MULTIMODAL MULTIPLAYER TABLETOP GAMING

Edward Tse[1,2], Saul Greenberg[2], Chia Shen[1], Clifton Forlines[1]

Abstract

There is a large disparity between the rich physical interfaces of co-located arcade games and the generic input devices seen in most home console systems. In this paper we argue that a digital table is a conducive form factor for general co-located home gaming as it affords: (a) seating in collaboratively relevant positions that give all equal opportunity to reach into the surface and share a common view, (b) rich whole handed gesture input normally only seen when handling physical objects, (c) the ability to monitor how others use space and access objects on the surface, and (d) the ability to communicate to each other and interact atop the surface via gestures and verbal utterances. Our thesis is that multimodal gesture and speech input benefits collaborative interaction over such a digital table. To investigate this thesis, we designed a multimodal, multiplayer gaming environment that allows players to interact directly atop a digital table via speech and rich whole hand gestures. We transform two commercial single player computer games, representing a strategy and simulation game genre, to work within this setting.

1. Introduction

Tables are a pervasive component in many real-world games. Players sit around a table playing board games; even though most require turn-taking, the 'inactive' player remains engaged and often has a role to play (e.g., the 'banker' in Monopoly; the chess player who continually studies the board). In competitive game tables, such as air hockey and foosball, players take sides and play directly against each other – both are highly aware of what the other is doing (or about to do), which affects their individual play strategies. Construction games such as Lego® invite children to collaborate while building structures and objects (here, the floor may serve as a 'table'). The dominant pattern is that 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 [10]. There is much talk between players, ranging from exclamations to taunts to 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. Consider the dominant game products: desktop computer games, and console games played on a television. Desktop computers are largely constructed as a single user system: the size of the screen, the standard single mouse and keyboard, and how people orient computers on a desk impedes how others can join in. Consequently, desktop computer games are typically oriented for a single person playing either alone, or with remotely located players. If other co-located players are present, they normally have to take turns using the game, or work 'over the shoulder' where one person controls the game while others offer advice. Either way, the placement and relatively small size of the monitor usually means that co-located players have to jockey for space [7]. Console games are better at inviting co-located collaboration. Televisions are larger and are usually set up in an area that invites social interaction, meaning that a group of people can easily see the surface. Interaction is not limited to a single input device; indeed four controllers are the standard for most commercial consoles. However, 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,

[1]Mitsubishi Electric Research Laboratories [shen, forlines]@merl.com and
[2]University of Calgary, Alberta, Canada [tsee, saul]@cpsc.ucalgary.ca

Tse, E., Greenberg, S., Shen, C. And Forlines, C. Multimodal Multiplayer Tabletop Gaming. Jointly available as Report 2006-823-16, Department of Computer Science, University of Calgary, Alberta, CANADA *and* MERL Technical Report TR2006-009, Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA, 02139, USA. February, 2006.

providing each player with one's own custom view onto the play. People sit facing the screen rather than each other. Thus the dominant pattern is that co-located people tend to be immersed in their individual view into the game at the expense of the social experience.

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.

The pervasive gaming community has shown a growing interest in bringing physical devices and objects into the gaming environment. For example, Magerkurth [12] tracked tangible pieces placed atop a digital tabletop. Akin to physical devices in arcades, the physical manipulation of game pieces supports rich visceral and gestural affordances (e.g., holding a gun).

Yet to our knowledge, no one has yet analyzed the relevant behavioural foundations behind tabletop gaming and how that can influence game design. Our goal in this paper is to take on this challenge, First, we summarize the behavioural foundations of how people work together over shared visual surfaces. As we will see, good collaboration relies on at least: (a) people sharing a common view, (b) direct input methods that are aware of multiple people, (c) people's ability to monitor how others directly access objects on the surface, and (d) how people communicate to each other and interact atop the surface via gestures and verbal utterances. From these points, we argue that the *digital tabletop* is a conducive form factor for co-located game play as it lets people easily position themselves in a variety of collaborative postures (side by side, kitty-corner, round table, etc.) while giving all equal and simultaneous opportunity to reach into and interact over the surface**.** We also argue that *multimodal gesture and speech input* benefits collaborative tabletop interaction. Second, we apply this knowledge to the design of a multimodal, multiplayer gaming environment that allows people to interact directly atop a digital table via speech and gestures, where we transform single player computer games to work within this setting via our Gesture Speech Infrastructure [18].

