Studying a Head Tracking Technique for First-Person-Shooter Games in a Home Setting

Torben Sko, Henry Gardner, Michael Martin

HAL Id: hal-01510552
https://inria.hal.science/hal-01510552

Submitted on 19 Apr 2017

Studying a Head Tracking Technique for First-Person-Shooter Games in a Home Setting

Torben Sko[1], Henry Gardner[1], and Michael Martin[2]

[1] The Research School of Computer Science, The Australian National University
[2] The Research School of Finance, Actuarial Studies and Applied Statistics, The Australian National University

{torben.sko, henry.gardner, michael.martin}@anu.edu.au

Abstract. This paper examines webcam-enabled head tracking for games in a home setting. A new head interaction technique was developed based upon prior laboratory-based research, with a focus on making it robust to the variable conditions of a home setting. Our technique was integrated into a test-bed game and 550 hours of gameplay data was collected from 2500 users, many of whom also provided formal feedback. The head tracking performed creditably and players reported that the experience was more immersive. Head tracking failed to enhance competitive playing performance, perhaps owing to familiarization effects. Nevertheless, the data revealed evidence of learning amongst users, suggesting that performance would improve with continued use. Key lessons that emerged in the home setting in contrast to the earlier laboratory study were a demonstrated need for clear guidance and feedback during system set-up, and greater caution regarding its deployment, having discovered a small population of users who became nauseous.

Keywords: head tracking; gestural interaction; online studies; games.

1   Introduction

In recent years, webcams have become a ubiquitous form of technology. When placed in a desktop setting – where a user sits up-close to their computer screens – webcams can be used to track the 3D location and orientation of the user's head position. Despite the availability of this technology, head tracking has seen relatively little adoption, even in computer games that have elsewhere seen a significant move towards gestural interaction. Our previous head tracking work [23] highlighted several barriers to adoption, with the most prominent of these being the perceived reliability of the tracking system. This issue was presented in the context of a laboratory setting and techniques of a prototype quality. Consequently, it could only be hypothesized what the real-world impact of this issue was – when considered in the context of a home setting and a consumer-grade technique.

This paper presents the first study of its kind to examine a refined head tracking technique in a household setting. We conducted a remote user study with a built-in logging system to allow us observe how our technique functioned when used by home

players. In doing so, we looked to examine whether the technique could work reliably enough within this uncontrolled environment and provide enough benefit that players would adopt it. In examining this adoption, we were also keen to explore what impact our head-tracked interaction technique had on the user's competitive playing performance, as compared to a button alternative. Such an investigation responds to the idea that "hardcore" PC players can be quite dismissive of gestural interfaces on the basis that they do not yield any playing performance benefits [11]. The public exposure the study entailed also provided an opportunity to examine how a large audience of players reacted to the idea of using head tracking in games – a concept that still remains quite novel. This paper examines this reaction in an effort to identify what qualities may help to make a head tracking technique more appealing.

The next section provides some background and further motivates the decision to conduct the study in a home setting. Sections 3 and 4 present a refined interaction technique and detail the study used to assess it. Section 5 presents the results of the study, while Section 6 summarizes our findings as a set of head-tracking guidelines, designed to help others develop their own interaction techniques.

2   Background

Head tracking was first reported in the literature in the late 1960's by Sutherland [26] who attached a mechanical arm to a user's head to detect their head pose. By contrast, modern head-tracking libraries, such as [22, 28], can now function with just a standard webcam (see [17] for a full head tracking review). Following these improvements, researchers have explored the use of head tracking within various desktop applications, including gaming [3, 12]. For example, the work of Wang et al. [27] presented a head-based leaning technique in a first-person-shooter (FPS) game that allowed players to dodge bullets. Their work found that the inclusion of head tracking improved a measure of player *presence*. The commercial game, *ArmA 2* [2], adopted a similar technique that also allowed players to look around independently of their aim. This title required use of the infrared-based hardware system, TrackIR [18]. Another tracking system more recently released, is the Kinect [16]. Despite the availability of this hardware, only a few games exist that specifically utilize the head as a gestural controller.

In the absence of any widespread adoption, our previous work [23] examined the ways in which head tracking could be utilized in an FPS game, by comparing Wang's leaning technique [27] to six other head-interaction techniques. In this work we implemented and evaluated our techniques using focus groups. The general reception was positive, although gamers were also concerned about the *reliability* and *robustness* of the tracking system. As this work was conducted in a controlled, laboratory setting, how these concerns might translate to real-world use within a household setting was a matter for conjecture.

Fig. 1: A user performing a right lean. The amount the player needs to tilt is kept very subtle to ensure the interaction technique is not too strenuous. During the lean the screen and gun model are both rotated and the player shifts sideways.

Many existing studies have acknowledged the importance of the real-world environment in assessing the usability of new technology. For example, Hartson [13] argues that both the test settings, along with properties of the system itself, are intrinsic factors that impact upon the usage patterns, making it difficult to replicate in a laboratory setting. The work of Brown [5] similarly points out (in the context of a geo-locating tool) that while the lab provides a good setting for exploring the technical feasibility, it provides little insight into *user acceptance*. These issues, that of the real-world realization and adoption, are particularly relevant when discussing webcam-based head tracking given the use of vision-based algorithms. The performance of the tracking is susceptible to environmental factors, such as lighting conditions and physical set up – making it necessary to use a home setting if we wish to know its real-world performance and adoption.

To help deliver studies within remote settings, researchers have increasingly taken to using the Internet. The benefits of online studies include the ability to reach potentially thousands of home users [7], across a large and varied sample [29]. For example, Kittur et al. demonstrated that is possible to utilize existing crowdsourcing systems (i.e. systems where users complete small tasks for monetary incentives) for the purposes of running a user study [14]. The level of sophistication employed varies dramatically, from simple surveys [10] to sophisticated remote studies where the participant and researcher are separated *physically* and *temporally* [13]. In these latter studies, a logging system is employed to capture how the user interacts with the study software, which frees the user to engage with the study in their own time [21]. For example, the work of Costanza et al. [8] used a logging system to capture the visual and auditory output of their software, allowing them to later replay its use. Some of the other feedback mechanisms that researchers have used include; text and video chat, questionnaires, interviews, forums, and diaries [1]. Our own study employed many of these concepts, as described in Section 4.

