An Evaluation of Game Controllers and Tablets as Controllers for Interactive TV Applications

Dale Cox, Justin Wolford, Carlos Jensen, Dedrie Beardsley
Oregon State University
Department of Computer Science
Corvallis, OR 97331
{cox, wolfordj, cjensen, beardsld}@eecs.oregonstate.edu

ABSTRACT
There is a growing interest in bringing online and streaming content to the television. Gaming platforms such as the PS3, Xbox 360 and Wii are at the center of this digital convergence; platforms for accessing new media services. This presents a number of interface challenges, as controllers designed for gaming have to be adapted to accessing online content. This paper presents a user study examining the limitations and affordances of novel game controllers in an interactive TV (iTV) context and compares them to "second display" approaches using tablets. We look at task completion times, accuracy and user satisfaction across a number of tasks and find that the Wiimote is most liked and performed best in almost all tasks. Participants found the Kinect difficult to use, which led to slow performance and high error rates. We discuss challenges and opportunities for the future convergence of game consoles and iTV.

Categories and Subject Descriptors
H.5.2Information Interfaces and Presentation (e.g., HCI): User Interfaces – Graphical User Interfaces, Input Devices and Strategies, Interaction Styles.

General Terms
Performance, Reliability, Human Factors.

Keywords
Digital convergence, interactive television, iTV, game controller, text entry, navigation, natural user interfaces.

1. INTRODUCTION
Interactive television promises viewers more flexibility and control over their viewing experience, while enriching it with Internet accessible content. By giving viewers a communication channel back to service providers, viewers can access new services and shape their viewing experience in unprecedented ways. As a result, services have emerged like video on demand (Netflix, Hulu, etc.), internet content accessibility on TV screens (YouTube & Flickr channels on Apple TV, etc.) and social networking (see Boxee.tv) allowing for content ratings, or contextual searching.

For trivial tasks like navigating a simple movie rental UI, or controlling streaming video content, a traditional remote control is often sufficient. However services have become more complex, requiring finer levels of control or interaction – such as navigating web page or GUI, drag and drop tasks, or extensive text entry for search or socialization. For this more sophisticated input devices may be required. Some of these tasks overlap with those integral to the modern gaming experience. Video game consoles, with dedicated game controllers, due to their pervasiveness, connectivity and processing power are often at the center of this digital convergence of TV and Internet content. It is therefore important to examine how suitable current systems are for bridging this gap.

The last generation of game consoles have each introduced some device capable of spatial gestures allowing the possibility for a natural user interface (NUI), and which could be especially helpful for navigating complex UIs. Microsoft released the Kinect for the Xbox 360 in 2010, a camera based system that tracks players' movement to allow for complex and natural interactions without holding any kind of controller. Nintendo released the Wii in 2006, which introduced the Wii Remote (nicknamed Wiimote), a wireless controller that tracks spatial movement through accelerometers and infrared sensors. Sony has a similar system for the Playstation 3.

These game consoles and their controllers have sparked the development of tools and solutions beyond those in traditional gaming. The motion-driven Nintendo Wiimote was the first to attract the attention of the hacker community outside the console market. The Bluetooth interface of the Nintendo Wiimote made it a simple and accessible device to "hack" and adopt for various uses. Soon after the Kinect was released, the open source community reverse engineered the device and released a driver package allowing others to develop systems that took advantage of its capabilities. Its USB interface makes it ideal for use with a PC. Today, several different open source SDKs exist, as well as an official Microsoft Kinect SDK for Windows.

In addition, tablets continue to evolve and increase in popularity. Some proposed game controllers are exploring the use of such touch interfaces (most notably the upcoming Wii U). Tablets are also being considered as companions to both gaming devices and iTV services (often referred to as second screen navigation). According to a recent Pew study [13], tablet ownership nearly doubled over just the 2011 holiday season. There are a number of applications currently available that allow a tablet to serve as an input device for another computer such as IntoNow (http://www.intonow.com) that detects which show or movie you're watching and provides additional media content and social networking capabilities. These applications allow the tablet to function like a touch screen display or a track pad found on most laptops and would serve as a suitable baseline.

This paper explores the challenges and opportunities of using game-related control technologies to control interactive television applications through the study of a hypothetical, but representative set of navigation, selection, and control tasks for an iTV application. We examine learnability and ease of use, as well as accuracy and error rates.

The rest of this paper is laid out as follows. First we discuss related work looking at the control of interactive television applications as well as gaming systems. We then discuss our user experiment and the design considerations we took into account. Finally, we present our findings, and a discussion of future work.

2. RELATED WORK

Much innovation has taken place in the design of new interaction techniques and devices for gaming devices. The research community is still catching up with the necessary evaluation of the potential, effectiveness and usability of these novel game input devices, both within their targeted use domain as well as in other environments like PC or interactive television.

Looking to the modern interactive TV interface, we see that it combines many types of user interaction. The areas we chose to focus on cover the core functionality of pointing, navigation, and text entry. Pointing is perhaps the more novel and difficult task with todays' hardware, but is a prerequisite for many of the more sophisticated types of applications and use cases. Individually, each of these topics has an established body of research, but in the context of a media center or iTV there is very little. We feel our work contributes in this area in a practical way by combining choosing an itv-like interface and a living room environment. We also present the recent adoption rate history of streaming services, which are at the center of modern interactive television systems.

2.1 Growing Popularity of Streaming Services

With the spread of broadband internet access throughout North America, high bandwidth services like high definition on-demand streaming video, previously limited in either quality or duration, have become commonplace. Between 2001 and 2009, broadband Internet use increased seven-fold, covering from 9% to 64% of American households [17]. A 2011 Nielsen study found that from 2008 to 2011 there was a 22% increase in the number of users watching video on the Internet, and an 80% increase in the average viewing time [15].