2. Behavioural Foundations

The rich body of research on how people interact over horizontal and vertical surfaces agrees that spatial information placed atop a table typically serves as conversational prop to the group. In turn, this creates a common ground that informs and coordinates their joint actions [2]. Rich collaborative interactions over this information often occur as a direct result of *workspace awareness:* the up-to-the-moment understanding one person has of another person's interaction with the shared workspace [10]. This includes awareness of people, how they interact with the workspace, and the events happening within the workspace over time. As summarized below, key behavioural factors that contribute to how collaborators maintain workspace awareness by monitoring others' gestures, speech and gaze. [10].

2.1 Gestures

Gestures as intentional communication. In observational studies of collaborative design involving a tabletop drawing surface, Tang noticed that over one third of all activities consisted of intentional gestures [17]. These intentional gestures serve many communication roles [15], including: pointing to objects and areas of interest within the workspace, drawing of paths and shapes to emphasise content, giving directions, indicating sizes or areas, and acting out operations.

Rich gestures and hand postures. Observations of people working over maps showed that people used different hand postures as well as both hands coupled with speech in very rich ways [4]. These

animated gestures and postures are easily understood as they are often consequences of how one manipulates or refers to the surface and its objects, e.g., grasping, pushing, and pointing postures.

Gestures as consequential communication. Consequential communication happens as one watches the bodies of other's moving around the work surface [16][15]. Many gestures are consequential *vs.* intentional communication. For example, as one person moves her hand in a grasping posture towards an object, others can infer where her hand is heading and what she plans to do. Gestures are also produced as part of many mechanical actions, e.g., grasping, moving, or picking up an object: this also serves to emphasize actions atop the workspace. If accompanied by speech, it also serves to reinforce one's understanding of what that person is doing.

Gestures as simultaneous activity. Given good proximity to the work surface, participants often gesture simultaneously over tables. For example, Tang observed that approximately 50-70% of people's activities around the tabletop involved simultaneous access to the space by more than one person, and that many of these activities were accompanied by a gesture of one type or another**.**

2.2 Speech and alouds.

Talk is fundamental to interpersonal communication. It serves many roles: to inform, to debate, to taunt, to command, to give feedback [2]. Speech also provides awareness through alouds.

Alouds are high level spoken utterances made by the performer of an action meant for the benefit of the group but not directed to any one individual in the group [11]. This 'verbal shadowing' becomes the running commentary that people commonly produce alongside their actions. When working over a table, alouds can help others decide when and where to direct their attention, e.g., by glancing up and looking to see what that person is doing in more detail [10]. For example, a person may say something like "I am moving this car" for a variety of reasons:

- to make others aware of actions that may otherwise be missed,
- to forewarn others about the action they are about to take,
- to serve as an implicit request for assistance,
- to allow others to coordinate their actions with one's own,
- to reveal the course of reasoning,
- to contribute to a history of the decision making process.

2.3 Combination: Gestures and Speech

Deixis: speech refined by gestures. Deictic references are speech terms ('this', 'that', etc.) whose meanings are disambiguated by spatial gestures (e.g., pointing to a location). A typical deictic utterance is "Put that…" (points to item) "there…" (points to location) [1]. Deixis often makes communication more efficient since complex locations and object descriptions can be replaced in speech by a simple gesture. For example, contrast the ease of understanding a person pointing to this sentence while saying 'this sentence here' to the utterance 'the 5th sentence in the paragraph starting with the word deixis located in the middle of page 3'. Furthermore, when speech and gestures are used as multimodal input to a computer, Bolt states [1] and Oviatt confirms [13] that such input provides individuals with a briefer, syntactically simpler and more fluent means of input than speech alone.

Complementary modes. Speech and gestures are strikingly distinct in the information each transmits. For example, studies show that speech is less useful for describing locations and objects that are perceptually accessible to the user, with other modes such as pointing and gesturing being far more appropriate [3,5,13]. Similarly, speech is more useful than gestures for specifying abstract or discrete actions (e.g., Fly to Boston).

Simplicity, efficiency, and errors. Empirical studies of speech/gestures *vs.* speech-only interaction by individuals performing map-based tasks showed that parallel speech/gestural input yields a higher likelihood of correct interpretation than recognition based on a single input mode [14]. This includes more efficient use of speech (23% fewer spoken words), 35% less disfluencies (content self corrections, false starts, verbatim repetitions, spoken pauses, etc.), 36% fewer task performance errors, and 10% faster task performance [14].