3   Technique Description

The head interaction technique presented in this paper is referred to as the Handy-cam and Leaning (HAL) technique (pictured in Fig. 1). It combines and improves upon the two most promising techniques identified in our earlier work [23]. The first of these was a *leaning* technique, which was first explored by Wang et al. [27]. This technique allowed a player to perform an in-game lean movement by physically leaning (that is, tilting and/or offsetting their own head). The resulting in-game action allows players peer out from behind cover so they can spy on enemies and better protect themselves during a gunfight. When featured in an FPS title, this form of control is typically mapped to the 'Q' and 'E' keys. By comparison, using head-movements to perform this form of control is potentially more intuitive [23] and can provide a finer level of control through its continuous nature [25]. In Section 5.2, we examine whether this increased level of control can indeed yield a stronger playing performance.

The second technique chosen was the *handy-cam* effect, which aimed to reinforce the intention that a first-person game is being viewed through the eyes of the protagonist. The technique worked by adding a "wobble" to the virtual camera so that the player's own subtle and involuntary head movements were mimicked by the virtual camera. For example, if the user was pitching their head slightly upwards, the virtual camera was subsequently pitched slightly upwards (likewise for the other head axes). A similar effect is used in modern cinema in films like "The Bourne Identity" (2002), where a subtle wobble is added to the view to make it look as if the action is shot on a handheld camera. This effect can make the viewer feel more *present*, as if they are viewing the action directly with their own eyes [15]. Adding a jitter has also been found to make the scene appear more *realistic*, by increasing the perception of self-motion [6]. Several studies have examined the effects that a *pre-animated* jitter can provide, both in pre-recorded video [20] and games [9]. Our earlier work was the first known attempt to create an interactive version of this effect, where it was unanimously well received by the focus group participants [23].

3.1   Technique Implementation

The design of HAL has seen our earlier prototype techniques [23] greatly matured into a technique designed for mass-use. More specifically, whereas our previous implementations were targeted at a single computer set up, HAL needed to work across a multitude of home computer configurations. To make the technique cope with variations in the set up, the following enhancements were added:

- No assumptions were made about the user's resting position in order to cater for differences in the physical configuration of the home systems.
- Time-based smoothing (rather than sample-size based) was used to account for computer speed variability.
- Smoothing of the head data was adapted based on the tracking quality, thereby allowing HAL to remain responsive on slower systems.
- An easing curve was added to the leaning for finer control over smaller lean movements.

- Several minor tweaks were added to avoid any negative visual impacts the head technique may create. These included constraining the range of head mapped movements and fadding out/in following a head loss/aqausition.

Our HAL technique can be thought of as a mapping from the user's head position to a normalized *leaning-amount* and a *camera offset*. The head data was obtained using Seeing Machines' faceAPI (v4) software [22], which provides both the position and orientation of the user's head. From this data, the leaning-amount was computed using both the roll and sideways head movement (as both axes are typically involved in a leaning action). To derive the leaning-amount the following logic was applied:

```
01.  rollResting = UpdateMean(rollResting, headData[ROLL], rollRange);
02.  roll = headData[ROLL] — rollResting;
03.  roll = Normalize(roll, rollMin, rollMin + rollRange * leanScale);
     ... // Repeat the above for sideways movements
04.  lean = Normalize(roll + sidew, 0, 1);
05.  lean = lean * (1 — easeAmount) + Ease(lean) * easeAmount;
06.  lean = MovingAverage(lean, leanSampleTimeframe);
07.  lean = FadeIn(lean, 2);
```

The bolded variables could be tuned using the settings panel, thereby allowing users to customize how the technique functioned. In lines 01 and 02, the head value is converted from an absolute position into an offset relative to the user's resting position, allowing us to identify how much they have moved. As the head position is stated relative to the webcam, the user's resting position will be a non-zero position vector (e.g. if the user sits 80cm away, the reported depth will be 0.8). There are several sources of variation that will influence this position, including the camera's position, the user's seating position, as well as how the software interprets the user's face. In our earlier work [23] this position was assumed, which caused the technique to behave incorrectly for some participants. Our HAL technique instead employed a running mean to compute this resting position. To avoid having this position shift too much during play, the mean was only updated when the new head data was within a certain range of the mean (*rollRange* = 30°, *sidewRange* = 15cm).

In line 03, the roll amount is normalized over a usage range. This range is scaled (*leanScale* = 1, *sidewScale* = 1) in order to control the sensitivity of the technique. A minimum is used (*rollMin* = 3°, *sidewMin* = 2cm) to introduce a "dead zone" or no-movement zone. This dead zone ensures subtle head movements avoid invoking a lean movement, which could otherwise interfere with the player's aim. In line 04, the two components of the lean movement are combined and normalized between 1 (ful left lean) and -1 (full right lean). In line 05, an ease-in/ease-out curve is applied to the lean-value based on the *easeAmount* (= 100%). The leaning value is then smoothed using a moving average (line 06, *leanSampleTimeframe* = 0.2sec) and faded (line 07) in the event that the user has only started to be tracked within the last two seconds.

The lean-amount obtained from the previous logic was then mapped to a leaning action. In the case of our game, this corresponded to a sideways player movement, a gun movement, and a slight reduction in the field of view (FOV). The gun movement

and FOV reduction together recreated some of the look-and-feel of the *iron sighting* technique, which was well received in the previous focus group studies [23]).

