Improving the Experience of Controlling Avatars in Camera-Based Games Using Physical Input

Na Li, Neema Moraveji, Hiroaki Kimura, Eyal Ofek
Microsoft Research Asia
Beijing, China
nli@zju.edu.cn, neemam@microsoft.com, kimura@ruri.waseda.jp, eyalofek@microsoft.com

ABSTRACT

This paper investigates two methods of improving the user experience of camera-based interaction. First, problems that arise when avatars are designed to mimic a user's physical actions are presented. Second, a solution is proposed: adding a layer of separation between user and avatar while retaining intuitive user control. Two methods are proposed for this separation: spatially and temporally. Implementations of these methods are then presented in the context of a simple game and evaluate their effect on performance and satisfaction. Results of a human subject experiment are presented, showing that reducing the amount of user control can maintain, and even improve, user satisfaction if the design of such a reduction is appropriate. This is followed by a discussion of how the findings inform camera-based game design.

Categories and Subject Descriptors

H.5.2 [Information interfaces and presentation]: User Interfaces

General Terms

Design, Experimentation, Human Factors.

Keywords

Camera-based interaction; avatar; interaction design; user study.

1. INTRODUCTION

A compelling reason to play video games is to immerse oneself into a character's situation. Increasing the visual realism of the virtual world is one way to strengthen this immersion. Another way is to improve control of the virtual characters using novel input devices. Camera-based controls as typified by EyeToy [11] allow bodily input without the restrictions of worn equipment.

Current methods of camera-based input map the player's physical motion to their virtual avatar using two different strategies: physical gestures (as in [1], [7], and EyeToy) and mimicry of human movements by an avatar (herein referred to as "direct-manipulation") as in [5]. The former requires the player to learn a set of physical gestures that initiate avatar actions. In the latter, the avatar mimics the exact motion of the gamer in real-time without requiring any learning on the user's part. Comparing the two approaches, physical gestures are an indirect (herein referred to as "disconnected") means of control. In this sense, direct-manipulation is a 'fully-connected' relationship where the avatar moves consistently with the user's actions.

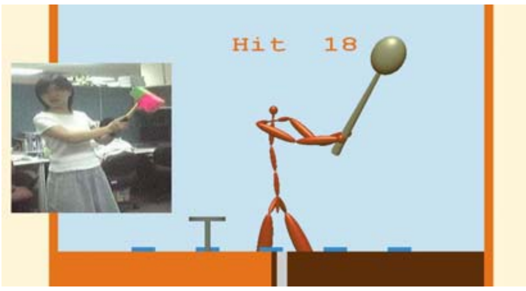

Figure 1. Using a physical hammer and a webcam, a player controls an avatar's hammer to hit nails in a game.

While simplifying control, direct-manipulation introduces many problems. Primarily, our bodies are not as precise as we would like our avatar's to be. That is, providing accurate input to the system using one's limbs can be frustrating. On the other hand, gesture-based control, common on devices like joysticks, introducing complexities by requiring users to memorize gesture commands. Neither gestures nor direct-manipulation always provide an ideal gaming experience. This paper proposes something in between the two: designing disconnected gestures into camera-based interaction such that the gestures are an extension of one's natural physical action and need not be memorized.

Two methods of disconnection are proposed and investigated: temporal and spatial. Examples of these methods are implemented in the context of a simple game and their effect on user performance and satisfaction is evaluated. Results show that providing proper disconnection is an effective way of increasing performance and decreasing frustration, without decreasing satisfaction. Further, the design of the disconnection is an important factor in predicting its utility. The work concludes with a discussion of how this research informs the design of camera-based techniques.

2. RELATED WORK

A great deal of work has been done on observing the effect of user presence on avatar and environment realism. However, little academic work proposes new methods of avatar control in games or tasks to enhance performance and satisfaction while maintaining intuitive physical movements as user input.

Höysniemi, et.al. [6] is representative of papers exploring interaction techniques for vision-based children's games, attempting to identify intuitive gestures for avatar actions.