One of the key players in the Internet-based video-on-demand area has been Netflix. Netflix debuted as a DVD-by-mail service in 1997, and has since introduced and popularized a broadly available Internet streaming service. By the end of 2011, Netflix had over 21 million paying streaming subscribers [14]. In a Fall 2010 report by Sandvine, Netflix was shown to account for 20.6% of all downstream prime-time Internet traffic in North America [19]. Just 7 months later, Netflix users were consuming 29.7% of all downstream prime-time Internet traffic in North America [19].

Other online video services such as Hulu, Amazon Instant Video and YouTube (though the latter still mostly offers shorter clips, it has branched into feature content delivery as well) have also grown in popularity. Internationally, over 4 billion videos are viewed on YouTube each day [22]. Hulu just passed 1.5 million paying subscribers of its paid Plus service [7].

In part this success is driven by the growth of systems that help these users bridge the gap between the computer and the TV experience. This includes a plethora of streaming devices like the Roku and Apple TV, a new generation of connected TV's and DVD/Blue-ray players, and last but not least game consoles. Each of the three leading game consoles have added mechanisms for viewing streaming Netflix content on their devices. Services like Netflix and Hulu that began as a PC experiences, can now be accessed from a number of different devices and platforms. This has made enabled these services to go from a niche technophile market to appealing to the average consumer. In a 2011 Nielsen study, 50% of all Netflix users were found to watch Netflix content through a gaming console [16]. In the same study, Nielsen found 162 million Americans own a game console. This means that these platforms are natural ways to deliver these experiences.

The need to manage users' media viewing experience has led to the development of media center applications like the Xbox Media Center (XBMC) and Windows Media Center. Internet-enabled set-top devices like Boxee and AppleTV have also appeared allowing easy streaming video viewing from a normal TV. These allow users to consolidate their media consumption, as well as manage their local library. Due to the interactive and highly customizable experience allowed by these services, the need for robust input methods will continue to gain importance.

2.2 Pointing and Navigation Using Novel UI Devices

Over the last few years there has been a growing trend to develop and evaluate what are being referred to as Natural User Interfaces (NUI's). These interfaces extend the basic direct manipulation paradigm by allowing users to interact with the computer with motions more closely resembling those we'd use in real life. Among the leading platforms for such interfaces we find game consoles. These techniques could help bridge the complexity gap between the new interactive TV applications and the interactions afforded by conventional remote controls. Because of space limitations we will only review some of the most directly applicable research to our study.

Starting with camera and motion based techniques, Cheng and Takatsuka [12] introduced dTouch, a finger pointing technique for large displays that uses an off-the-shelf webcam. Using the concept of a "virtual touchscreen", dTouch enables users to manipulate onscreen objects in an absolute coordinate system. They performed a user study comparing dTouch to a method using the EyeToy camera, used on the PlayStation console. Results indicated the two methods were comparable with users preferring dTouch.

Lee [10] described a cursor technique using the Wiimote that enabled finger-tracking through the use of reflective tags taped to the fingers of users. Rather than holding the Wiimote in the hand, they used the IR camera built into the Wiimote with an IR LED array to allow almost bare-hand operation. Lin et al. [11] demonstrated a technique similar to Lee's, but using a second Wiimote for additional functionality.

Using more traditional controllers, Natapov et al. [4] performed a comparative study evaluating the Wiimote and traditional gamepad for pointing and selecting tasks. Although the error rate was higher, 14 out of 15 participants said they preferred the Wiimote in a home entertainment environment. They found that the Wiimote had a 75% performance increase over the traditional gamepad when comparing speed and accuracy.

Finally, turning to smart phones and tablets, McCallum et al. [5] developed a hybrid system called ARC-Pad, which combined absolute and relative positioning techniques for use with large displays. A smart phone screen was used like a touchpad. ARC-Pad was compared against a traditional touchpad style interface, which employed cursor acceleration. ARC-Pad performed slightly better (166ms faster) than the relative in completion time. The results suggested as pixel distance increased beyond what was studied, ARC-Pad performance would change minimally while the relative touchpad would continue to worsen.

2.3 Text Entry With Keyboard Alternatives

Over the last two decades, the need for text entry without a traditional mechanical keyboard has increased. With the introduction of PDAs and smart phones, text entry presents a challenge due to a limited input area. Most interactive TV systems attempt to minimize the necessity of text entry through the use of various widgets and interface choices. Though the need may be reduced, it is difficult to completely do away with text entry for applications such as search or social media.

This has led TV manufacturers like Samsung to market 2-sided remote controls; one side having normal remote control functions and the reverse a full keyboard, or Sony to merge a PlayStation controller and a full keyboard in their Google TV products. While such solutions may provide speed advantages, they lead to cumbersome and intimidating user experiences. We examined alternatives to keyboard text entry, focusing on touchscreens, game controllers and freehand gesture techniques.

The Graffiti pen-based gesture alphabet made popular by Palm in the late 90s, allowed users to quickly input text using a proprietary alphabet. MacKenzie and Zhang [9] analyzed the learnability and accuracy of Graffiti. Participants were given practice time using a reference chart showing the gesture alphabet. After practice they repeated the entire alphabet 5 times without having a reference available and again 1 week later. The results showed a nearly 97% character accuracy rate after 5 minutes of practice.

Ni, et al. [21] adapted the Graffiti alphabet to a freehand gesture-based text entry system called AirStroke. They compared two AirStroke implementations, one with word completion and one without. Airstroke with word completion averaged 11 wpm while no word completion was at 6.5 wpm. The error rate with word completion averaged 6.6% compared to 11.8% without. Some participants reported arm fatigue, which lessened as their proficiency increased.

