Perceptual Thresholds for Foot Slipping in Animated Characters

by

Jordan J. Strawn

A PROJECT

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Computer Science (Honors Associate)

Presented June 1, 2006
Commencement June 18, 2006

AN ABSTRACT OF THE THESIS OF

Jordan J. Strawn for the degree of Honors Baccalaureate of Science in Computer Science presented on June 1, 2006. Title: Perceptual Thresholds for Foot Slipping in Animated Characters.

Abstract approved: ____
Ronald Metoyer

The computer game industry continues to progress toward realistic-looking character motion. However, even in state-of-the-art games, the use of motion capture data in character animation may result in errors such as “foot slipping,” where the feet do not match up with the floor properly during translation. Various algorithms have been proposed to minimize foot slipping, including one which changes limb lengths. While foot slipping decreases the realism of character motion, there must be some threshold below which this error is imperceptible; devoting further processor time in these cases is wasteful. We apply the classical method of perception threshold determination using a set of motion clips with parameterized slipping error. From this experiment, we develop guidelines for acceptable error. Furthermore, we show that introducing simple camera motion may increase the perceptual threshold, and thus could be used to “mask” foot slipping errors.

Keywords: visual perception, human character animation, motion capture processing

© Copyright by Jordan J. Strawn
June 1, 2006
All Rights Reserved

Perceptual Thresholds for Foot Slipping in Animated Characters

by

Jordan J. Strawn

A PROJECT

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Computer Science (Honors Associate)

Presented June 1, 2006
Commencement June 18, 2006

Honors Baccalaureate of Science in Computer Science project of Jordan J. Strawn presented on June 1, 2006.

APPROVED:

Mentor, representing Computer Science

Committee Member, representing Psychology

Committee Member, representing industry

Director, School of Electrical Engineering and Computer Science

Dean, University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, University Honors College. My signature below authorizes release of my project to any reader upon request.

Jordan J. Strawn, Author

ACKNOWLEDGEMENTS

I am deeply grateful to Dr. Metoyer for his insights and confidence in this study, and for making time for our enjoyable weekly meetings. We both owe thanks to Dr. Lien for her instruction in the fundamentals of perception research, and to her and her graduate students for feedback on the implementation in its early stages. I also appreciate Frank Cloutier's willingness to serve on my defense committee, as a representative from industry and mentor.

Bryce Hutchings graciously provided the code library for generating the test motion, and Madhu Srinivasan came to our aid several times. Jessica Hodgins (Carnegie Mellon) and Michiel van de Panne (University of British Columbia) provided early direction, on how exactly to go about studying foot slipping with camera motion.

I am grateful to my wife, Kellie, for her patient encouragement, and for showing last year that surviving the Honors thesis is possible. I would also like to thank my parents, W. Gregg and Pam Strawn, for participating in the pilot study and for many years of love and support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

PERCEPTUAL THRESHOLDS FOR FOOT SLIPPING IN ANIMATED CHARACTERS

1. Introduction

With ever-faster processors and evolving graphics hardware, the state of computer games has improved at an astonishing rate over the past decade. Imagery in games continues to move closer to the goal of photorealism, as more polygons can be rendered in each frame with complex shaders afforded by recent hardware advances. At the same time, research into motion animation continues to produce more realistic, more human action sequences for interactive entertainment. In most cases, motion capture data is blended to produce smooth transitions between motion clips. Like Hollywood, the computer game industry is based on the concept of “suspension of disbelief,” and this goal keeps getting closer.

However, even in state-of-the-art games, graphics are far from perfect. Processor speed remains a limiting factor in image quality. Similarly, animated motion is somewhat limited by the richness of the available motion capture data. Transitions between captured sequences can create motion discontinuities, or jumps in joint angle values. Footskate, or foot slipping, is another common occurrence in games, in which a character’s feet do not match up with the ground during translation (for example, running).

The branch of psychology dealing with visual perception has recently been applied to computer graphics. Studies describing the limits of the human visual system are useful in producing better animated motion. In short, if a human observer is likely to notice something, it should look good; if not, then computation should be focused elsewhere. This brings the related psychological field of attention into the picture.

In the tradition of recent studies combining computer graphics and research methods from psychology, we have investigated perceptual thresholds for foot slipping with simple animated motion. In particular, we hoped to find a relationship between camera motion and perception of foot slipping. We hypothesize that camera motion may be useful in masking fairly large animation errors.

