Exploring Interaction with Multi User Speech and Whole Handed Gestures on a Digital Table

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 T3H 2V1
[2]201 Broadway, Cambridge, Massachusetts, USA, 02139
Tel: [1]403-210-9502, [2]617-621-7500
[tsee, saul]@cs.ucalgary.ca, shen@merl.com

ABSTRACT
While highly collaborative environments involve the use of speech and rich whole handed gestures, current operating systems assume that there will only be one person interacting with the computer at any given time using a keyboard and a mouse over an upright display. In this demonstration we explore interaction with multi user speech and whole handed gestures on a digital table. We illustrate the interaction potential of three different single user computer games: Warcraft III, The Sims and Virtual Surgery. Our examples illustrate how multi user multimodal interaction can mimic some of the affordances traditionally seen when manipulating physical objects (such as a racing wheel in an arcade).

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors, Experimentation

Keywords: Multimodal Interaction, Co-located Interaction and Tabletop Interaction, Whole handed gestures, Verbal Alouds

INTRODUCTION
Tabletop games invite co-located interpersonal play, where players are engaged with both the game and each other. People are tightly coupled in how they monitor the game surface, and each other's actions [3]. There is much talk between players, ranging from exclamations to taunts, instructions, and encouragement. Since people sit around a digital table, they can monitor both the artefacts on the digital display as well as the gestures of others.

Oddly, most home-based computer games do not support this kind of play. Even with console games designed to support multiple players, co-located interaction is limited.

On some games, people take turns at playing game rounds. Other games allow players to interact simultaneously, but they do so by splitting the screen, providing each player with one's own custom view onto the play. People sit facing the screen rather than each other. Thus co-located people tend to be immersed in their individual view into the game at the expense of the social experience.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
UIST '06, October 15–18, 2006, Montreux, Switzerland.
Copyright © 2006 Tse, Greenberg, Shen.


Figure 1. A two handed bracketing gesture in Warcraft III

We believe that a digital table can offer a better social setting for gaming when compared to desktop and console gaming. Of course, this is not a new idea. Some vendors of custom video arcade games (e.g., as installed in video arcades, bars, and other public places) use a tabletop format, typically with controls placed either side by side or opposite one another. Other manufacturers create special purpose digital games that can be placed atop a flat surface.

While multimodal interaction such as Bolt's Put That There system [1] provides natural interaction with large digital surfaces, there exist some benefits that can only be realized in the multi user environment. For example, the use of hand gestures and speech commands over a digital table produces consequential communication, awareness, and cues for validation and assistance.

While previous demonstrations of this system have focused on motivating speech and gesture interaction on a multi user digital tabletop [4, 5, 6], this demonstration focuses on its potential as a novel interaction technology. By demonstrating three unique games we hope to convince the viewer that multi user multimodal tabletop interaction is a rich platform for multi user interaction over many diverse gaming genres and application domains.

Tse, E., Greenberg, S., Shen, C. (2006)
Exploring Interaction with Multi User Speech and Whole Handed Gestures on a Digital Table. Demonstration, Adjunct Proceedings ACM UIST, October.

INTERACTION CAPABILITIES

This section summarizes some of the new interaction capabilities provided by multi user speech and gesture interaction on a digital table. The interaction capabilities are illustrated through three existing computer games: Warcraft III, The Sims, Virtual Surgery.

Rich Hand Gestures: We designed our gestures so that they can be recognized as input and easily understood as a explicit communicative act producing consequential communication to other collaborators. We emphasise that our choice of gestures is not arbitrary. Rather, we examined the rich multimodal interactions reported in ethnographic studies of brigadier generals in real world military command and control situations [2].

For example, observations revealed that multiple controllers would often use two hands to bracket a region of interest. We replicated this gesture in our wrapper by making a two handed bracket command simulate a selection in Warcraft III. Figure 1 shows a Warcraft III player selecting six friendly units within a particular region of the screen using a two-handed selection gesture.

Our gestures are also designed to mimic the actions that naturally occur over physical objects in a workspace. For example, we support the use of five fingers to pick up and place a digital hot tub. In addition, players can use a fist stamping gesture to paste multiple instances of a tree in The Sims.

Meaningful Speech: As with gestures, speech serves as an explicit communicative act (a meaningful 'aloud') that must be informative. Thus a player's speech commands must be constructed so that (a) a player can rapidly issue commands to the game table, and (b) its meaning can be understood by other players within the context of the visual landscape and the player's gestures. In other words, speech is intended not only for the control of the system, but also for the benefits of one's collaborators.

For example, a player can select a scalpel in Virtual Surgery by saying the "scalpel" verbal command. This easily understood phrase makes the normally private act of selecting a tool from a toolbar more easily understood by people around the table.

Combining Speech and Gesture Together: Speech and gestures can interact to provide a rich and expressive language for interaction and collaboration, e.g., through deixis [1]. For example, a person may select a unit, and then say 'Build Barracks' while pointing to the location where it should be built. This intermixing not only makes input simple and efficient, but makes the action sequence easier for others to understand.

These multimodal commands greatly simplify the player's task of understanding the meaning of an overloaded hand posture. A user can easily distinguish different meanings for a single finger using utterances such as 'unit two, move here [point]' or 'next worker, build a farm here [point]' (Figure 2).

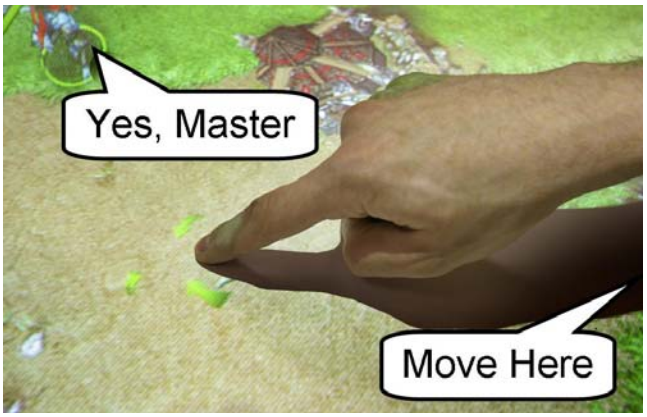


Figure 2. Using speech and gestures on a digital table

Interleaving Actions: Multiple people can split the group decision making process by interleaving their actions above the table. For example in The Sims, one person can start a multimodal speech and gesture command using the "create a piano" multimodal command. The other person can complete this command by specifying where that piano will be placed with their finger on the table. Interleaving actions makes group interaction much more engaging as each person feels like a part of the group decision making process. In the previous example, everyone around the table had taken part of the decision of creating a piano and specifying a place for it in the virtual home.

CONCLUSION

This demonstration explores interaction with multi user speech and whole handed gestures on a digital table. This work represents an early step towards providing rich collaborative experiences in co-located environments.

For more information visit http://grouplab.cpsc.ucalgary.ca

REFERENCES

1. Bolt, R.A., Put-that-there: Voice and gesture at the graphics interface. Proc ACM Conf. Computer Graphics and Interactive Techniques Seattle, 1980, 262-270.
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. 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.
4. Tse, E., Greenberg, S., Shen, C. and Forlines, C. (2006) Multimodal Multiplayer Tabletop Gaming. PerGames '06, May 7th Dublin, Ireland, pp. 139-148.
5. 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. Proc. AVI '06, May 23-26, Venezia, Italy, pp. 336-343.
6. Tse, E., Greenberg, S. and Shen, C. (2006) Motivating Multimodal Interaction around Digital Tabletops. Video Proc. ACM CSCW Conf, Computer Supported Cooperative Work.