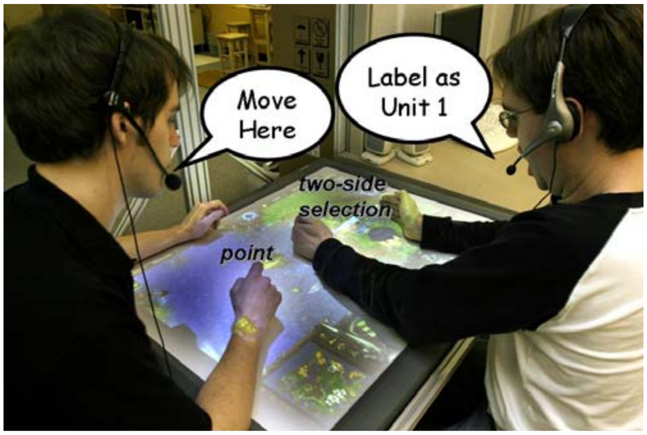

Figure 2. Two people micro turn taking over Warcraft III.

3. CONCLUSION

This video describes motivation for 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.

4. ACKNOWLEDGMENTS

Special thanks to: Sheelagh Carpendale, Kathryn Elliot, Clifton Forlines, Cheng Guo, Gregor McEwan, Stephanie Smale, Nelson Wong, Daniel Wigdor, Jim Young and our sponsors: Alberta Ingenuity, iCORE, and NSERC.

5. REFERENCES

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

[2] Cohen, P. Speech can't do everything: A case for multimodal systems. *Speech Technology Magazine*, 5(4), 2000.

[3] 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.

[4] 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.

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

[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