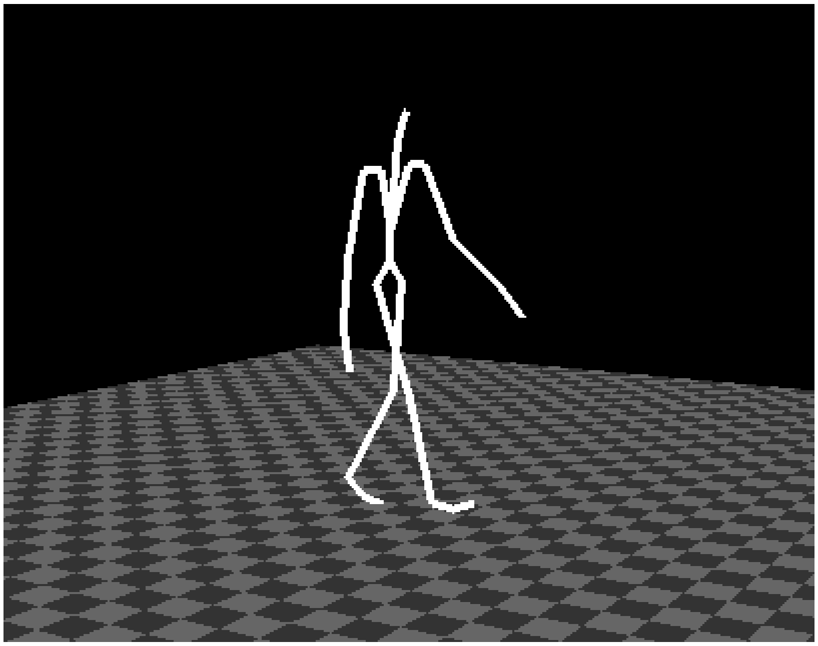
2. Background

2.1 *Motion Capture Data*

Two primary techniques are used in producing data for computer-animated human motion: key-framing and motion capture. With key-framing, the position of a character at certain "key frames" is specified by hand by an animator, and motion is produced by interpolating body positions in the intermediate frames ("tweening"). The motion capture process involves outfitting actual human actors on a stage with expensive equipment that records the body position at each frame of motion. This approach is commonly used for developing motion suites in contemporary games.

Motion capture data, in the .bvh ("Biovision Hierarchical") file format, is based on an hierarchical structure of a character [Thingvold 1999]. The root node of the character hierarchy, typically at the pelvis, defines the global position of the character. From this root, child nodes containing offset and frame information specify protruding bones (e.g. the right and left thighs) (see Figure 1). A pose, then, consists of the complete specification of the position of each child node in relation to each parent, and the global position of the root node [Parent 2002]. In the motion capture session, markers are placed at strategic points on the actor's body. The global positions of these markers, constituting an actor's pose, are monitored at each frame. This raw marker data can be subsequently processed into hierarchy and joint angle information, which is often stored in the .bvh format. This data format can then be subsequently read as input, and played back in real-time.

FIGURE 1. Example of Hierarchical Skeleton Figure. The pelvis is the root node, and the hands, feet and head are leaves in the hierarchy.



In the context of a game, the best available data that fits the character’s current motion is applied and rendered in real-time. Recognizing the need for post-capture motion editing, Witkin and Popovic described a method for motion warping, by key-framing certain joints in a motion sequence, to adapt existing data to new situations [1995]. Transitions between motions can be produced through interpolation [Bruderlin and Williams 1995], and new motions can even be generated through inter-mixing existing data. For example, Ikemoto and Forsyth present an off-line technique for transplanting body parts between motion sequences to produce new combinations of motions [2004]. Unfortunately, motion editing frequently introduces undesirable side effects into the resulting motions. Limited by the suite of motion capture data, the animation in games is often noticeably unrealistic.

2.2 *Implications from Psychology*

With a perfect visual system, human viewers would require perfect motion for believable animation. However, studies in visual perception have consistently shown that the human visual system is more limited then we may imagine. A primary limiting factor is attention. This established theory in psychology provides the basis for perceptually-based rendering.

It is estimated that the optic nerve accounts for 38% of the nerve fibers which connect to the central nervous system. Perhaps even more astounding, scientists estimate that approximately 75% of the information that enters the brain is visual [Hanson 1987 in Berger 1998]. However, the process of sensation and perception involves data filtering, as a person only consciously perceives a portion of what he or she physically “sees” [Simons and Chabris 1999]. Visual perception, then, is subject to both high-level (conscious) goals, such as targets in a game, and certain low-level (unconscious) stimuli [James 1891 in Franconeri and Simons 2003].

The salience of both human forms and motion in the visual system helps explain why observers frequently notice foot slipping error in games. Furthermore, Hodgins et al.

note that “few movements are as familiar and recognizable as human walking and running” [1998]. However, we must balance the experiential belief that foot slipping is easily noticeable in games with the well-supported understanding that human vision is highly limited. That is, there must be some threshold below which foot slipping consistently goes unnoticed, and is thus acceptable in animated motion.