Several techniques have been developed enabling text input using a traditional gamepad. Költringer et al. [20] designed and evaluated TwoStick, a novel text entry system using both analog joysticks on an Xbox 360 controller. TwoStick was compared to a traditional selection keyboard. Initially, users typed slower and had a higher error rate using TwoStick, but after 15 sessions TwoStick averaged 14.87 wpm while the selection keyboard had a mean of 12.9 wpm. Wilson and Agrawala [1] also created a dual joystick QWERTY method, which showed modest improvement upon the traditional single stick selection keyboard.

Shoemaker et al. [6] compared 3 techniques for mid-air text input. A circle keyboard, QWERTY keyboard and cube keyboard all used a Wiimote as an input method. The QWERTY method performed best in accuracy and performance; this method is similar to our Wiimote text entry task. A questionnaire taken after the study revealed users preferred the QWERTY method overall.

Castellucci and MacKenzie [18] presented an alternative to an on-screen keyboard using the Wiimote called UniGest. UniGest is a technique that takes advantage of the motion-sensing capabilities of the controller to capture movement and rotation. A gesture alphabet is proposed which maps the gestures to character input. Their results predict an upper-bound of 27.9 wpm using the UniGest technique.

3. METHODOLOGY

This section describes a user study designed to measure the effectiveness of video game and tablet input methods in a iTV context. We used 4 input methods: the Microsoft Kinect, Nintendo Wiimote and 2 methods using an Android tablet; a condition where subjects had to scroll (relative coordinate condition), and one using an absolute coordinate space (mirror condition). The idea was that in the relative condition subjects would have to scroll around like when using a mouse pad, and in the absolute coordinate condition, the whole TV image would be represented on tablet at once. Participants completed pre and post-experiment questionnaires and also a post-experiment interview. All sessions were recorded using a video camera and screen capturing software.

Participants were recruited in pairs from a college campus and surrounding community. Examining pairs offered a realistic social context in the living room. There were 33 male and 29 female participants with ages ranging from 18 to 57 years. The mean age was 24.5 and all but 4 were right-handed. Demographic data and media viewing frequency were gathered prior to the study. Two devices were randomly assigned to each pair.

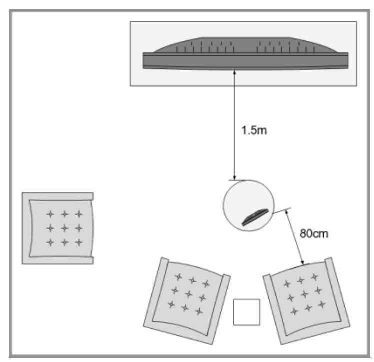

Figure 1: Room configuration used during the experiment.

The system ran on a PC hooked up to a 55" HDTV, in a mock living room environment (see room layout in Figure 1). The subjects sat in the two center seats, while the experimenter sat off to the side. The center table was only purposed as a platform for the Kinect sensor. A small table (15x15cm surface area) placed between the two chairs, was large enough to hold a drink or a plate, but not both at the same time.

A Nintendo wireless sensor bar was used in combination with a standard Nintendo Wii Remote for relevant conditions. A Microsoft Kinect sensor was used for the Kinect tasks, and a 10.1" tablet running Android 2.3 was used for the tablet tasks. A windows application called GlovePIE was used to control the cursor using the Wiimote. A GlovePIE script enabled the IR camera in the Wiimote to control the mouse cursor and the 'A' button to control the left mouse click.

A custom application was created to allow the Kinect to control the mouse cursor and left button. The application was written in C# using the OpenNI framework. To move the cursor, participants moved their right hand, which positioned the cursor similar to a traditional mouse. To initiate a drag, participants moved their left hand forward to cross the threshold of a virtual plane 30-40cm in front of them. This action is equivalent to a left mouse down event. To initiate a drop, participants would simply pull their arm back and break the plane in the opposite direction. This action is equivalent to a left mouse up event. To initiate a left click, participants move their left hand quickly through the plane and back out in one fluid motion.

The software used in the relative tablet condition was an open source Android application called RemoteDroid. This application turns the entire tablet into one large touchpad similar to what is found on most laptop computers (see Figure 2). This is application was paired with an application that runs on the host computer.

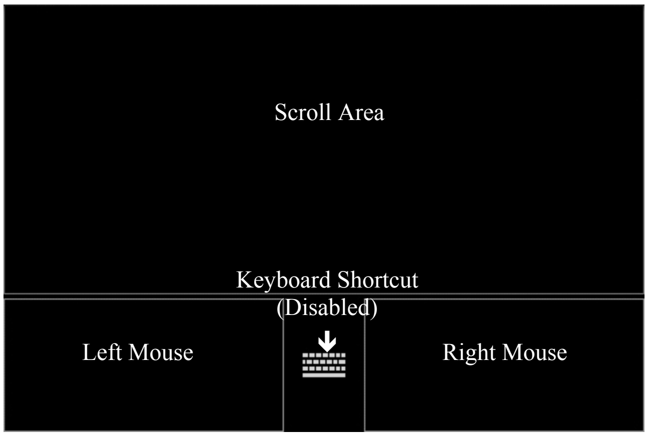

Figure 2: The configuration of the relative tablet app.

The mirrored condition used a modified version of the RemoteDroid application. It continually updates the tablet display with a screenshot of the TV. Instead of using a relative coordinate system where the cursor movement corresponds to relative changes in cursor position, an absolute coordinate system is used. By using an absolute coordinate system, a user can click on any location of the mirrored display and have it mapped to the equivalent location on the PC display.