Linebarger and Kessler [8] showed that for direct object-focused tasks changing avatar representation did not affect task performance while demonstrating that interaction design in these scenarios can have a large effect on performance.

Pressing buttons on a control pad can be seen as a form of gesture, but we are looking specifically at physical gestures that require movement of limbs and camera-based input. Physical gestures used in the camera game Eyetoy:Antigrav [11] were well-received in commercial game design. However, the precise gestures required for control is not an easy task, which motivates another game, EyeToy:Kinetic. The game was described as the ″Most Innovative Design of 2005″ by IGN [12], a large gaming websites. The spirit of the game is that mirroring the player into the game gives her/him full control (direct-manipulation) when accomplishing a task. The problem with this method is that the user has to exert a large amount of effort during the course of the game, which is laborious and limits the story.

We propose a control method between direct-manipulation and physical gestures: the user can move freely before the camera, but the control of the avatar is disconnected from the body movement by extracting meaningful gestures from the natural movement.

3. SYSTEM DESIGN

To investigate techniques of disconnection, we created a framework that allows users to control an avatar in real-time using a single common web-camera. The framework is composed of following components: 1) a motion database indexed by vision features that are extractable with a single camera, 2) robust and real-time feature-tracking by detecting a set of colorized markers with a common camera, and 3) translating the tracked features of human movement into features of desired avatar motion; 4) synthesizing a motion pose from the poses in the motion database indexed by the translated features.

Our implementation of this framework uses a database of continuous 'swing' motions. The user swings a physical hammer to control the hammer swung by the avatar, in a 2D plane, at virtual nails on a table (Figure 1). The physical hammer is painted in bright colors and is the only object detected and tracked by the camera. When the user swings the object outside the camera's field of view, the skeleton fades to grey and the hammer disappears, freezing the avatar until the object re-enters the field of view.

The object-tracking system used in this research is based on computer vision research in our lab. The tracking algorithm is robust and the motion synthesis of the avatar is near real-time. User log videos show virtual objects keeping up with the motions of the user's tracked object.

4. INTERACTION DESIGN

Using direct-manipulation, controlling a virtual avatar with one's body requires precise movements that are difficult and frustrating because the subtle physical feedback of the real world is often removed. One real example of the utility of subtle physical feedback is hitting a physical nail with a physical hammer: it seems simple but requires a high amount of precision, coordination, visual input, and physical feedback. Our idea is to scaffold users by preserving intuitive direct-manipulation as much as possible while reducing the frustrations. We identified several situations where doing so could be useful for designers.

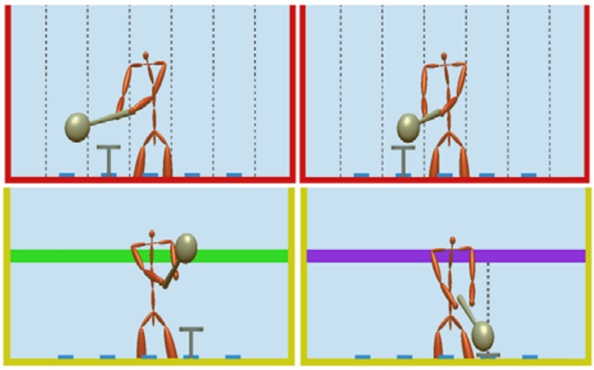

Figure 2a, 2b. Spatial disconnection (top) uses discrete columns for horizontal motion. Spatial-temporal (bottom) shows the trigger line and the path of the hammer.

The first reason to disconnect control is that it's difficult, if not impossible, for designers to make the infinite number of potential virtual interactions between avatar and game targets feel natural. The second is that game designers should be able to vary the required level of input-precision as they see fit (level difficulty). A third situation is when the user is not able to perform the avatar's stylized actions, e.g. a perfectly executed round-house kick, there must be some level of disconnection between the user's motion and the avatar's action. A fourth reason is that requiring precise input while using limbs and physical objects can be frustrating. Depending on the game design, another reason is that targets in a situation may not allow for spatially-continuous movement (e.g. game pieces in chess must be placed on spatially discrete targets).