2.3 *Perceptually-Based Rendering*

Given the limits of the human visual system, the technique of perceptually-based rendering has been proposed as a solution to the problem of computationally expensive graphics. Perceptually-based rendering focuses resources on regions of the display that are most likely to capture the observer’s attention. At a high level, this approach is logical, but its implementation requires the ability to predict both attention and the perceptual cost of rendering errors [Horvitz and Lengyel 1997].

Predicting attention may seem to be a difficult problem, but there are known factors that contribute to which objects in a scene will capture attention. As noted above, high-level goals and low-level salience should be considered. Researchers have already demonstrated success in predicting attention in some limited cases. For example, a computer model has been developed that predicts up to 85% of visual fixations in an observer searching for people in a scene [Oliva et al. 2003 in Harrison et al. 2004]. One difficulty is that eyes move constantly to survey a scene [Ornstein 1972 in Berger 1998], but such a model could at least predict which objects should have *low* rendering priority.

With the goal of directing rendering where it is most valuable, perceptual thresholds serve as guidelines for the point of diminishing returns. In other words, just as emphasis is placed where attention is most likely, there is no reason to devote processor time when attention and detection are unlikely.

Perceptually-based rendering is already commonly applied through the technique of levels of detail (LOD) [e.g. Lindstrom et al. 1996]. Essentially, the geometry of a terrain, object or character is refined as it approaches the viewer, and thus becomes more prominent in the display. Characters or objects in the background may be rendered quite simply, with the justification that they will receive minimal attention. Sattler et al. investigate perceptual thresholds for shadow approximation with fewer polygons, since shadow rendering is computationally intensive [2005]. They find that 90% of observers cannot tell the difference between the shadow of an object, and an approximation simplified to 1% of the original triangles.

LOD could be applied to animated motion, as well. Just as objects far from the viewer or insignificant to game-play may be rendered more simplistically, the motion generation could be simplified as well, to focus resources on prominent motion in the foreground. Recent studies have investigated perceptual thresholds for common types of error in character animation, as guidelines for acceptable error [e.g. Reistma and Pollard 2003]. Quality of motion has vast implications, as Oesker et al. show that unconscious processing of levels of detail in motion can affect high-level judgments about characters, such as their skill level in soccer [2000]. O’Sullivan et al. suggest that levels of detail can be useful in crowd simulation, as applied to character geometry, motion, and even behavior [2002].

2.4 *Origins of Foot Slipping*

One common problem with animated motion that persists in state-of-the-art games is footskate, or foot slipping. This is where the feet of a character do not plant cleanly on the ground, but rather slip as though the character is on a slick surface (compare Figures 2 and 3). Foot slipping is an artifact of the motion capture/playback process, with various causes. When the raw motion data is processed, the calculated joint angle solutions can produce secondary motion at the extremities. In a particularly obvious form, the feet may slip on the ground when the character is rotated about its root node. It is also common that the best motion capture sequence for an in-game motion was captured at a different translational velocity than that desired. For example, the actor for a walking sequence walked at velocity $v0$, but the character in the game is walking at some $v' < v0$. Because the joint motion is the same, but the velocity of the root node must be modified, the feet will not match up properly with the ground, and the resulting animated motion looks unnatural. In this case, the foot slips backward during each step, rather than planting firmly on the ground. With the goal of making games look more realistic, it is desirable to eliminate such artifacts. However, these improvements come at a computational cost.

FIGURE 2. Sample Frames from Motion-Captured Normal Walking (Frames 40, 45, 50, 55, 60). Note how the right toes plant on the floor marker for the duration of a step, in the absence of slipping.