Subjects were trained on the devices they were going to use and given time to practice on a screen that allowed dragging and dropping an object and clicking a button. When they felt comfortable with the device they began the drag-and-drop task.

The only actions allowed in the drag and drop task were dragging and dropping a widget into a target box (see Figure 3). If the widget was dropped fully within the boundary of a target presented at a random point on the screen, then a hit was recorded, and the subject would be presented with a new, slightly smaller target. A miss was recorded if the user missed the widget when attempting to select it, or if they released the widget outside of the target box. They were able to keep trying until they ran out of time for the trial. If the user was unable to place the widget in the target within 16 seconds, the box and widget were moved to random locations on the screen and the target box got bigger. If the user hit the target, then both the widget and target were randomly moved and the target shrank, with the minimum size for the target being 3 pixels wider and taller than the widget.

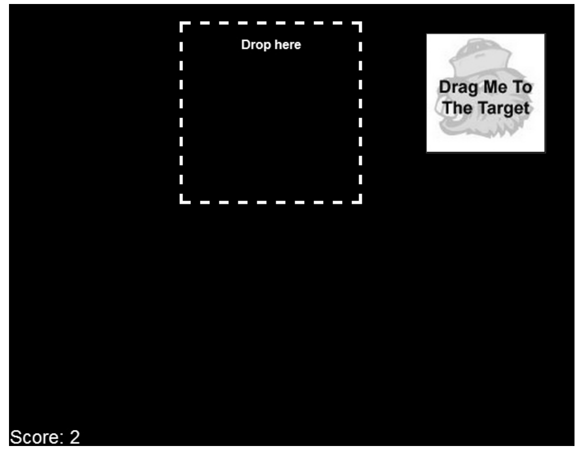

Figure 3: Sample screen for drag and drop task.

After completing the drag-and-drop accuracy task, we asked subjects to complete a number of navigation tasks on an iTV environment, simulated using the popular XBMC media center.

The navigation tasks included 3 different activities in XBMC. The first required the subject to navigate from the main menu to the weather settings screen and change the city name. This may have been the most challenging task due to the text entry requirement. Next, subjects would go to the weather screen and change the city currently displayed. This was difficult at times because it required clicking on a very small button. After changing the city, the subject would navigate to the movie selection screen and select a movie using a scrollbar. Finally, after the movie started, a slider was used to adjust movie volume. In all navigation tasks, an error was recorded if a subject clicked on a non-interactive item or if they clicked on the wrong UI widget.

We measured users' time to complete tasks and their error rates. Additionally we collected qualitative data in a post experiment interview and survey. Finally we used screen capture software and a video camera to record subjects. Subjects performed the experiment in pairs, taking turns with each device (subject A would try device 1, then subject B would use the same device. Next Subject A would try device 2, and then Subject B would do the same). Pairs were randomly assigned two input methods.

Because of previous research showing the importance of studying the effectiveness of UI techniques under similar manual loads [2], and the oft-informal nature of TV viewing, we decided to give each subject a slice of pizza and a drink to hold and consume during the course of the experimental tasks. Subjects were not allowed to place the food items on the floor or on the larger central table, but had to balance them on their seat or lap.

After both subjects completed all tasks using the first device assigned to them, they were introduced to the second device, and the process started anew, from the training period onward. After both subjects completed both conditions, they were asked to complete a short survey asking them about learnability, ease of use and practicality of the devices they had been assigned. All questions were on a 5-point Likert scale, 1 meaning strongly agree and 5 meaning strongly disagree.

Finally, they were interviewed to get a deeper understanding of their experience. We were interested in their satisfaction with the various input devices. This included the ease with which the subjects could use the device along with their enjoyment of using the device. Additionally we asked about their comfort level using the device in a social context where others were observing them.

4. RESULTS

4.1 Drag and Drop Task

The main task we used to measure the efficiency of a UI technique for manipulation was the timed drag and drop task, as it combined selection, movement, as well as accuracy. The more drops a subject managed within the time allotted, the more accurate their manipulation of the widgets on the screen. The highest mean number of targets hit was with the mirror tablet, where subjects hit an average of 14.09 targets (see Figure 4). Subjects using the Wiimote and relative tablet scored 12.97 and 11.90 hits respectively. Those using the Kinect averaged a score of 7.37 hits. The Kinect did significantly worse than all other devices (One-way ANOVA $F(3,19)=54.5$, $P<0.001$ with Tukey's HSD for Post Hoc analysis). The relative tablet also did significantly worse than the Wiimote and mirror tablet ($P<0.05$). There was no significant difference between the Wiimote and mirror devices.

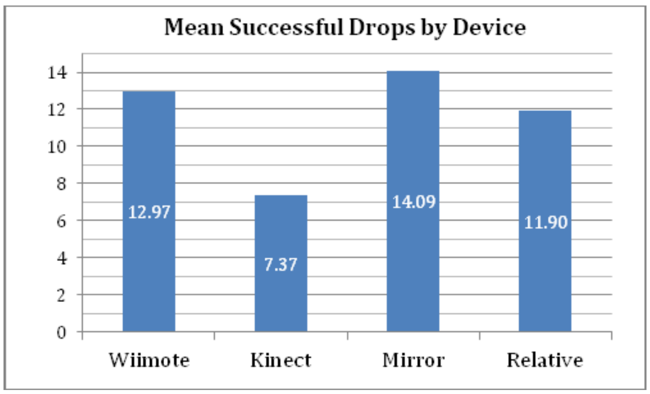

Figure 4: More drops show that the user was quicker and more accurate than users with fewer successful drops.