The other key aspect of HAL was the handy-cam effect. To compute the camera offset, the following logic was applied:

```
01.    leanMean = MovingAverage(lean, 0.5);
02.    pResting = UpdateMean(pResting, headData[PITCH]);
03.    p = headData[PITCH] - pResting;
04.    p = p * handycamScale * pitchScale;
05.    p = p * (1 - (leanMean * dampenWhileLeaning));
06.    p = MovingAverage(p, handycamSampleTimeframe);
07.    p = FadeIn(p, 2);
08.    handycamOffset[PITCH] = p;
       ... // Repeat the above for all axes except head depth
09.    cameraPosition += handycamOffset;
```

As before, the head data is converted into an offset (lines 02 and 03, above). In line 04, the effect is scaled using both a global scale (*handycamScale* = 1) and an individual scale for each axis (*pitchScale* = 1, *rollScale* = 1, etc). These scaling factors allowed the user to control the strength of the technique. In line 05, the offset is potentially dampened during times that the user was leaning (*dampenWhileLeaning* = 50%). This was included to avoid having the effect interfere with the player's aim whilst leaning. The value was then smoothed (line 06, *handycamSampleTimeframe* = 0.2sec) and faded (line 07). This was repeated for all the other head axes, except for movements along the forward-facing axis as these were found to negatively impact on the appearance of the weapon model positioned in the foreground of the scene. The computed offset vector was added to camera's position (line 09) to create the handy-cam effect. A complete implementation of the HAL technique can be found online[1]. It has been implemented using freely available software.

4   Study Design

To study HAL we wanted to observe; how it performed, to what degree it was adopted by players, and what players thought of it. To achieve this we implemented HAL into the commercial FPS title, Half-Life 2 (HL2). This allowed us to utilize a wealth of existing game content, which provided a rich setting in which to examine our technique. As HL2 does not include leaning as part of its control set, we modified the game to include it. We also constructed a custom lean-centric game mode to explore the hypothesis that a greater focus on head-control would elicit a more positive user response [23]. This custom mode was called FaceOff Paintball (FOP). The following summarizes the differences in these game modes:

---

[1] https://github.com/torbensko/HAL, *accessed Apr '13*

- *HL2 mode:* existing game levels of Half-Life 2. The levels were not designed with leaning in mind, but can be played with leaning (controlled using either the keyboard or head movements) with our implementation.

- *FOP mode:* a novel paintball game level where players try to survive for as long as possible. The player (and enemies) only take one shot to be "marked" (i.e. eliminated) making it important to use the provided cover, which we felt could be best done through the use of leaning.

We added several mechanisms to both game modes that allowed us to examine how the HAL technique was performing and being used. Firstly, we monitored and logged characteristics of participants' use of the system including their head position, the leaning value, their playing performance (e.g. game score), system performance (e.g. frame-rate), the hardware (e.g. camera model), and potentially a picture of their set up as taken with their webcam. These data were only sent if the user opted into the study aspect of the game. At the time of downloading the game we surveyed the participants (via a sign-up form), asking them about their *age*, *gender*, *gaming experience* and their *enthusiasm towards head tracking*. We were able to associate this information to the logged data using a sign-in panel within the game.

To collect the player's opinion of the technique, the game was modified to include a post-play questionnaire, made up of 45 questions, most of which requested responses on a five-point Likert scale. We also incorporated an optional comparison test, which provided us a more formalized way of examining the impact head tracking had on the game experience. Within the test, users played the game twice, once with the *head tracking enabled*, and once with it *disabled* (randomized order between users). This test could be played under either HL2 or FOP game modes. Under the HL2 mode, players could choose to play with either the traditional control set (without leaning) or with the extended control set (including leaning). This meant there were
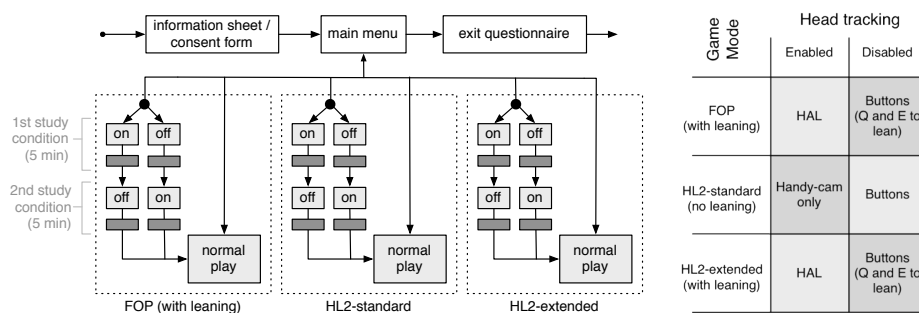

Game Mode	Head tracking	
	Enabled	Disabled
FOP (with leaning)	HAL	Buttons (Q and E to lean)
HL2-standard (no leaning)	Handy-cam only	Buttons
HL2-extended (with leaning)	HAL	Buttons (Q and E to lean)

Fig. 2: A flow diagram for the online experiment (left) and resulting comparison conditions (right). In the flow diagram, the mid-gray boxes represent states where the user is playing. The black dots indicate where the system auto-assigned users to a condition. The 'on' boxes denote the condition where head tracking was enabled, while in 'off' head tracking was disabled. The darkest boxes represent the in-game survey. The system forced the surveys to be viewed but did not prevent users dismissing it without answering any questions. The player could freely switch between any of the game modes and could complete the comparison test multiple times.

two conditions: head control (*enabled / disabled*) and game mode (*FOP / HL2-standard / HL2-extended*). When the playing under the HL2-standard mode (without leaning), only the handy-cam aspect of the head-interaction technique was in effect, allowing us to study its impact on its own. At the end of each condition, an in-game survey was presented to the users. A summary of the study is provided in Figure 2.

Our study system was placed online and promoted through the use of two YouTube videos. The release of the work generated a considerable amount of online discussion, which was found to provide a valuable source of feedback [21]. When combined with the open text response taken from the exit questionnaire (479 comments), a collection of over 2000 comments was accumulated. These comments reflected how HAL was both *perceived* (prior to using) and *received* (after using it).