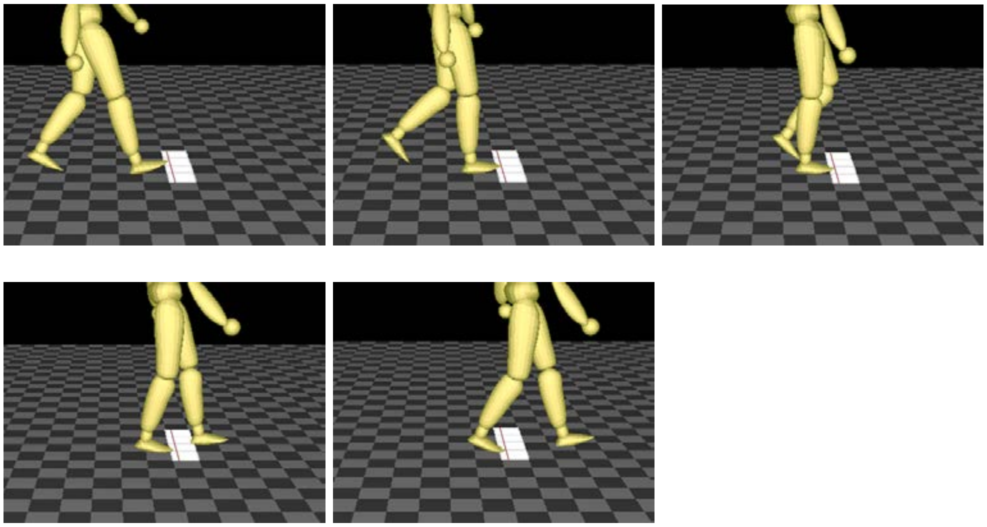
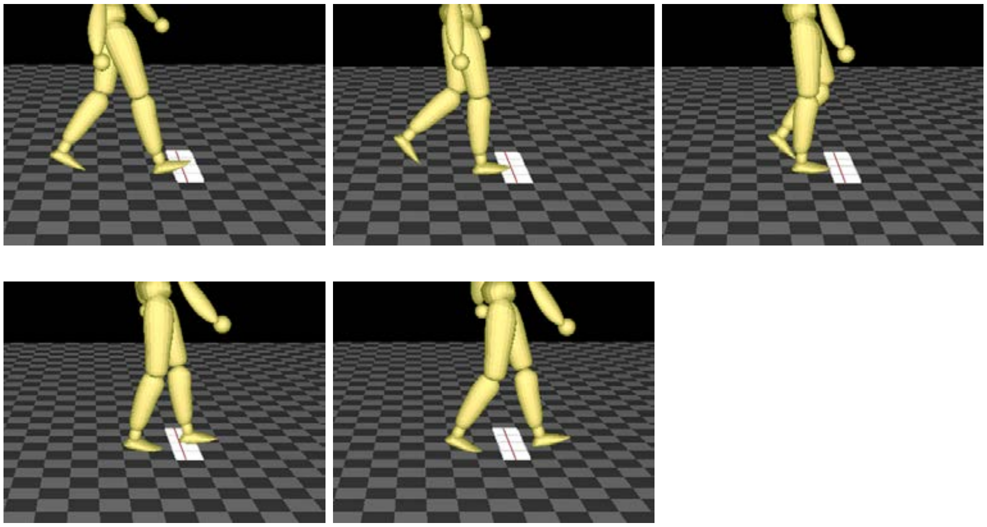


FIGURE 3. Sample Frames from Motion-Captured Walking with 40% Foot Slipping Error introduced. Because the root velocity has been decreased by 40%, the right foot slides backward on the floor during the step.



2.5 *Proposed Solution to Eliminate Foot Slipping*

Kovar et al. present an algorithm for cleaning up foot slipping, or “footskate,” which effectively eliminates the phenomenon [2002]. The processing algorithm selects a fixed location on the ground for each foot-fall of the motion. It then modifies the motion so that in the frames surrounding a foot-fall, the corresponding foot is fixed at that location. This is accomplished by adjusting the length of the leg. Their work is a departure from previous solutions, which assume a rigid skeleton [Kovar et al. 2002]. They report that the cost of such linear processing in real-time applications is only a 0.5-1.0 second delay, which is justifiable for characters in the background of an interactive environment.

However, this solution for eliminating foot slipping introduces a new type of error: limb length change. To characterize this trade-off, Harrison et al. subsequently investigated perceptual thresholds for length changes in a rotating two-dimensional arm [2004]. Their findings echo earlier work in visual perception: the human visual system is truly quite limited, particularly by attention. With fixed attention, expectant observers began perceiving length change at approximately 2.7%. Naïve observers produced a threshold of 5.6%. Perception degrades significantly in the absence of attention; the study suggests a possible worst-case threshold near 20%. These numbers suggest that Kovar et al.’s solution of introducing limb length error is defensible in many cases.

Although Harrison et al. present useful guidelines for length changes in animated motion, there is no available data regarding the perception of foot slipping itself. In cases where foot slipping is an obvious detractor to the realism of the scene, Kovar et al.'s footskate cleanup algorithm merits consideration. However, there is no need to clean up faulty motion when it is below the perceptual threshold of the viewer. With perceptually-based rendering, the problem of foot slipping could best be tackled with numeric guidelines for both acceptable limb-length changes and foot slipping error. In their closing remarks, Kovar et al. mention this need explicitly:

> "…This leads to the interesting general question of how various artifacts in a motion – footskate, over-stretched limbs, sudden changes in joint orientation, etc. – may be balanced so as to produce a desired change while minimizing visual disturbance" [2002].

This is a question of optimization which requires numeric limits on acceptable motion error. Through careful construction of psychology experiments, perceptual thresholds can be determined statistically for these noted artifacts.

2.6 *Goals of the Study*

In this study, we add to the existing body of knowledge in visual perception of animated motion. We designed an experiment to establish perceptual thresholds for foot slipping, both with and without camera motion. Such findings could be used to establish guidelines for motion processing in future applications, for the optimal usage of computational resources. That is, numeric guidelines are useful in eliminating perceptible error, without wasting processor time in fixing the imperceptible errors.