Natural interaction. During observations of people using highly visual surfaces such as maps, people were seen to interact with the map very heavily through both speech and gestures. The symbiosis between speech and gestures are verified in the strong user preferences stated by people performing map-based tasks: 95% preferred multimodal interaction *vs.* 5% preferred pen only. No one preferred a speech only interface [13].

2.4 Gaze awareness

People monitor the gaze of a collaborator [11,10]. It lets one know where others are looking and where they are directing their attention. It helps monitor what others are doing. It serves as visual evidence to confirm that others are looking in the right place or are attending one's own acts. It even serves as a deictic reference by having it function as an implicit pointing act [2]. Gaze awareness happens easily and naturally in a co-located tabletop setting, as people are seated in a way where they can see each other's eyes and determine where they are looking on the tabletop.

2.5 Implications

The above points, while oriented toward any co-located interaction, clearly motivates digital multiplayer tabletop gaming using gesture and speech input. Intermixed speech and gestures comprise part of the glue that makes tabletop collaboration effective. Multimodal input is a good way to support individual play over visual game artefacts. Taken together, gestures and speech coupled with gaze awareness support a rich choreography of simultaneous collaborative acts over games. Players' intentional and consequential gestures, gaze movements and verbal alouds indicate intentions, reasoning, and actions. People monitor these acts to help coordinate actions and to regulate their access to the game and its artefacts. Simultaneous activities promote interaction ranging from loosely coupled semi-independent tabletop activities to a tightly coordinated dance of dependant activities. It also explains the weaknesses of existing games. For example, the seating position of console game players and the detachment of one's input from the display means that gestures are not really part of the play, consequential communication is hidden, and gaze awareness is difficult to exploit. Because of split screens, speech acts (deixis, alouds) are decoupled from the artefacts of interest.

In the next section, we apply these behavioural foundations to 'redesign' two existing single player games. As we will see, we create a wrapper around these games that affords multimodal speech and gesture input, and multiplayer capabilities.

3. Warcraft III and The Sims

To illustrate our behavioural foundations in practice, we implemented multiplayer multimodal wrappers atop of the two commercial single player games illustrated in Figure 1: Warcraft III (a command and control strategy game) and The Sims (a simulation game). We chose to use existing games for three reasons. First, they provide a richness and depth of game play that could not be realistically achieved in a research prototype. Second, our focus is on designing rich multimodal interactions; this is where we wanted to concentrate our efforts rather than on a fully functional

Figure 1. Two People Interacting with (left) Warcraft III, (right) The Sims

game system. Finally, we could explore the effects of multimodal input on different game genres simply by wrapping different commercial products. The two games we chose are described below.

Warcraft III, by Blizzard Inc., is a real time strategy game that portrays a command and control scenario over a geospatial landscape. The game visuals include a detailed view of the landscape that can be panned, and a small inset overview of the entire scene. Similar to other strategy games, a person can create *units* comprising semi-autonomous characters, and then direct characters and units to perform a variety of actions, e.g., move, build, attack. Warcraft play is all about a player developing strategies to manage, control and reposition different units over a geospatial area.

The Sims, by Electronic Arts Inc., is a real time domestic simulation game. It implements a virtual home environment where simulated characters (the Sims) live. The game visuals include a landscape presented as an isometric projection of the property and the people who live in it. Players can either control character actions (e.g., shower, play games, sleep) or modify the layout of their virtual homes (e.g., create a table). Game play is about creating a domestic environment nurturing particular lifestyles.

Both games are intended for single user play. By wrapping them in a multimodal, multiuser digital tabletop environment, we repurpose them as games for collaborative play. This is described next.

4. Multiplayer Multimodal Interactions over the Digital Table

For the remainder of this paper, we use these two games as case studies of how the behavioural foundations of Section 2 motivated the design and illustrated the benefits of the rich gestures and multimodal speech input added through our multiplayer wrapper. Tse et. al. [18] provides technical aspects of how we created these multi-player wrappers, while Dietz et. al. [6] describes the DiamondTouch hardware we used to afford a multiplayer touch surface.