5 Results

Playing data was collected from 2593 users from across 100 countries, with the US, Russia and the UK being the most common countries of origin. The most represented group was young males with "lots" of FPS experience. Through the use of a pre-play survey, administered during the sign-up process, we found that 75% professed to being "excited" about the use of head tracking. The median playing time was 7 minutes. A total of 550 hours of playing data was logged.

The remaining part of this section is split based on the main research goals; Sec. 5.1 examines how the technique performed in an uncontrolled environment, Sec. 5.2 explores how players adopted the technique and its impact on their playing performance, while Sec. 5.3 examines the players' opinion of HAL as a way of identifying the qualities that have most influenced its reception.
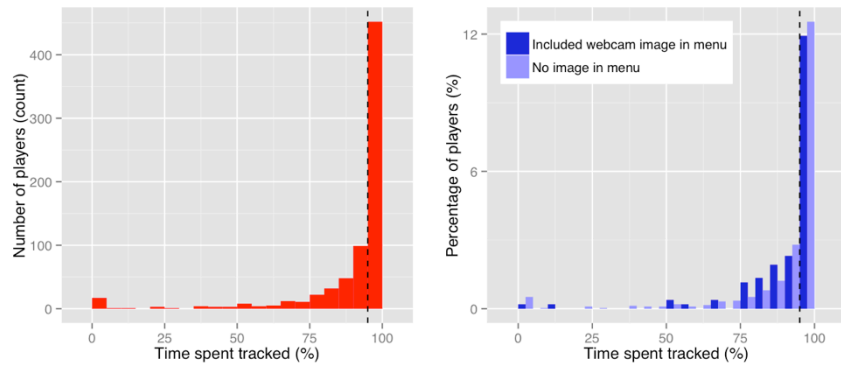

Fig. 3: System reliability: the players have been binned based on the percentage of time they were successfully tracked during the *tracking enabled* condition of the comparison test. Two versions of the study software were distributed, with one including an image of the user featured on the main menu – in the hope it would help them recognize deficiencies in their set up. No improvement can be observed when this image was included (as supported by a statistical comparison).

5.1 Technique Performance

By logging HAL's functional performance, we were able to examine how well it performed in a household setting. Figure 3 illustrates the reliability of the system by binning participants based on the tracking rates during the *tracking enabled* condition of the comparison test. Of the 726 participants that played under this mode, 62% (452) were tracked for 95% of their playing session or more. By contrast, 2% (17) were never tracked.

When examining the *complete* participant sample, it was noted that 23% (595) were *never* successfully tracked. Examining the hardware details of these players revealed that 34% (202) of these non-tracked players did not have a working camera – meaning that their camera was incompatible, incorrectly configured, or non-existent. The collected forum comments highlighted that built-in cameras were the most troublesome. The other 66% of the non-tracked players (393 users) did a have recognized camera, but were still never successfully tracked. Several images were collected from the user's own webcams to help explain why the tracking failed. Figure 4 demonstrates that some players used the system under far-from-ideal conditions. Some of the common issues encountered included: partially cropped and obscured faces (e.g. some users wore obscuring apparel, others sat too close); dimly lit settings (e.g. some users were lit via their monitor only); high contrast images (e.g. over-exposure of the image due to bright lighting); blurry/unclear images; and pre-processed images (e.g. with some users featuring partial overlays and virtual avatars). Because many of these issues could have been corrected on the user's side, the software was modified during the study to display an image from the user's own webcam on the main menu of the system in the hope that users would identify and rectify deficiencies in their own setups. As Fig. 3 shows, this had limited impact on tracking reliability, an observation verified by performing a classification of user images based on the apparent issues.
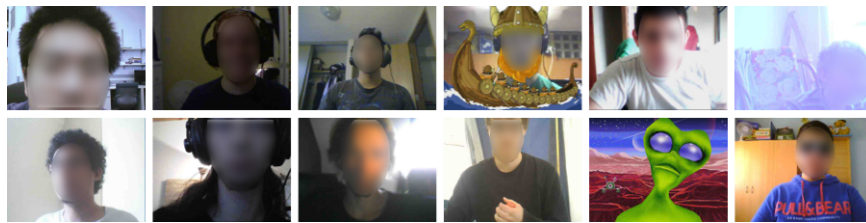

Fig. 4: Some of the webcam images (with faces blurred to protect identities). Many of these images show conditions that would make a user's head difficult to track: dim lighting, overexposure, obscured facial features (e.g. sun-glasses), a highly angled camera. In the case of the alien head and Viking boat, the webcam feed has been pre-processed before being received by the head-tracking library. Explicit permission was sought from users before collecting these images. The study was approved under the ethics protocol, 2010/032.

5.2   Player Adoption

The log data from the successfully tracked players provided an insight into how HAL was used by players. The data illustrated that the FOP game mode was successful in encouraging more leaning, with the head leaning being more popular than button leaning under both game modes. As one participant noted, *"I rarely make efficient use of the lean function in games because it usually requires some finger gymnastics to strafe and lean at the same time. I found myself leaning much more when all I had to do was move my head. I'm very pleased with this idea."* Button-based leaning saw almost no use within the HL2 game mode. As the game content did not encourage leaning, this outcome would suggest that the novelty of the leaning control alone was not enough to encourage players to use it. However, when the leaning control was performed via head tracking its use increased and was well sustained across the users who chose to play for more than an hour under the HL2 mode (30 players).