The above reasons for investigating some level of disconnection led us to identify three basic methods of disconnecting control from users. We explain how we used them to implement two disconnections for evaluation: 'discrete' and 'trigger'.

4.1 Spatial control

This type is used to allow designers to adjust the required precision of input as they see fit. The user continues to move naturally, but the avatar is restricted to certain useful spatial positions. In our implementation, the avatar's horizontal movements are restricted to discrete positions where a nail can appear (Figure 2a). This allows the user to focus less on horizontal placement and more on the vertical swinging. We call this disconnection (spatially) 'discrete'.

4.2 Temporal control

This type allows designers to create stylized actions and super-human talents for avatars. Temporal disconnection suspends the user's control of the avatar for a pre-determined duration then passes it back.

4.3 Spatial-temporal control

In our implementation, we combined temporal with spatial disconnection to create a disconnection called 'trigger'. When the user swings the hammer down from a certain vertical position (the

	Non-discrete	Discrete
Non-trigger	A (L0) Direct manipulation	B (L1) Discrete
Trigger	C (L1) Trigger	D (L2) Discrete with trigger

Table 1. Experiment design with condition labels.

'trigger line') and with a certain minimum velocity, the system takes control of the avatar, swinging the virtual hammer down at the same horizontal position and with a pre-defined velocity (Figure 2b). The trigger line is indicated via a thick line that visually indicates activation by changing color. During this short period of temporal disconnection, the user's physical actions are ignored by the system. This user must strike at the correct horizontal location, but does not need to worry about subtle shifts in angle during the swing itself.

5. EXPERIMENT

The goal of our experiment was to learn about designing disconnection techniques and to observe their effect on performance and satisfaction. We hypothesized that using disconnection would increase performance in the game and not only would satisfaction *not decrease* with disconnection, it would *increase* satisfaction significantly, as the frustration of requiring precise input was alleviated.

We utilized a within-subjects design with four conditions: direct-manipulation, discrete avatar, trigger, and discrete avatar with trigger (Table 1). The order of the games was randomized. Two independent variables (discrete, trigger) manipulated the level of disconnection. The dependent variables were the participant's effectiveness, accuracy, and rankings of the conditions.

We originally thought that the two disconnection methods, 'discrete' and 'trigger', were both 'low' in degree and, when combined together, would result in 'high' disconnection. Therefore, we expected the lowest scores to be for L0 (A), followed by near-equivalent scores for L1 (B, C), and highest for L2 (D) (see Table 1).

5.1 Subjects

24 technical university students (6 female), aged 21-26, participated and were given a reward, with the top 3 scores receiving a bonus. 5 ranked themselves as video game 'experts' and 10 as 'average'. 14 reported previous experience with camera-based input systems. Each participant completed 4 tests, creating 96 trials. All avatar and user actions were logged and videotaped.

5.2 Procedure

We implemented a modification of the classic "whack-a-mole" game: use a physical hammer to control the avatar's hammer to hit as many virtual nails into the table as possible in the time allotted (Figure 1). We chose a simple game to eliminate conditions that might influence performance. The physical hammer weighed 3 pounds and was used 6-8 feet away from a 54-inch plasma television. The virtual nails were placed in discrete horizontal locations on a tabletop, popping up after the previous one is hit down. If the avatar's hammer hits the nail with insufficient velocity, the nail goes down only halfway then pops back up, eliciting a sound. Likewise, if the avatar's striking angle is not sufficiently vertical, the nail 'bends' and returns, eliciting a sound. In this way, the user was given visual and auditory feedback.