3. Methodology

3.1 *Method of Limits*

The perceptual threshold for some stimulus or event is the point at which a human observer will detect the event 50% of the time. Naturally, the probability of detection varies with the magnitude of the stimulus, such that an observer is less likely to detect the stimulus as its magnitude approaches 0, and is more likely to detect it as the magnitude increases. Thus, if the magnitude of the stimulus is initialized to a small value and gradually incremented, the magnitude at which the observer first detects the stimulus is an estimate of the perceptual threshold. This is the basis for Fechner's classical method of determining perceptual thresholds, called the "Method of Limits" [Proctor and Van Zandt 1994].

3.2 *The Stimulus*

In this study, the stimulus is foot slipping. Foot slipping often occurs in animated motion as a result of playing back motion capture data with a root velocity different from that of the motion capture actor during the capture session. That is, the locomotion is captured at some average velocity $v0$, and is subsequently displayed at a different average velocity v'. We can define the magnitude of the slipping as an error percentage:

$$\%error = \left|(v0 - v')/v0\right| \times 100\%$$

Characterizing foot slipping error as a percentage of the captured velocity is desirable, because the study results are therefore independent of the sample motion used. Results can be applied generally to motion data with some arbitrary velocity. This follows the concept of a Weber fraction, which “capture[s] the relationship between the size of the estimation error and the size of the stimulus, which is expressed as a proportion” [Harrison et al. 2004].

It should be noted that we generated foot slipping by playing back motion capture data with a decreased root velocity $\left(v' < v0\right)$. However, it is easy to imagine cases where applying motion capture data would require increasing the root velocity in play-back $\left(v' > v0\right)$. This would produce foot slipping as well; it is not immediately apparent whether perceptual thresholds would be similar between the two cases.

3.3 *Trial Construction*

As mentioned above, Harrison et al. demonstrate that expectation can affect perceptual thresholds: if an observer knows what to look for, resulting perceptual thresholds are lower [2004]. In interactive environments, foot slipping is unexpected (and frequently undetected). We wished to maintain naïve observers, by avoiding asking about foot slipping explicitly. We accomplished this by measuring a difference threshold, between “good” motion and motion with foot slipping error. Each trial consisted of the presentation of two short motion clips, one containing the original captured motion, and one with parameterized slipping error introduced. In each trial, participants were asked whether the character motion in the two clips was the same or different. If the participant answered “different,” he/she detected the stimulus, the slipping present in one of the two clips. The order of the motion clips within each trial (original motion and slipping motion) was randomized to minimize learning.

3.4 *Block Construction*

In the method of limits, the magnitude of the stimulus is incremented or decremented by a constant step, until the observer changes his/her answer, thereby crossing the threshold. The method specifies that the threshold should be approached alternately from above and below [Proctor and Van Zandt 1994].

A block of trials consists of the trials necessary to reach the threshold once. That is, in an increasing block, the stimulus magnitude is initialized well below the threshold estimate to-date, and gradually increased, until the observer detects that the two motion clips are different. Conversely, in a decreasing block, the magnitude is initialized above the estimated threshold, and gradually decreased, until the observer can no longer discriminate between the two presentations.

3.5 *Threshold Estimation*

Each block, increasing or decreasing, provides one estimate of the perceptual threshold for the presence of slipping. The current magnitude of the slipping error is saved each time the threshold is reached. With a sufficient number of blocks (over a range of participants), an overall estimate of the perceptual threshold can be calculated by averaging the results, with some statistical level of confidence, assuming a reasonably normal distribution.

3.6 *Implementation*

We developed a simple web-driven interface in PHP to lead each participant through the blocks of trials [PHP Group 2006]. As much as possible, generality was maintained, so that the interface implements the method of limits with two video presentations per trial. Therefore, the video files could be replaced and the source files easily modified to implement a similar study in the future.