The playing logs also revealed how much players tended to move while using the technique. Figure 5 shows the amount of head movement under three different control schemes: HAL (both head leaning and handy-cam enabled), handy-cam only and head-controls disabled. In the disabled condition the player's head position was still logged despite having no apparent effect on the game, allowing us to observe the natural head movements that occur during play and establish a control group. The data in Fig. 5 show that players moved their head the least when the head control was disabled. Users appeared to move least in the vertical axis and the roll axis. By comparison, the pitch and yaw axes exhibited a greater amount of movement, making it more likely that they would yield false-positives if used for head controls. This result affirms our decision to develop a technique that predominantly relied on the roll axis.
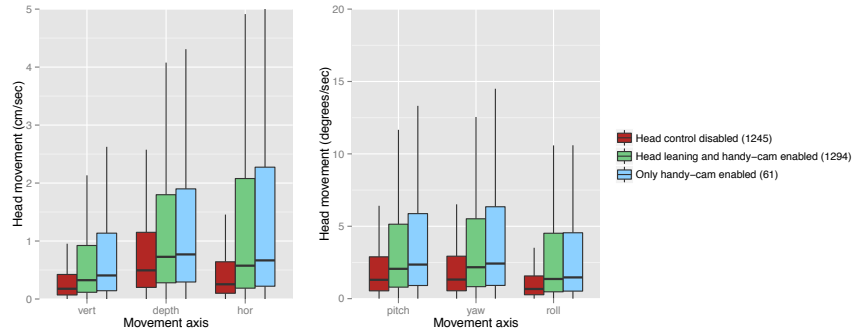

Fig. 5: The amount of head movement under different control schemes. The data has been sourced from the HL2 game mode during the comparison test. A half-second smoothing window has been used to account for tracking noise. The control schemes were taken from the different test conditions (see Fig. 2 for a list of conditions). The numbers in parenthesis indicate the number of applicable players, noting that only a few players chose to disable the leaning control, hence why the *handy-cam only* category is less populated. Nonparametric Wilcoxon tests reveal the *disabled/HAL* and *disabled/handy-cam* conditions to be significantly different ($p < 0.01$). By contrast, the *HAL/handy-cam* were not found to be significantly different.

The data shows that the amount of roll movement increased when the head leaning was enabled, however, even under this condition, this axis still exhibited less energy than the other two rotational axes.

Unexpectedly, the charts reveal that players tended to move most when only the handy-cam effect was enabled. It was expected that the HAL condition would exhibit the most movement, as part of performing the leaning control. By comparison, the handy-cam was intended to utilize the user's subtle and involuntary head movements. This finding suggests that users were trying to pro-actively use the handy-cam effect, presumably to look around, illustrating that its design intention was not well conveyed. When used in conjunction with the leaning, the amount of movement dropped. This may suggest that users were not as inclined to use the handy-cam for practical purposes if a more obvious use of head tracking was also present. For this reason, the handy-cam is probably best used alongside another head-interaction technique.

Returning to the topic of HAL's adoption, another aspect to consider was the impact that head tracking had on the users' playing performance. Past research shows that an interface that provides performance benefits is more appealing to players [11]. To explore HAL's performance impact, our system included an online leaderboard within the FOP mode, which globally ranked players based on their score. By including this indirect competition, it was hoped players would focus on their playing performance. A total of 24,000 attempts were recorded, lasting an average of 36 seconds. It was interesting to note that the top five players consisted three who chose not to lean, one player who mainly used button leaning (3rd place) and one head-leaning player (4th). Three of the top 10 players mainly used head leaning. A mixed-effects model was used to examine how button and head leaning affected the score.

For each attempt the number of marks (i.e. eliminations) was divided by the attempt duration to yield a *marks-per-second score*. The one-shot-eliminates nature of the FOP mode meant that the scores were not normally distributed, so a log of the marks-per-second score was taken (after removing scores of zero). The resulting "*log-score*" performance measure was normally distributed. We also added to the model aspects like *prior use* of the system (e.g. prior duration in FOP mode, prior head/button leaning), information about the attempt (e.g. duration of play, percentage of time tracked), as well as demographic and hardware information. The usernames were fit as a random effect in the model as observations from the same user were likely to be correlated with one another, while observations between users were assumed independent. All other covariates were fit as fixed effects.

After an initial 'full' model fit, all clearly non-significant variables were excluded and reduced model was assessed using both deviance-based criteria and AIC

Table 1: The coefficients of the mixed-effects model. The percentage of time spent head leaning includes the periods where the user performed only a partial lean movement.

	(Intercept)	Attempt duration	Time tracked (%)	Head leaning (%)	Button leaning (%)	Computer speed (frame-rate)	FPS experience: "lots"	Sex: "male"
Estimate	-2.6673	0.004	0.3273	-0.2456	0.5304	0.0008	0.0479	0.1627
Std. Err	0.0942	0.0002	0.0176	0.0166	0.0255	0.0001	0.0216	0.0918

(Akaike's Information Criterion). This reduction process was continued, until a final model was reached for which the remaining model terms were significant. The estimated coefficients, with standard errors, for the final model are shown in Table 1. The model indicates that gender (male) had a positive influence on log-score, as did having extensive FPS experience, a faster computer, as well as being reliably tracked. Our mixed-effects modeling revealed that button leaning had a positive influence on the leaderboard score, but that head-tracked leaning *negatively impacted* it. With users only playing for an average of seven minutes, it is possible that this detrimental effect was the result of a familiarization period. A correlation test found the accumulated duration of play had a significant positive correlation with the score (0.147), illustrating that a learning effect was present. This learning effect was also observed when examining the two forms of leaning. The accumulated duration spent button leaning and head leaning were both found to have a significant positive correlation with the score (0.123 and 0.107 respectively), albeit with the head leaning being harder to learn. This learning effect is encouraging, as it suggests that HAL may be able to eventually facilitate a better performance over time.