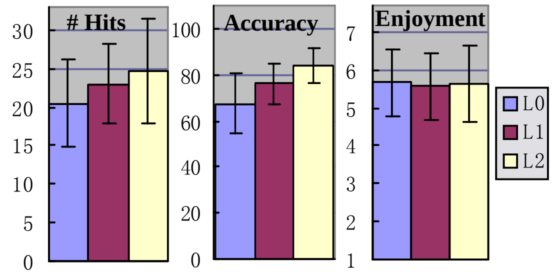

Figure 3. As the number of disconnection methods used increases, accuracy improves significantly while effectiveness follows the same trend. Enjoyment was unaffected.

Before the tests, participants completed a 10-second calibration routine, a 30-second learning stage as well as 2 practice games using direct-manipulation. A brief video introducing the trigger mechanism was shown but no practice was permitted. No description of the discrete disconnection method was given.

After each test, subjects completed a brief questionnaire consisting of Likert scales of 7. After completing all 4 conditions, participants completed another questionnaire and were able to make more verbal comments. Among other questions, we asked users to rate the "enjoyment" for each condition. As "enjoyment" is difficult to quantify and the game is very simple (and not very fun) to begin with, we assert that participants commonly equated "enjoyment" with satisfaction with avatar control.

6. RESULTS

6.1 General

We used a single-factor, two-level, independent-measures analysis of variance (ANOVA) to determine significance. Figure 3 shows average scores for the groups of conditions with error bars indicating standard deviation. An increase in the number of disconnection methods used correlated directly with a significant increase in accuracy ($p=0.012$ for L0 to L1, $p=0.0022$ for L1 to L2). Effectiveness (# Hits) followed a similar trend but was not statistically significant. Even with control of the avatar reduced, satisfaction ratings did not decrease significantly but they also did not increase.

Informally, we observed less visible frustration when participants used the discrete and trigger methods. Use of the trigger method also exhibited less physical exertion, as expected. However, participants were often seen holding the hammer in awkward physical positions while using this method.

6.2 Preference groups

When we inspected how the participants ranked the conditions, two distinctive groups emerged (see Figure 4): Group I, 15 participants, strongly preferred no trigger (ranking A and B above C and D) while Group II strongly preferred the trigger (ranking C and D above A and B).

The difference between Group I and II became evident when we found that game performance and satisfaction ratings differed significantly between them (see Figure 5). In particular, Group II

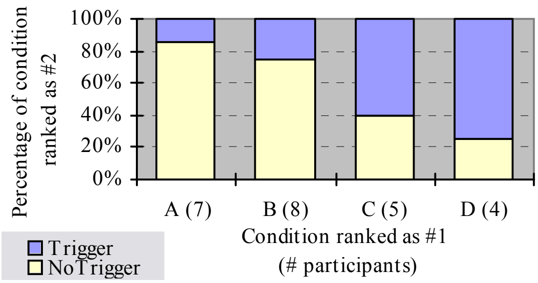

Figure 4. Participants ranking non-trigger conditions #1 ranked them #2 as well. The same was true for trigger.

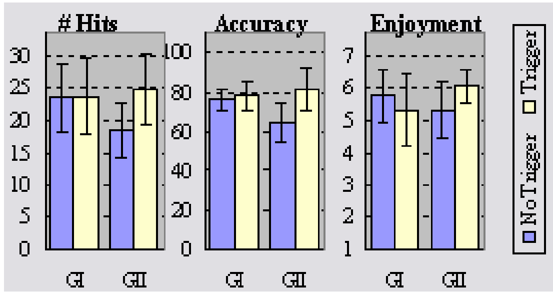

Figure 5. Group II, understanding the trigger and using it effectively, significantly improved in all metrics.

significantly increased their number of hits (18.6 to 24.7) ($p=0.018$, $F(4,49)=6.98$) and accuracy (64.8% to 81.4%) ($p=0.003$, $F(4.49)=12.2$) in 'trigger' conditions. Correspondingly, Group II participants rated the trigger as more satisfactory (mean=6.06; SD=0.58) than 'no trigger' (mean=5.33; SD=0.87) ($p=0.05$, $F(4.49)=4.31$). In general, they rated disconnection higher than Group I rated it ($p=0.0005$, $F(4.30)=16.6$). The user comments that lent the most insight to our results were the 8 complaining that the user interface for the trigger was confusing. Six of these came from Group I.