There is of course a direct relation between accuracy and speed in this task. The quicker the manipulation, the more likely you are to be able to complete the task, and even try multiple times in case of failure. Therefore an inaccurate but very quick technique could lead to misleading results. To investigate this we decided to look at the average target size for the last 5 targets subjects successfully hit. This allowed us to give subjects some additional practice time, and allowed subjects' performance to plateau. The results are shown in figure 5.

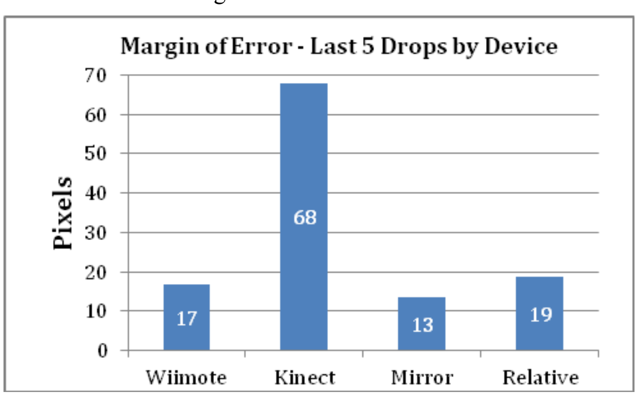

Figure 5: Average final margin of error in pixels by condition

Users were much less accurate with the Kinect than with any other device. Users of the mirror, relative and Wiimote conditions averaged a 13 to 19-pixel difference between the widget dimensions and the target dimensions. With the widget being 152px square, this meant a margin of error of less than ±9-12% of the widget size, or ±1-2% of the total screen real estate (1920x1080). For the Kinect the margin of error was ±45% of the widget, and ±6% of the total screen real estate. There were significant differences (One-way ANOVA $F(3,120)=13.77$, $P<0.001$ with Tukey's HSD for Post Hoc analysis) between the Kinect and all the other devices ($P<0.001$). Other differences were not significant.

During the experiment and in the post experiment interview, several subjects mentioned that the sensitivity for the relative tablet was low and that they would have to slide their finger across the device more than once to get the cursor to traverse from one side of the screen to the other. Users needed to swipe 3 times to go from one side of the screen to the other. No enhancements such as cursor acceleration were implemented; this could potentially improve performance of the relative condition. This may in part explain why the relative tablet scored worse than the mirror and Wiimote. However, because speed and accuracy are often traded off against each other, it is not a given that acceleration would lead to better results. This is something that should be investigated more in-depth. Users were not able to move the cursor rapidly enough to hit the same number of targets.

The issues with using the Kinect were more pronounced and deep-rooted. Subjects were observed having a difficult time both beginning a drag (selecting and dragging the target) and dropping the widget into the target (widget would often be dropped prematurely and unintentionally). A less common but also real problem was that in order to establish a difference between a click event and a drag event users had to press forward and hold for 0.5 seconds before beginning the drag. It was common to see users attempt to drag before the drag event had been registered. They were told about this in the training but as the user began the trial and were trying to rush through the task, they would often not pause long enough. Some visual indicator to let them know that the event had been registered could have made a difference.

The more fundamental problem with the Kinect condition was the 2-handed operations. Subjects usually had little trouble placing a cursor over a target using one hand, though fatigue was mentioned as a concern in some trials. However, the action of bringing or removing the second hand from the camera plane often caused subjects to inadvertently rotate their bodies to retain balance, even while seated. This of course would make their targeting hand move, resulting in a missed target. This same phenomenon was observed time and again across tasks and subjects. The only effective remedy we saw was for subjects to plant their elbows in the seat, and use this to counter the natural body rotation action. Though effective, this led to a very restrictive seating position.

4.2 Nudging

There is a tradeoff between speed and accuracy, and with a sufficiently fast UI, users can home in on the target effectively. We referred to this behavior as "nudging". In our experiment, this turned out to be a relatively common strategy; if a subject failed to hit the target on the first try, they would rethink their strategy (a longer pause) and then pick up the widget and home in through a series of rapid follow-up moves. This was especially common with the smaller targets, where the margin for error was low. The majority of times subject were able to hit the target in one or two attempts. However, sometimes it took longer.

Figure 6 shows how subjects using the Wiimote employed this strategy. Wiimote users were among the most successful and accurate, and the technique allowed for quick and easy nudging, or homing in on the target. As we can see, after a longer rethink following an initial miss, subjects engaged in a lot of rapid moves aimed at trying to hit the target. Subjects in this condition were still among the most accurate and successful. We see a very similar behavior among subjects using the mirror tablet application, though there is less of a long-tail (see Figure 7).

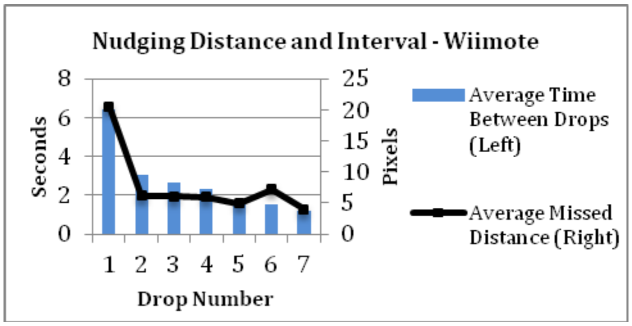

Figure 6: Nudging interval (left axis, blue columns, in seconds) and distance (black line, right axis, in pixels) over number of tries to hit one target – Wiimote.