To supplement this test, we examined whether the participants' use of the leaning technique changed over time. Little change was observed over a one hour period, suggesting it probably takes several hours of play to become reasonably skilled using HAL. Such an outcome suggests a longitudinal study would be required to properly examine the playing performance impacts of head tracking. Such a study of competitive performance may benefit from adopting a richer control set than that used by us in the FOP mode in order to mitigate the trade-off that occurs between the *physical affordance* of an input method, the perceived *in-game benefit* and the *reliability* of the head tracking system. As one participant noted, *"The best keys for leaning are "Q" and "E", but I would rather use those keys for more important actions. If it works right this would free up some valuable real estate."* Similarly, having players compete directly against each other in the gameplay may help to address the familiarization performance slump. As one participant noted, *"It would be interesting to see if the "I'll stick to my keys" players would change their tune if they started getting owned by head tracking player because of this difference."*

5.3 Player Opinion

With head tracking remaining a relative novel concept within gaming, we wanted to discover what player's thought of HAL so that we may better understand what influences its appeal. The players' opinions were, in part, collected via the comparison test surveys and the exit questionnaire. An exploratory examination of the data was conducted using a classification tree (CART) approach [4]. This approach seeks to predict (classify) response variables based on the covariate values by recursively partitioning the covariate space through a sequence of binary splits on individual covariates. Each step in the algorithm chooses a covariate that provides the most homogeneous groups, delivering a tree-like structure for the model. The tree model approach does not require the pre-specification of interaction terms; it allows for complex multi-way interactions; and it caters for a mix of continuous and categorical variables

without imposing any particular parametric structure, making it well suited to exploring rich datasets, such as those yielded by online studies [24].

Some of the covariates included within our tree modeling included: the *tracking reliability* and the *stated prior enthusiasm for head tracking* (to help explore the presence of an enthusiasm bias), *hardware performance* (e.g. frame-rate, to help explore whether a minimum hardware requirement exists), and the *prior playing time* (to test whether longer exposure improved the player's opinion). For the comparison test questions, the model also included the *game mode* (to test whether a greater focus on the head-technique yielded a more positive result), and the *head tracking condition* (i.e. enabled/disabled).

The CART trees for the exit questionnaire and in-game surveys both told a very consistent story. From the exit questionnaire responses, participants were clearly in favor of head tracking. For example, when asked, *"What are your thoughts about having the head tracking technique demonstrated in this game incorporated into other first-person-shooters?"* the resulting tree, comprising of 899 responses (85% in a positive category), first split on *prior enthusiasm*, with both first-level nodes classified as a positive response. A similar result was found when asking the more contentious question: *"Would you be inclined to use this head tracking technique in more demanding situations, such as when playing on harder difficulties or in competitive online games?"* Based on the 825 responses (55% in a positive category), again the tree split on *prior enthusiasm*. Encouragingly, when examining the response of just the "skeptical" or "undecided" players (i.e. the *non-excited* players), 65% (72) were still in favor of having HAL included in other games, however, only 26% (27) would use it in a competitive scenario.

For the comparison-test, the *head tracking* condition (disabled/enabled) was consistently the most dominant factor (primary splitting variable, or best-discriminating covariate) for each of the 18 questions. For example, one of the questions asked players how much they "enjoyed" the game. It received a total of 1276 responses from 888 users (includes responses to the different study conditions and under the different game modes). The resulting tree can in Fig. 6. The tree demonstrates that the *tracking condition* is the primary predictor, with the *prior enthusiasm* being a secondary effect relevant for users with tracking enabled, with no improvement in classification for users for whom tracking was disabled.

Although these results show that players liked head tracking, it seems that a prior enthusiasm bias was present. This phenomenon was probably a result of the recruitment approach, which was based on drumming up interest in the work by posting videos on the Internet. With nodes at lower levels of the tree exhibiting high misclassification error rates, this enthusiasm effect appears to have swamped other possible interactions within the data, thus limiting the ability of the model to reveal finer structure. In other words, the preponderance of enthusiastic participants in the study meant that the covariate space was insufficiently heterogeneous for variation within the other covariates to be able to adequately explain variation in the response after the enthusiasm effect was accounted for.
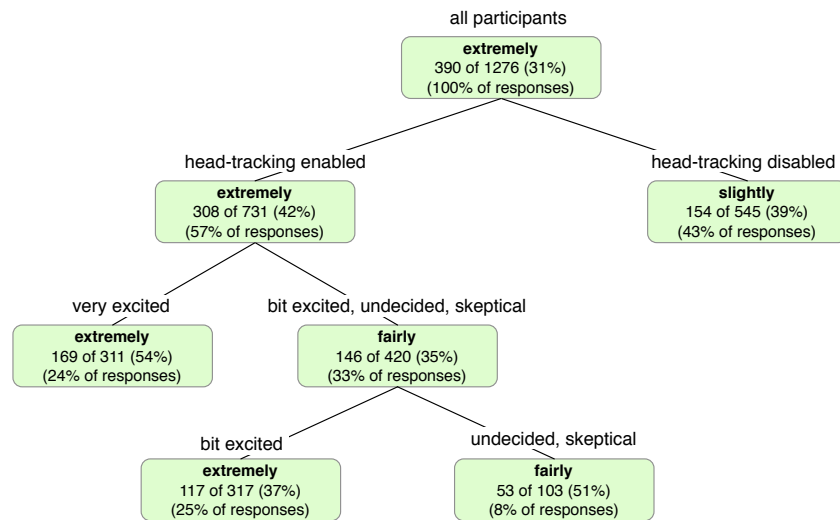
Fig. 6: Fitted decision tree model: users were asked to state how much they *enjoyed* the game experience using a five-point Likert scale ("extremely", "fairly", "moderately", "slightly", "not at all"). The tree initially splits on tracking condition (enabled/disabled), with users for whom tracking was disabled classified as enjoying it only "slightly". For users for whom tracking was enabled, predicted enjoyment depended on their reported prior enthusiasm for head tracking (categorized as "very excited" to "very skeptical"), with "excited" tracked users predicted to "extremely" enjoy it, and non-excited tracked users predicted to only "fairly" enjoy it. The details recorded at each node are the predicted classification, the proportion of data at the node consistent with that classification (x out of y, the classification at the node being the result of a majority vote) from which a classification rate can be calculated, and the proportion of the entire data at that node. The terminal nodes represent the final fit.