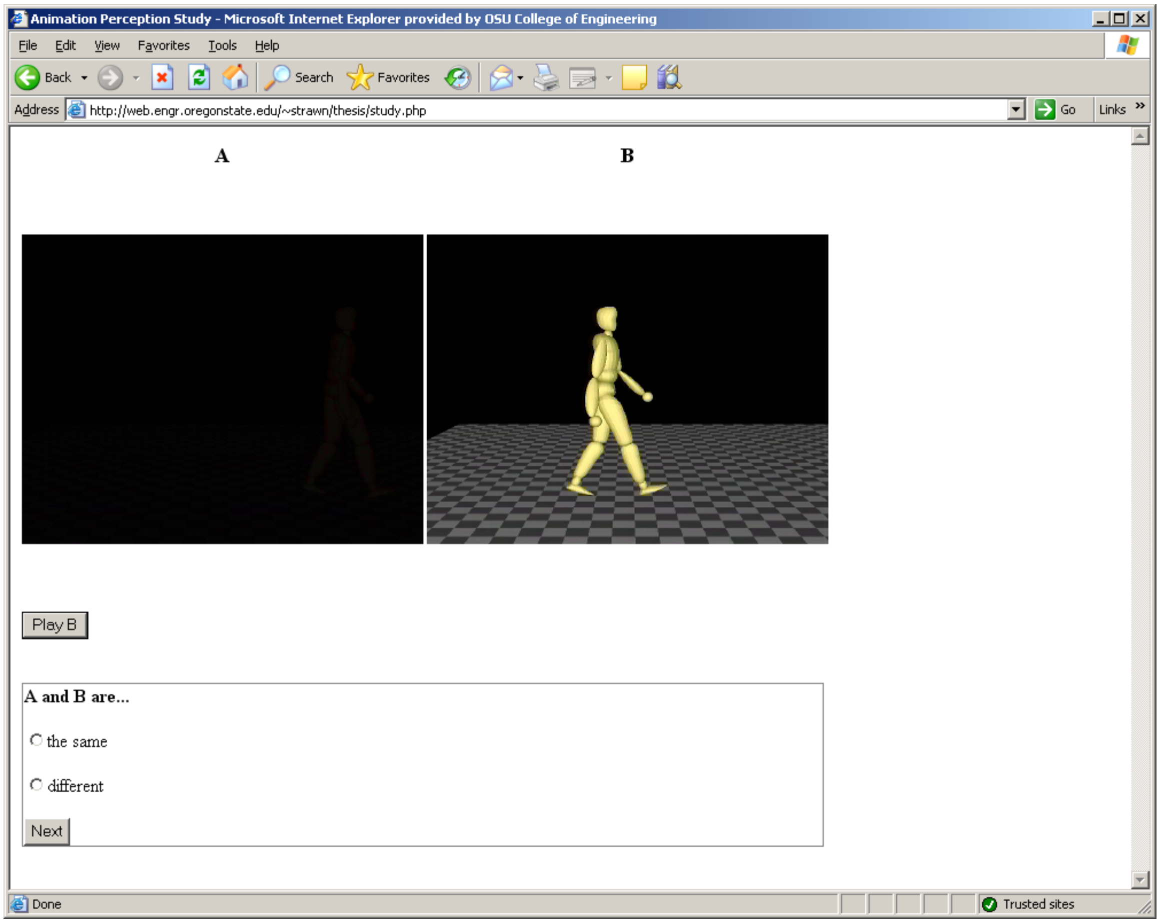
For this study, there were only two thresholds to measure: slipping with a stationary camera and slipping with simple camera motion. Eight blocks of trials (ascending and descending, alternately) were allocated to each of these cases, for a total of sixteen blocks per participant.

After the participant logs in with a given numeric ID, he/she is presented with instructions and a practice trial (stationary camera, with 0% and 40% error). The participant is explicitly instructed to “watch the motion of the character’s feet in each movie, and determine if the two movies are the same or different. Do not look for differences in the background or floor…” The practice trial is intended to present the observer with two clearly different motion clips, so that he/she may become familiar with the format of the trials and the motion itself before actual data collection begins.

After completion of the practice trial, and feedback with the correct answer for the practice trial, the interface then proceeds with eight blocks of trials with a stationary camera. The first block initializes the magnitude of error in the first trial to $\left(0+STEP\right)$, where STEP is the change in magnitude between consecutive trials. In other words, the first trial portrays a motion clip with no slipping error and a motion clip with the minimum error greater than 0.

The participant clicks the “Play A” button to start the first motion clip. Once it has completed, the participant clicks “Play B” to watch the second clip. He/she must then answer the question “A and B are: i) the same; ii) different.” Within each trial, simple JavaScript error checking ensures that the participant has watched both movies and has answered the question, before continuing. The participant does not have the option of replaying movies.

FIGURE 4. Screen Shot of Experiment Interface. The observer has already watched motion clip “A,” and is now watching clip “B.”



A and B are...

○ the same

○ different

In an increasing block, if the participant chooses “the same,” the error magnitude is increased by STEP in the next trial, and the process is repeated. When the participant detects the stimulus and selects “different,” the error magnitude and information about the block are inserted into a MySQL table [MySQL AB 2006], with each row representing one estimate of the perceptual threshold x_i. In the event that the observer fails to detect the difference with the maximum error presented, the block is logged as an error condition.

The interface then reverses direction and begins a descending block. The magnitude of the error for the next trial is initialized randomly on the interval

$$\min\left(rand\left(3,4,5\right)*STEP+x_{i-1}, MAX\_ERROR\right),$$

to ensure that the stimulus magnitude in the next trial is well above the threshold that was just produced. The interface proceeds with decreasing magnitude, storing the threshold and reversing direction when the threshold is reached from above. The error magnitude initialization is similar for subsequent ascending blocks:

$$\max\left(x_{i-1}-rand\left(3,4,5\right)*STEP, MIN\_ERROR\right).$$

After eight blocks of trials with a stationary camera, the experimental interface presents eight blocks of trials with identical character motion, and simple camera motion introduced.