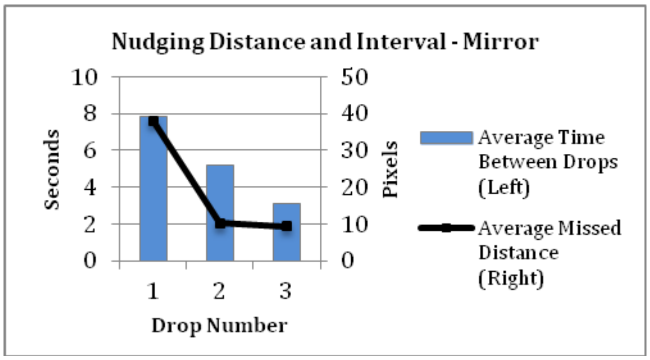

Figure 7: Nudging interval (left axis, blue columns, in seconds) and distance (black line, right axis, in pixels) over number of tries to hit one target – Mirror Tablet.

This strategy however seemed to be less successful, even counterproductive in the other two conditions (see Figure 8 and Figure 9). In the case of the Kinect condition, accuracy was an enormous issue, and though subjects were more successful with repeated tries, they did not home in on the target, but rather hit random new points. In the case of the relative tablet application, the nudging strategy appears to be counterproductive. Subjects would after the second try engage in very rapid moves that rather than take them closer to the target would distance them more. To us this is an important distinction between these two groups of techniques. Our subjects naturally gravitated to this strategy, and therefore it should be supported.

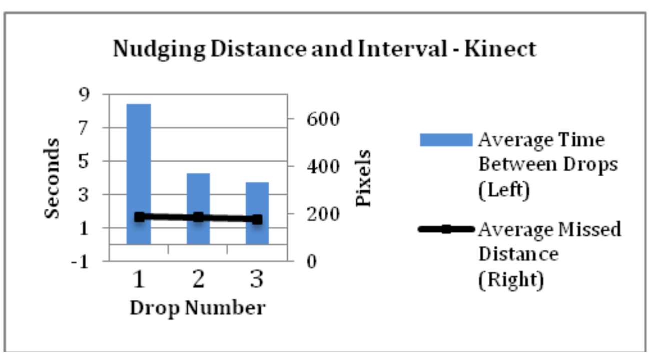

Figure 8: Nudging interval (left axis, blue columns, in seconds) and distance (black line, right axis, in pixels) over number of tries to hit one target – Kinect.

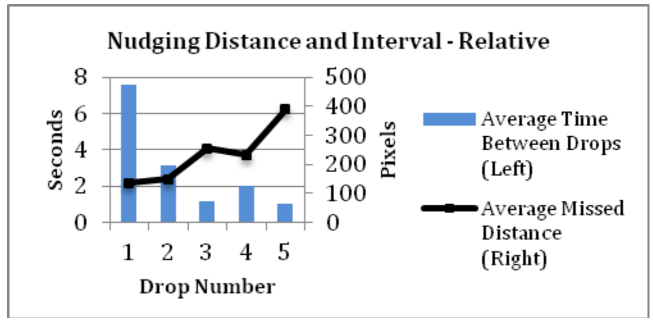

Figure 9: Nudging interval (left axis, blue columns, in seconds) and distance (black line, right axis, in pixels) over number of tries to hit one target – Relative Tablet.

4.3 Text Entry

An important task for iTV applications is text entry, as it allows more rapid customization, search, etc. We chose to examine two factors, the number of clicks that landed off of the intended target (key) and the time it took users to input the text string (a 9 character string).

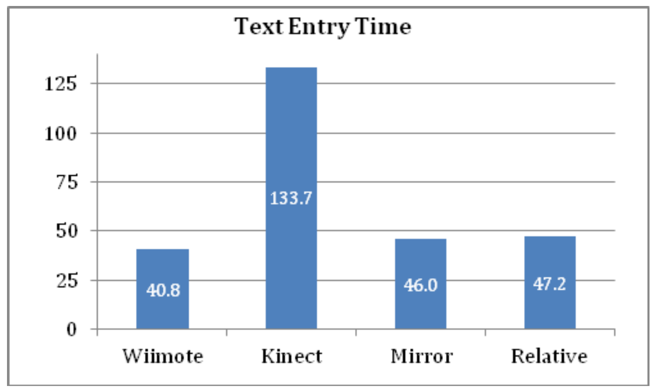

Figure 10: Average text entry time in seconds

As we see in Figure 10, text entry was significantly slower with the Kinect when compared to the other devices (One-way ANOVA $F(3,115)=19.53$ $P<0.001$ with Tukey's HSD for Post Hoc analysis). The Kinect was significantly slower than all other devices ($P<0.001$). There were no other significant differences.

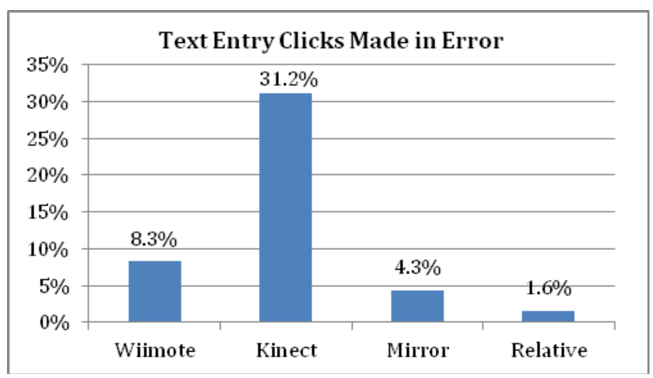

Figure 11: Text entry mistakes using virtual keyboard

Again, because a lack of accuracy can lead to slower task completion, we chose to look at how many mistakes subject made. A mistake in this case could be a subject hitting the wrong letter, or trying to click on something other than a letter on the on-screen virtual keyboard. As in just about every task in our experiment, the Kinect fared most poorly with 31.2% of clicks missing their target. This is significantly worse than the other three devices (One-way ANOVA $F(3,116)=19.14$ $P<0.01$ with Tukey's HSD for Post Hoc analysis).