To further explore what people thought of HAL, we examined the written feedback collected via both the exit questionnaire and from the online discourse following the study. The comments helped identify the head tracking *benefits* that appeared to most appeal to players. Players frequently mentioned its ability to improve the game's *immersion* and *realism*. As one player stated, *"Even if it doesn't really change gameplay all that much, it certainly enhances the immersiveness of the 3D game environment, more so than higher frames per second, or higher resolution, or better shaders could hope to do at this point."* Several comments attributed these qualities to the *handy-cam* effect, feeling that the technique may be suited to horror style games, with the title *Amnesia* being repeatedly mentioned. The leaning aspect of the technique was also noted as being very *natural*; *"People have been doing this forever, leaning to the side in a game, hoping for a "better view", but this time your natural movement leads to an in-game result."*

There was strong evidence that head tracking appealed to a spectrum of PC players, including those initially skeptical about gestural control. One article about the work opined: *"Afraid that new camera-based interfaces herald the death of traditional gameplay? Fear not! This video is whispering in your ear, Core Gamer, and say-*

ing, 'Don't be afraid. Everything is going to be alright.'"[2] In another example, the popular YouTube user, Robbaz: The King of Sweden, placed a self-narrated video of himself playing the study software onto YouTube. As dedicated FPS player – as evidenced by his 300 odd videos, most of which feature FPS games – he is very praising of the technique, stating at one point, *"Oh my god. Oh that's satisfying … it feels amazing. Like the first piss in the morning."* One explanation for HAL's acceptance is the design decision to *augment* traditional controls, rather than replace them. As one player noted, *"I'd rather see more touches like this that augment controls in hardcore games than having control schemes that are all-or-nothing in the motion control department."* In another example, a gamer appealed to other players: *"...you do realise that this isn't so much motion controls so much as an addition to a traditional control scheme on the PC?… If you're a PC fan, you may actually want to check it out."*

Although the response to HAL was mostly favorable, several recurring concerns were noted. For example, several commentators feared that the use of head tracking could cause neck strain. This *perceived* issue was, however, scarcely reported by our study participants, with only three noting any *actual* strain, one of whom stated; *"My neck hurts, but in a good way!"* Although the study did not explore the use of head tracking over the very long durations that some players claim to play for, encouragingly, none of the 30 participants who played for more than an hour complained of any neck strain. Another health concern pertained to HAL's ability to induce feelings of motion sickness. As one participant stated, *"Dizziness! My main complaint is that the head tracking almost made me sick!"* This comment was reflected by a few other participants, with some specifically attributing these feelings to the handy-cam effect: *"Possibly being rid of the camera shake. It makes me dizzy fast."* Notably, these concerns were not raised in our previous focus group studies [23]. With only a small percentage of the participant sample raising the issue, its occurrence may be a result of the high probability of even a low prevalence event happening amongst a very large participant sample. The issue appears to have some precedent in the literature; with the work of [19] finding the inclusion of visual jitter can increase the realism of a scene at the cost of hastening the onset of motion sickness. The experiential benefits of the handy-cam were also noted to have a tradeoff with playing performance. One participant noted, *"I liked head tracking, it added a sense of more realism and immersion into the game. But sadly it was kind of frustrating to use since my accuracy greatly decreased."*

6   Head Tracking Guidelines

To help summarize many of the findings discussed in the previous section, we present a set of head tracking guidelines. These guidelines extend upon those presented in our earlier work [23] to reflect the insights we have gained in having examined head tracking within a real-world setting. Their intention is help other researchers with the design and creation their own head tracking techniques.

---

[2] http://www.joystiq.com/2009/07/01/a-peek-and-head-tilt-into-the-future-head-tracking-in-fps-gam/, *accessed Apr '13*

Provide Motivation and Alternatives: A game experience designed around the use of head tracking and its benefits (i.e. the ability to perform partial movements alongside traditional controls) will see its use increase (see Sec. 5.2 for more details). Despite this, potentially quite a large proportion of players (up to a quarter) will still not use head tracking, either through choice (e.g. for playing performance reasons) or circumstance (e.g. inadequate lighting), so proper consideration still needs to be given to non-gestural alternatives.

Make it Adaptable: A home setting introduces a *very large* amount of variation with regards to the tracking quality and hardware performance. The following can help make a technique more consistent: convert the user's absolute head position into one relative to a mean position, use time-based (rather than sample-size-based) logic, make the logic more conservative under less-than-desirable conditions, and clamp the head data to a desirable range (see Sec. 3).

Not All Head Movements Are Equal: Avoid using the pitch or yaw axes for critical controls as these are noisy and, at extremes, make observing the screen awkward. By contrast, horizontal and roll movements are natural to perform and are more reliably tracked (see Sec. 5.2). Experimenting with the tracking software will help you determine which movements (i.e. axes and speeds) are better tracked than others [23].

Know Your Audience: Techniques that utilize a natural movement will be better received, remembering that some movements might not be natural to all users (e.g. cocking your head to peer down the barrel of a gun - see [23]). Consider what properties attract a player to a game and whether your technique enhances these (e.g. the handy-cam effect could probably enhance the immersion of a horror game, but interferes in a performance driven game). If players are particularly skilled in using a traditional control scheme, try to avoid replacing them entirely (sec 5.3). Finally, only by observing players use your technique will you know whether its intention has been properly conveyed (sec 5.2).

Put the Health of the Player First: Avoid long awkward movements and quick snapping actions as these may place the player under duress [23], as well as reduce the perceived appeal of the game (sec 5.3). Be mindful that a technique that alters the view perspective may cause some players to feel nauseous. Consider making the player have to proactively enable a technique to use it (sec 5.3).

Guide the Player: Guide the player through the set up of the head-tracking system. Written instructions and demonstrative images may be ignored by some players, so consider having the system dynamically highlight less-than-desirable issues as they arise (sec 5.1).