4.1 Meaningful Gestures

We added a number of rich hand gestures to player's interactions of both Warcraft III and The Sims. The important point is that a gesture is not only recognized as input, but is easily understood as a communicative act providing explicit and consequential information of one's actions to the other players. We emphasise that our choice of gestures are not arbitrary. Rather, we examined the rich multimodal interactions reported in ethnographic studies of brigadier generals in real world military command and control situations [4].

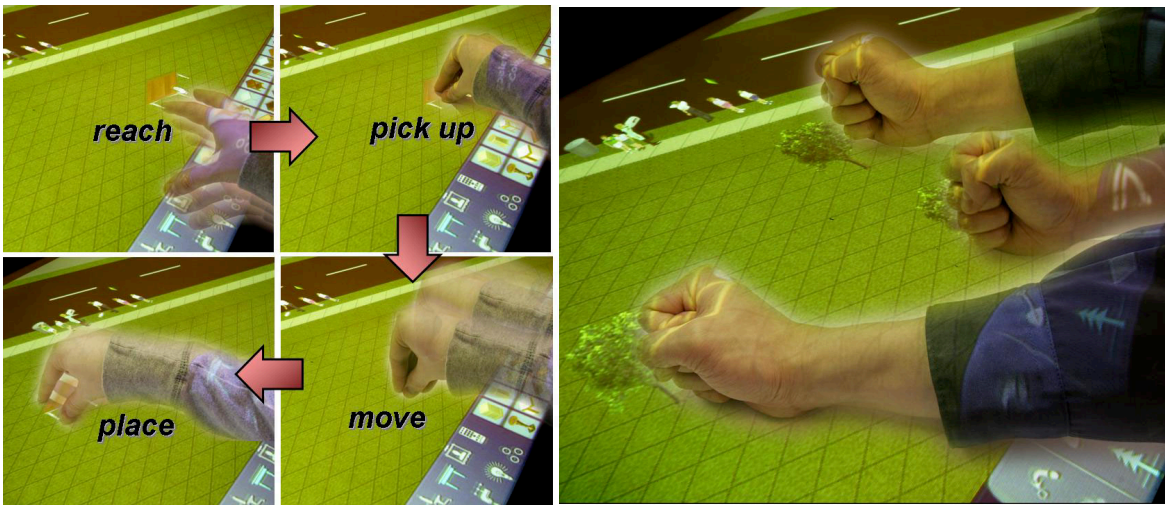

Figure 2. The Sims: five-finger grabbing gesture (left), and fist stamping gesture (right)

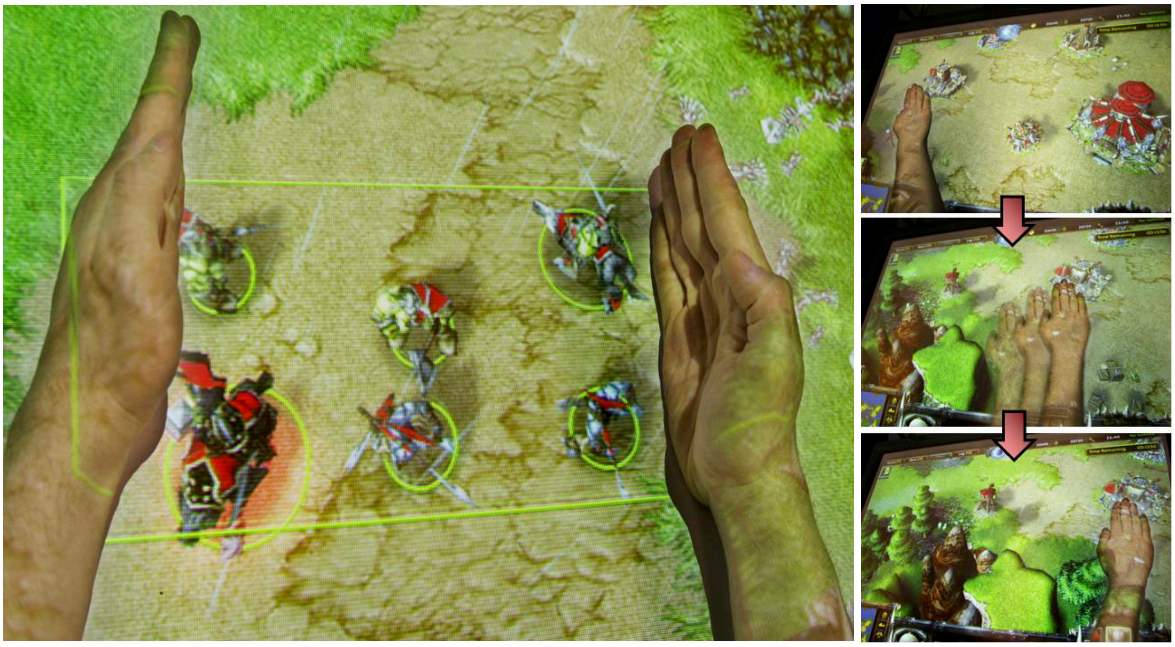

Figure 3. Warcraft III, 2-hand region selection gesture (left), and 1-hand panning gesture (right)