Of the remaining devices the Wiimote fared the worst with an error rate of 8.3%, though this did not affect completion time. The mirror app came next, with an error rate of 4.3%, likely caused by the small size of the keys when shown on the tablet. The error rate for the relative app was a surprisingly low 1.6%. The difference between these devices was not significant.

4.4 Effects of Prior Experience

Prior experience can have a significant impact on performance, especially when dealing with the novel. Subjects were asked to rate their prior experience with devices similar to those used in our study. A linear regression was used to compare the number of successful drops in the drag and drop task and the speed of text entry versus their prior experience (see Table 1).

Table 1: Slope of linear models. Steeper slopes indicate stronger experience effect.

	Hits vs. Experience	Text Entry vs. Experience
Wiimote	0.61	0.33
Kinect	0.51	-5.42
Relative	0.39	-0.79
Mirror	0.25	-7.85

We see that the prior experience played the largest role in the Wiimote case. Despite being seen as universally easy to use, subjects were able to effectively leverage prior experience to improve further. The same was the case in the Kinect condition, though here, novices really suffered, and even experts performed marginally. There may have been a floor effect here as fewer subjects had experience with Kinect compared to other devices, and those that did have experience tended to have less experience than with the Kinect than with other devices.

More surprising, there was only a relatively mild learning effect for the two tablet solutions. While experience did help, it seems that these two techniques were so universally well-known and intuitive that all subject were able to use them effectively regardless of experience level.

5. DISCUSSION

In our post experiment interviews we focused on understanding the limitations and advantages of the different approaches. One thing we stumbled on were issues related to the sensitivity, or lack thereof for some of our conditions. 76% of users of the relative tablet mentioned that sensitivity was an issue (it took too long to scroll from one side of the display to the other). 43% of Wiimote subjects complained about the device being too sensitive, reacting to slight hand tremors.

Despite the negative results, 13% of Kinect subjects commented positively about its usefulness. This was obviously surprising, but shows that people like the concept of this technique, if not the implementation. We also found that 40% of those who used the mirror tablet and 47% of those who used the Wiimote commented positively about these. Surprisingly only 13% of the relative tablet users commented positively about it despite the high performance.

When asked if the input method could be learned quickly, the Wiimote won out, and it was also rated as the least awkward to use. Subjects liked the simplicity of the Wiimote, both in interface navigation and physically. One participant said, "It was simple. Just point and click. You just aim at it and it's right there." One of the most common answers about what people liked about the Wiimote input method was that it had only 1 button. People also liked the familiarity with holding the Wiimote, that it felt like a remote control and had a physical button.

By far, the least liked device was the Kinect. The most frequent negative comments had to do with physical fatigue and issues with sensor range and sensitivity. Having to hold their right arm up to position the cursor and left arm for click control resulted in almost all Kinect users complaining of arm fatigue. Finding a one-handed method for controlling the system could result in a significant improvement, as indicated by 60% of Kinect users. 60% also complained about the sensor range or sensitivity. To limit interference, the sensor was placed 80cm away from the subject. As a consequence, subjects felt they were unable to move their hands as far to the left and right as they would like. Although most comments focused on why the Kinect was not effective, several participants liked how it did not require them to hold a physical device. One participant talked about how nice it would be to have no remotes and control everything with gestures. Most subjects however indicated that they would be embarrassed to use this technique in front of friends and family.

The two tablet techniques achieved roughly the same ratings, which were generally good. 30% of mirror application users commented positively on being able to directly manipulate the interface. One person said "I really liked being able to click on exactly what I can see on the screen." Others disagreed, saying how they prefer to only have one screen to interact with. One participant commented "Occasionally I found myself not knowing which screen to look at."

A side-effect of our implementation of the mirror app was noticeable "lag" between the TV and the tablet images of between 0.25 and 0.5 seconds. This delay was often mentioned as an annoyance. Likewise, nearly everyone who used the relative tablet app disliked its low sensitivity and the lack of acceleration techniques. With appropriate tweaking, both of these techniques would have likely scored higher on both likability and effectiveness.

6. FUTURE WORK

This was meant as an exploratory study examining the relative merits of a number of gaming-related UI methods, and their usefulness in an iTV setting. Looking forward, there are several improvements worth exploring based both on user feedback and our findings. As mentioned in the results and discussion sections, implementing motion smoothing for the Wiimote, acceleration for the relative tablet app, and reducing the lag for the mirror app are natural next steps. We believe all of these could drastically improve the user experience.

In implementing smoothing for the Wiimote, a slight delay will be introduced. Pavlovych and Stuerzlinger [3] studied the relationship between jitter and latency and their results could help inform an appropriate balance.

A more tricky problem was the noisiness and occasional false positive for hands for the Kinect. We were unable to use the Kinect API's native skeleton tracking because only the upper half of the users' body was visible to the sensor. Instead, we used a hand tracking method and filtered based on depth field data. Even with these precautions, a knee or other object could register as a hand. This led to a very frustrating user experience. In future revisions, we would look for a more robust tracking solution. We did not use the official Kinect SDK as it was not available in time.

When asked what they would change about the Kinect method, several people said they would prefer a one-handed solution. We think this would substantially decrease the physical fatigue and provide a more intuitive experience. Due to the Kinect's 640x480 resolution depth camera, robust finger tracking was difficult. Perhaps with a different library or algorithm, a more feasible approach could be found. One option might be the work of Oikonomidis et al. [8], who have demonstrated complex finger articulation, though not in real-time.