7   Conclusion

The new contribution of this research is two-fold. First, we extended and refined our earlier work [23]. This development required significant effort to modify the earlier prototype techniques to make HAL robust to the highly variable environment of home use. The home-use context set a rich background for assessing HAL, raising many

challenges not found in the laboratory. These challenges led to new insights, which we have presented here. Foremost among these is the importance of the user's set-up in determining the effectiveness of the technique. Figure 4 is a simple illustration of the extent of the challenge facing widespread adoption of this technology – users, either intentionally or not, often present conditions that are almost insurmountable for reliable head tracking, and so coaching users to optimize their set-ups appears a necessary precursor to widespread use. The use of the home setting and the consequent large sample of participants also allowed us to identify a small but meaningful group of users who experience nausea while using our software, an result that did not occur during small-scale laboratory testing. The use of online recruitment and survey tools allowed us to amass a large amount of data quickly and easily, and our analyses revealed structure that showed that users were enthusiastic about head tracking and that even initially-skeptical users were swayed by its effectiveness in immersing them in the game. Head tracking did not produce an immediate performance boost – in fact, performance declined, perhaps because of users' unfamiliarity with the technique – although there was evidence of improvement with learning. This encourages us to consider a future longitudinal study to examine the long-term effectiveness of head tracking in home gaming.

References

1. Andreasen, M., Nielsen, H., Schrøder, S., and Stage, J.: What happened to remote usability testing?: an empirical study of three methods. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 1405--1414. ACM (2007)
2. Bohemia Interactive Studio: ARMA 2. http://www.arma2.com [accessed Apr '13] (2009)
3. Bradski, G.: Computer vision face tracking for use in a perceptual user interface. (1998)
4. Breiman, L., Friedman, J., Olshen, R., Stone, C., Steinberg, D., and Colla, P.: CART: Classification and regression trees. Wadsworth, Belmont (1983)
5. Brown, B., Reeves, S., and Sherwood, S.: Into the wild: Challenges and opportunities for field trial methods. In: Proceedings of the 2011 annual conference on Human factors in computing systems, pp. 1657--1666. ACM (2011)
6. Bubka, A., and Bonato, F.: Natural visual-field features enhance vection. In: Perception 39, no. 5, pp. 627--635 (2010)
7. Cooper, S., Khatib, F., Treuille, A., Barbero, J., and Lee, J.: Predicting protein structures with a multiplayer online game. In: Nature 466, vol. 7307, pp. 756--760 (2010).
8. Costanza, E., Giaccone, M., Küng, O., Shelley, S., and Huang, J.: Ubicomp to the masses: a large-scale study of two tangible interfaces for download. In: Proceedings of the 12th ACM international conference on Ubiquitous computing, pp. 173--182. ACM (2010)
9. De Barros, P. G., Lindeman, R. W., and Loughlin, T. J.: Characterizing Head Movement in First-Person Games. Poster, ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (2009).

10. Evans, J. R., and Mathur, A.: The value of online surveys. Internet Research 15, no. 2, pp. 195--219 (2005)

11. Gkikas, K., Nathanael, D., and Marmaras, N.: The evolution of FPS game controllers: How use progressively shaped their present design. In: 11th Panhellenic Conference on Informatics, pp. 18--20 (2007)

12. Gorodnichy, D., Malik, S., and Roth, G.: Nouse 'Use your nose as a mouse'-a new technology for hands-free games and interfaces. In: Proc. Intern. Conf. on Vision Interface, pp. 354--361 (2002)

13. Hartson, H., Castillo, J., Kelso, J., and Neale, W.: Remote evaluation: the network as an extension of the usability laboratory. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 228--235. ACM (1996)

14. Kittur, A., Chi, E. H., and Suh, B.: Crowdsourcing user studies with Mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 453-456. ACM (2008)

15. Lombard, M., and Ditton, T.: At the heart of it all: The concept of presence. Journal of Computer-Mediated Communication 3, no. 2 (1997)

16. Microsoft Corporation: Kinect. http://xbox.com/kinect [accessed Apr '13] (2010)

17. Murphy-Chutorian, E., and Trivedi, M. M.: Head Pose Estimation in Computer Vision: A Survey. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on 31, no. 4, pp. 607--626 (2009)

18. NaturalPoint Inc.: TrackIR. http://www.naturalpoint.com/trackir [accessed Apr '13] (2001)

19. Oman, C. M.: Motion sickness: a synthesis and evaluation of the sensory conflict theory. Canadian Journal of Physiology and Pharmacology, 68, no. 2, pp. 294--303 (1990)

20. Palmisano, S., Gillam, B. J., and Blackburn, S. G.: Global-perspective jitter improves vection in central vision. Perception 29, no. 1, pp. 57--67 (2000)

21. Robinson, K. M.: Unsolicited Narratives from the Internet: A Rich Source of Qualitative Data. Qualitative Health Research 11, no. 5, pp. 706--714 (2001)

22. Seeing Machines: faceAPI. http://www.faceapi.com [accessed Jan 2013] (2001)

23. Sko, T., and Gardner, H. J.: In: Human-Computer Interaction–INTERACT 2009, pp. 342--355. Springer, Heidelberg (2009)

24. Sko, T. Gardner, H. J. and Martin, M.: Non-parametric Decision Trees for Online HCI. To be presented at: The 2013 ACM annual conference on Human Factors in Computing Systems, ACM (2013)

25. Stellmach, S., and Dachselt, R.: Designing gaze-based user interfaces for steering in virtual environments. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 131--138. ACM (2012)

26. Sutherland, I.: A head-mounted three dimensional display. In: Proceedings of the fall joint computer conference, part I, pp. 757-764. ACM (1968)

27. Wang, S., Xiong, X., Xu, Y., Wang, C., Zhang, W., Dai, X., and Zhang, D.: Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 1097--1106 (2006)

28. Willow Garage: OpenCV. http://opencv.willowgarage.com [accessed Apr '13] (2000)

29. Wood, R., Griffiths, M., and Eatough, V.: Online data collection from video game players: Methodological issues. CyberPsychology & Behavior 7, no. 5, pp. 511--518 (2004)