To illustrate, observations revealed that multiple controllers would often use two hands to bracket a region of interest. We replicated this gesture in our tabletop wrapper. Figure 3 (left) and Figure 1 (left) show a Warcraft III player selecting six friendly units within a particular region of the screen using a two-handed selection gesture, while Figure 3 (right) shows a one handed panning gesture similar to how one moves a paper map on a table. Similarly, a sampling of other gestures includes:

- a 5-finger grabbing gesture to reach, pick up, move and place items on a surface (Figure 2, left).
- a fist gesture mimicing the use of a physical stamp to paste object instances on the terrain (Figures 1+2, right).
- pointing for item selection (Figure 1 left, Figure 4) .

Table 1. The Speech and Gesture Interface to Warcraft III and the Sims

Speech Commands in Warcraft III		Speech Commands in The Sims	
Unit <#>	Selects a numbered unit, e.g., one, two	**Rotate**	Rotates the canvas clockwise 90 degrees
Attack / attack here [point]	Selected units attack a pointed to location	**Zoom <In / Out>**	Zooms the canvas to one of three discrete levels
Build <object> here [point]	Build object at current location, e.g., farm, barracks	**<First / Second> Floor**	Moves the current view to a particular floor
Move / move here [point]	Move to the pointed to location	**Return to Neighbourhood**	Allows a saved home to be loaded
[area] Label as unit <#>	Adds a character to a unit group	**Create <object> here [points / fists] okay**	Creates object(s) at the current location, e.g., table, pool, chair
Stop	Stop the current action	**Delete [point]**	Removes an object at the current location
Next worker	Navigate to the next worker	**Walls <Up / Down>**	Shows / Hides walls from current view

4.2. Meaningful Speech

A common approach to wrapping speech atop single user systems is to do a 1:1 mapping of speech onto system-provided command primitives (e.g., saying 'X', the default keyboard shortcut to attack). This is inadequate for a multiplayer setting. If speech is too low level, the other players would have to consciously reconstruct the intention of the player. As with gestures, speech serves as a 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 is easily 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.

To illustrate, our Warcraft III speech vocabulary was constructed using easily understood phrases: nouns such as 'unit one', verbs such as 'move', action phrases such as 'build farm' (Table 1). Internally, these were remapped onto the game's lower level commands. As described in the next section, these speech phrases are usually combined with gestures describing locations and selections to complete the action sequence. While these speech phrases are easily learnt, we added a 2nd display to the side of the table that listed all available speech utterances; this also provided visual feedback of how the system understood the auditory commands by highlighting the best match.

4.3 Combining Gesture and Speech together

The speech and gesture commands of Warcraft and The Sims are often intertwined. For example in Warcraft III, a person may tell a unit to attack, where the object to attack can be specified before, during or even after the speech utterance. As mentioned in Section 2, speech and gestures can interact to provide a rich and expressive language for interaction and collaboration, e.g., through deixis. Figure 1 gives several examples, where deictic speech acts are accompanied by one and two-finger gestures and by fist stamping; all gestures indicate locations not provided by the speech act. Further combinations are illustrated in Table 1. For example, a person may select a unit, and