Because the test movies could not be generated in real-time, the experimental constraints STEP, MIN_ERROR and MAX_ERROR were defined at experiment implementation, based on the results of an informal pilot study. Ideally, the STEP value of the error between consecutive trials should generate a just noticeable difference in the stimulus between trials. The optimal value should allow the observer to reach the threshold within several trials each block; with too large a STEP, resulting data lacks precision, and with too small a STEP, the observer has to proceed through endless indistinguishable trials.

Naturally, the MIN and MAX magnitudes of the stimulus should be substantially displaced from the threshold, so that the observer will consistently reach the threshold prior to reaching MIN or MAX. This implies that MIN and MAX may vary in different cases, as a higher threshold would require a higher MAX value, for example.

Based on pilot study results, we chose different (MIN-MAX) ranges for the cases of stationary camera and moving camera. To optimize the STEP value as described above, we chose different STEP values for the two cases as well, by a factor of 2. This means that the precision for the estimates of perceptual threshold is off by a factor of 2 as well, but the data for each case can still be statistically analyzed.

FIGURE 5. Parameters for Implementation of Method of Limits.

	ERROR STEP	MIN_ERROR	MAX_ERROR
Stationary camera	5%	5%	50%
Moving camera	10%	10%	100%

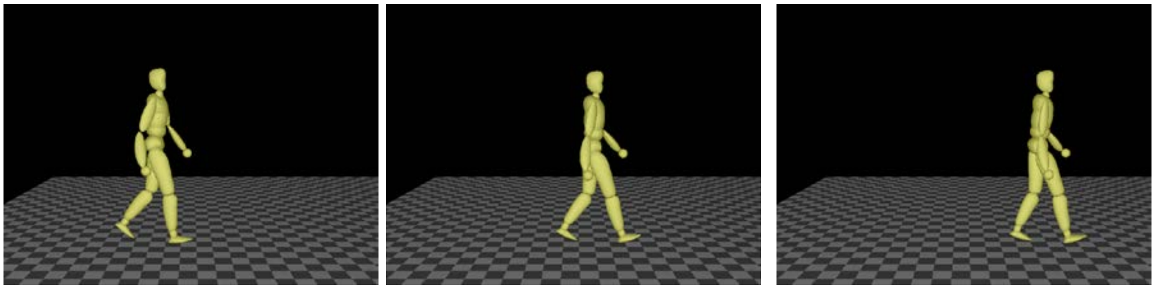
3.7 *Movie Generation*

The stimulus could not be generated dynamically in this experiment, due to the complexity of the motion and the time required to generate a movie playable in a browser. A single .bvh (motion capture data) file was used as the source of animated human motion for every movie. In the 75-frame sequence (at 30 frames per second), an adult male walks along a straight line at approximately 0.96 m/sec. We used a C++ library for playback and editing of the motion sequence using OpenGL [Hutchings 2006].

Hodgins et al. show that the structure of an animated human body affects perceptual thresholds for motion: in their experiment, observers were better able to distinguish between similar motion clips with a rigid-body character, than with a simple stick figure [1998]. Therefore, the single character in this study was rendered with rigid-body, ellipsoid limbs, chest, hips, etc. To direct attention to the character, the background was simply black; a checker-board floor was used to provide the observer with a reference to more easily detect foot slipping. Participants were explicitly instructed to not look for changes in the background or floor between clips in the trial, to direct attention and prevent unexpected results from rendering artifacts.

In the simple case, the character walks from the left side to the right side of the window; the camera is stationary. Foot slipping was produced by varying the root velocity of the character. This means that the character's steps were exactly the same in each clip, so as the translational velocity was decreased, slipping increased. This implementation introduces the uncertainty of differences between clip presentations of root velocity: the observer may not perceive slipping, but may notice a difference in distance traveled between the two clips. By instructing the observer to look at the feet, we attempted to mitigate this problem.

FIGURE 6. Frames of Animation Sequence with Stationary Camera. Frames 2, 39, 75 from the walking sequence with 50% slipping error.