We chose not to investigate the use of voice commands in our experiment, in part because it would be difficult to filter noise and could be socially awkward. As the introduction of the Siri system on the iPhone, and the flurry of interest this has caused, these assumptions and prejudices may need to be revisited in future work.

7. CONCLUSIONS

With the growing availability of broadband Internet access, highly extensible game consoles, and the increasing popularity of social and streaming online entertainment services, their convergence is presenting a number of new challenges for HCI researchers. To the best of the authors' knowledge, there has been very little research on the adoption and use of novel game controller technology in a media center or iTV context. While our research was largely exploratory, we hope it will serve as a base for future work in this area.

Looking at the results we see that devices designed for gaming have the potential to be effective input devices for a typical iTV interface. We also see that some devices are better suited to this task than others. The Wiimote was effective, well liked, and very easy to learn. At the same time, it offered ample room for improvement as users gain experience. On the other hand, it is potentially limiting UI-wise, as all information has to be displayed on the primary display. The tablet systems were both well liked and effective as well, though they potentially offer more flexibility and exploration, albeit at a much higher hardware cost.

The Kinect, though appealing to many subjects due to its novelty and the promise of device-free interaction, proved to be too unreliable and cumbersome to use for any extended period of time. While it may be refined with better hardware and algorithms, its suitability and desirability for a social lean-back viewing experience may be limited. Fear of ridicule as much as physical fatigue and the problem of interference from others' movement are serious problems that need to be overcome.

REFERENCES

[1] A. D. Wilson and M. Agrawala. "Text entry using a dual joystick game controller." In Proc. of the SIGCHI conf. on Human Factors in computing systems (CHI '06), ACM, New York, NY, USA, 475-478.

[2] A. Oulasvirta and J. Bergstrom-Lehtovirta. "Ease of juggling: studying the effects of manual multitasking." In Proc. of the 2011 annual conf. on Human factors in computing systems (CHI '11). ACM, New York, NY, USA, 3103-3112.

[3] A. Pavlovych & W. Stuerzlinger. "The tradeoff between spatial jitter and latency in pointing tasks." In Proc. of the 1st ACM SIGCHI symposium on Engineering interactive computing systems (EICS '09). ACM, New York, NY, USA, 187-196.

[4] D. Natapov, S.J. Castellucci, and I.S. MacKenzie. "ISO 9241-9 evaluation of video game controllers." In Proc. of Graphics Interface 2009 (GI '09). Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 223-230.

[5] D.C. McCallum and P. Irani. "ARC-Pad: absolute+relative cursor positioning for large displays with a mobile touchscreen." In Proc. of the 22nd annual ACM symposium on User interface software and technology (UIST '09). ACM, New York, NY, USA, 153-156.

[6] G. Shoemaker, L. Findlater, J.Q. Dawson, and K.S. Booth. "Mid-air text input techniques for very large wall displays" In Proc. of Graphics Interface 2009 (GI '09). pp. 231-238.

[7] Hulu, 2012. "Hulu Financial Results:" http://blog.hulu.com/2012/01/12/2011-2012-and-beyond/

[8] I. Oikonomidis, N. Kyriazis, and A. Argyros. "Efficient model-based 3d tracking of hand articulations using kinect." Procs. of BMVC, Dundee, UK, pages 1-11, 2011.

[9] I.S. MacKenzie and S.X. Zhang. "The immediate usability of graffiti." In Proc. of the conf. on Graphics interface '97, Toronto, Ont., Canada, Canada, 129-137.

[10] J. C. Lee. Hacking the Nintendo Wii Remote. IEEE Pervasive Computing, 7(3):39-45, July 2008.

[11] J. Lin, H. Nishino, T. Kagawa, and K. Utsumiya. "Free hand interface for controlling applications based on Wii remote IR sensor." In Proc. of the 9th ACM SIGGRAPH Conf. on Virtual-Reality Continuum and its Applications in Industry (VRCAI '10). ACM, New York, NY, USA, 139-142.

[12] K. Cheng and M. Takatsuka. "Initial evaluation of a bare-hand interaction technique for large displays using a webcam" In Proc. of the 1st ACM SIGCHI symposium on Engineering interactive computing systems (EICS '09). ACM, New York, NY, USA, 291-296.

[13] L. Rainie. "Tablet and E-book reader Ownership Nearly Double Over the Holiday Gift-Giving Period" Pew Internet. 2012.

[14] Netflix, "Netflix Q4 2011 Letter to Shareholders." http://goo.gl/4vgP1

[15] Nielsen, 2011. "Nielsen: The Cross-Platform Report." http://goo.gl/JxEpn

[16] Nielsen. "State of the Media: Consumer Usage Report," 2011. http://goo.gl/PzAzc

[17] NTIA, 2010. "Exploring the Digital Nation: Home Broadband Adoption in the United States." http://goo.gl/ErsnS

[18] S.J. Castellucci and I.S. MacKenzie. "Unigest: text entry using three degrees of motion" In extended abstracts of the 2008 Conf on Human factors in computing systems; CHI '08. ACM, New York, NY, USA, 3549-3554.

[19] Sandvine, 2011. "Sandvine. Global Internet Phenomena Spotlight: Netflix Rising." http://goo.gl/dbkPz

[20] T. Költringer, P. Isokoski, and T. Grechenig. "TwoStick: writing with a game controller." In Proc. of Graphics Interface 2007 (GI '07). ACM, New York, NY, USA, 103-110.

[21] T. Ni, D. Bowman, and C. North. "AirStroke: bringing unistroke text entry to freehand gesture interfaces." In Proc. of the 2011 annual conference on Human factors in computing systems (CHI '11). ACM, New York, NY, USA, 2473-2476.

[22] YouTube, 2012. YouTube Press Statistics: http://www.youtube.com/t/press_statistics