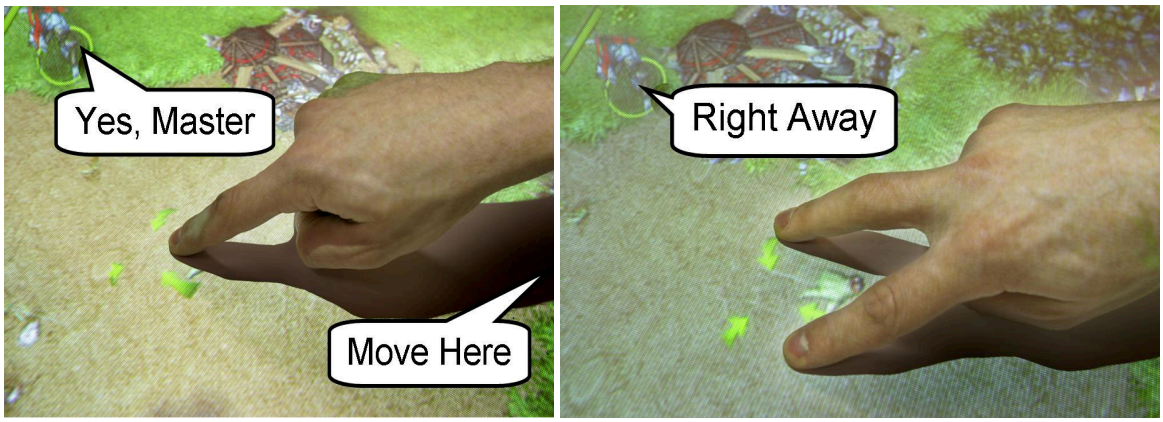

Figure 4. Warcraft III: 1-finger multimodal gesture (left), and 2-finger multimodal gesture (right)

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' and 'next worker, build a farm here' (Figure 4, left).

We should mention that constraints and offerings of the actual commercial single player game significantly influences the appropriate gestures and speech acts that can be added to it via our wrapper. For example, continuous zooming is ideally done by gestural interaction (e.g., a narrowing of a two-handed bounding box). However, since The Sims provides only three discrete levels of zoom it was appropriate to provide a meaningful aloud for zooming. Table 1 shows how we mapped Warcraft III and The Sims onto speech and gestures, while Figure 1 illustrates two people interacting with it on a table.

4.3. Feedback and feedthrough

For all players, game feedback re-enforces what the game understands. While feedback is usually intended for the player who did the action, it becomes feedthrough when others see and understand it. Feedback and feedthrough is done by the visuals (e.g., the arrows surrounding the pointing finger in Figure 4, the bounding box in Figure 3 left, the panning surface in Figure 3 right). As well, each game provides its own auditory feedback to spoken commands: saying 'Unit One Move Here' in Warcraft III results in an in-game character responding with phrases such as 'Yes, Master' or 'Right Away' if the phrase is understood (Figure 4). Similarly, saying 'Create a tree' in The Sims results in a click sound.

4.4. Awareness and Gaze

Because most of these acts work over a spatial location, awareness becomes rich and highly meaningful. By overhearing alouds, by observing players' moving their hands onto the table (consequential communication), by observing players' hand postures and resulting feedback (feedthrough), participants can easily determine the modes, actions and consequences of other people's actions. Gestures and speech are meaningful as they are designed to mimic what is seen and understood in physical environments; this meaning simplifies communication [2]. As a player visually tracks what the other is doing, that other player is aware of where that person is looking and gains a consequential understanding of how that player understands one's own actions.

4.5 Multiplayer Interaction

Finally, our wrapper transforms a single player game into a multi-user one, where players can interact over the surface. Yet this comes at a cost, for single player games are not designed with this in mind. Single player games expect only a single stream of input coming from a single person. In a multiplayer setting, these applications cannot disambiguate what commands come from what person, nor can they make sense of overlapping commands and/or command fragments that arise from simultaneous user activities.

To regulate this, we borrow from ideas in shared window systems. To avoid confusion arising from simultaneous user input across workstations, a *turn taking* wrapper is interposed between the multiple workstation input streams and the single user application [8]. Akin to a switch, this wrapper regulates *user pre-emption* so that only one workstation's input stream is selected and sent to the underlying application. The wrapper could embody various turn taking protocols, e.g., explicit release (a person explicitly gives up the turn), pre-emptive (a new person can grab the turn), pause detection (explicit release when the system detects a pause in the current turn-holder's activity), queue or round-robin (people can 'line up' for their turns), central moderator (a chairperson assigns turns), and free floor (anyone can input at any time, but the group is expected to regulate their turns using social protocol) [9].