This led us to conclude that those users who understood and felt comfortable with the trigger preferred it to direct-manipulation. The result, while encouraging, implies that if the interaction and visual design of the system is clear enough, users will prefer it to a direct-manipulation (traditional) camera-based game.

7. DISCUSSION AND CONCLUSION

Overall, participants were not adversely affected by removing some control from avatars, the design of the disconnection techniques impacted performance heavily. All participants understood discrete control but many complained about the trigger. There were three reasons for this: 1) the visual feedback of the discrete method, 2) controlling the trigger was not always straightforward, and 3) the trigger forced some users to modify their natural behavior (inaccurately supporting their intention). One reason the discrete method is less liable to cause frustration was that it is easier to rectify. Further, the visual design of the 'trigger line' (explained in 4.3) seemed to confuse some participants. However, those participants who mastered the trigger derived great benefit, significantly increasing score and satisfaction. This is promising, as it shows that reducing the amount of control freedom does not necessarily lead to dissatisfaction. Indeed, *supporting the user's intention* may be more important than increasing the level of control. The balance between intuitive, natural control and artificially supporting user intention is a 'sweet spot' for game designers.

7.1 Implications for design

First-person shooters, arcade games, and role-playing games are good candidates for direct-manipulation. The present research informs design by addressing a common task: controlling an avatar with a handheld item to hit a virtual target. It showed that, by introducing interaction changes, one can maintain and even increase satisfaction without resorting to physical gestures. Hence, though the user may expect direct-manipulation of the avatar some form of disconnection when high precision is required.

7.2 Validity of the testing system

The hammer-and-nail game was chosen because physically hitting nails with a hammer is not as easy as one might assume. One must move the hammer rather close to the nail at first, and it's easy to miss the target. Though we see the nail, and our swinging motion seems straight, our arm is not as precise as we want it to be. That's why we introduced 'disconnection' into the interaction, thereby reducing the frustration that comes with minute mistakes. This is an interaction design problem that has not been greatly explored.

7.3 Future investigation

We plan to study several disconnection techniques in the context of more mature and immersive games to create a list of heuristics. As control of avatars using common web-cameras has only recently been feasible, such a list could be very useful to the field.

8. REFERENCES

1. Barrientos, F. Continuous Control of Avatar Gesture. *Proc. MULTIMEDIA '00*, ACM Press (2000).
2. Boulic, R., et al. Integration of Motion Control Techniques for Virtual Human and Avatar Real-Time Animation. *Proc. VRST '97*, ACM Press (1997).
3. Crawford, C. The Art of Computer Game Design. Berkeley, CA. Osborne/McGraw-Hill (1984).
4. D'Hooge, H. Game Design Principles for the Intel Play Me2Cam Virtual Game System, Intel Tech. Q4, 2001.
5. Hämäläinen, P., et al. A. Martial Arts in Artificial Reality. *Proc. CHI 2005*, ACM Press (2005).
6. Höysniemi, J., et al, Children's Intuitive Gestures in Vision-Based Action. *Comm. of the ACM*, Jan 2005.
7. Lee, C., et al. The Control of Avatar Motion Using Hand Gesture. *Proc. VRST '98*, ACM Press (1998).
8. Linebarger, J. and Kessler, G. The Effect of Avatar Connectedness on Task Performance, Lehigh Univ TR, 2002.
9. Perlin, K. and Goldberg, A. Improv: A System for Scripting Interactive Actors in Virtual Worlds. *Proc. SIGGRAPH '96*, ACM Press (1996).
10. Yonemoto, S., Nakano, H. and Taniguchi, R. Avatar motion control by user body postures. *Proc. MULTIMEDIA '03*, ACM Press (2003).
11. Sony Inc. EyeToy. www.eyetoy.com
12. IGN. http://bestof.ign.com/2005/ps2/20.html