In the distributed setting of shared window systems, turn taking is implemented at quite gross levels (e.g., your turn, my turn). Our two case studies reveal far richer opportunities in tabletop multimodal games for social regulation by *micro turn-taking*. That is, speech and gestural tokens can be interleaved so that actions appear to be near-simultaneous. For example, Figure 1 (left) shows micro turn taking in Warcraft III. One person says 'label as unit one' with a two hand side selection, and the other person then immediately directs that unit to move to a new location. Informal observations of people playing together using the multimodal wrappers of Warcraft III and The Sims showed that natural social protocols mitigated most negative effects of micro turn taking over the digital table. Players commented about feeling more engaged and entertained after playing on the tabletop as compared to their experiences playing these games on a desktop computer.

5. Summary and Conclusion

While video gaming has become quite pervasive in our society, there is still a large gulf between the technologies and experiences of arcade gaming versus home console gaming. Console games and computers need to support a variety of applications and games, thus they use generic input devices (e.g., controllers, keyboard and mouse) that can be easily repurposed. Yet generic input devices fail to produce meaningful gestures and gaze awareness for people playing together for two reasons: First, everyone is looking at a common screen rather than each other, thus gaze awareness has the added cost of looking away from the screen. Second, generic input devices lock people's hands and arms in relatively similar hand postures and spatial locations, thus people fail to produce useful awareness information in a collaborative setting.

Conversely, arcade games often use dedicated tangible input devices (e.g., gun, racing wheel, motorcycle, etc) to provide the behavioural and visceral affordances of gestures on real world objects for a single specialized game. Yet specialized tangible input devices (e.g., power glove, steering wheel) are expensive: they only work with a small number of games, and several input devices must be purchased if multiple people are to play together. Even when meaningful gestures can be created with these tangible input devices, people are still looking at a screen rather than each

other; the spatial cues of gestures are lost since they are performed in mid air rather than on the display surface.

This paper contributes multimodal co-located tabletop interaction as a new genre of home console gaming, an interactive platform where multiple people can play together using a digital surface with rich hand gestures normally only seen in arcade games with specialized input devices. Our behavioural foundations show that allowing people to monitor the digital surface, gesture and speech acts of collaborators, produces an engaging and visceral experience for all those involved. Our application of multimodal co-located input to command and control (Warcraft III) and home planning (The Sims) scenarios show that single user games can be easily repurposed for different game genres. Consequently, this work bridge the gulf between arcade gaming and home console gaming by providing new and engaging experiences on a multiplayer multimodal tabletop display. Unlike special purpose arcade game, a single digital table can become a pervasive element in a home setting, allowing co-located players to play different game genres atop of it using their own bodies as input devices.

6. 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] CLARK, H. *Using language.* Cambridge Univ. Press, 1996.

[3] COHEN, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L. and Clow, J., QuickSet: Multimodal interaction for distributed applications. *Proc. ACM Multimedia*, 1997, 31-40.

[4] 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.

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

[6] DIETZ, P.H.; Leigh, D.L., DiamondTouch: A Multi-User Touch Technology, ACM Symposium on User Interface Software and Technology (Proc ACM UIST), 2001, 219-226.

[7] GREENBERG, S. Designing Computers as Public Artifacts. *International Journal of Design Computing: Special Issue on Design Computing on the Net (DCNet'99)*, November 30 - December 3, University of Sydney. 1999.

[8] GREENBERG, S., Sharing views and interactions with single-user applications. *Proc ACM COIS,* 1990, 227-237

[9] GREENBERG, S. Personalizable groupware: Accommodating individual roles and group differences. *Proc ECSCW*, 1991,17-32,

[10] 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.

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

[12] MAGERKURTH, C., Memisoglu, M., Engelke, T. and Streitz, N., *Towards the next generation of tabletop gaming experiences*. Proc. Graphics Interface, 2004, 73-80.

[13] OVIATT, S. L. Ten myths of multimodal interaction, *Comm. ACM*, 42(11), 1999, 74-81.

[14] OVIATT, S. Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction* 12, 1997.

[15] PINELLE, D., Gutwin, C. and Greenberg, S. Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM TOCHI*, 10(4), 2003, 281-311.

[16] SEGAL, L. Effects of checklist interface on non-verbal crew communications, NASA Ames Research Center, Contractor Report 177639. 1994

[17] TANG, J. Findings from observational studies of collaborative work. *Int. J. Man-Machine. Studies.* 34 (2), 1991.

[18] TSE, E., Shen, C., Greenberg, S. and Forlines, C. (2005) Enabling Interaction with Single User Applications through Speech and Gestures on a Multi-User Tabletop. MERL Technical Report TR2005-130, Cambridge, MA.