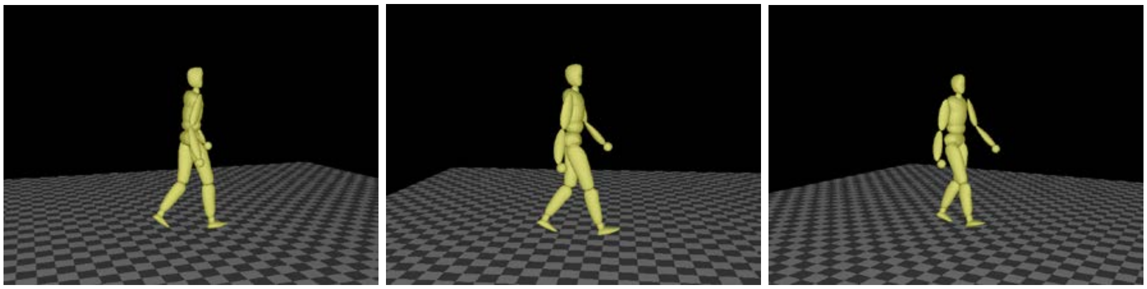


The camera field of view was set to 35% (with a 4:3 aspect ratio) to make the character take up the majority of the display. The camera was placed approximately 4.25m from the character, perpendicular to the path of the motion.

In the second case, we maintain the character motion described above, but introduce simple camera motion. The camera translates to its right, on a line parallel to the character’s motion, and adjusts its orientation each frame to remain centered on the character (Figure 7). The camera translates at a velocity of 2.1 m/sec, which is intentionally low, to minimize disorientation. Also, we chose to keep the camera centered on the character, to aid the observer in directing attention to the feet.

The selected camera motion is fairly representative of third-person sports games, in which the camera may follow the ball or a key player. For example, in football, the camera may translate right as the offense runs an “option-right.” The camera motion used is also similar to “strafing” in a first-person shooter (where the camera is from the point of view of the player’s eyes). In this case, the player is moving side-ways, and aiming his/her weapon and attention at another character.

FIGURE 7. Frames of Animation Sequence with Moving Camera. Frames 2, 39, 75 from the walking sequence with 50% slipping error. Note how the character and floor rotate as a result of the camera motion, although the character walks in a straight line.



The foot slipping code was instrumented to dump each frame to a bitmap file. These frames were then reassembled in Adobe Premiere, and exported in QuickTime format, for playing in a browser. Each motion clip begins and ends with a blank frame, so that the observer cannot compare the initial or final frames of the two presentations.

3.8 *Data Collection*

No specific population was targeted in recruitment. Participants volunteered for a 30- to 45-minute session, one at a time, in the Graphics and Imaging Technologies Lab on the campus of Oregon State University. After informed consent was obtained, the participant was seated in front of a flat-screen monitor, and instructed to log-in with a given numeric ID and follow the instructions on-screen. The web interface led the participant through the duration of the experiment. In all, 21 observers participated in the experiment. As described above, each participant was naïve to the specifics of the study and the focus on foot slipping.

At the conclusion of the experimental trials, the participant filled out a web form questionnaire, eliciting demographic information and prior experience with computer animation in games and film. The primary purpose of the survey was to determine any factors which may have explained departures from a normal distribution in the resulting data, in the event of data irregularities. The format of the questionnaire, and a summary of the results, is included in Appendix B.

4. Results

4.1 *Statistical Analysis*

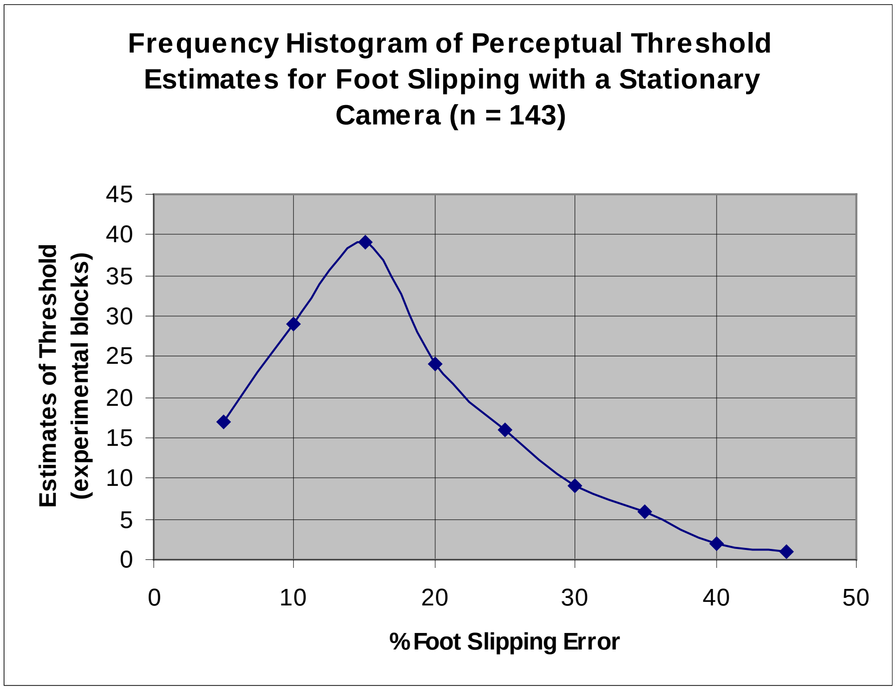
Twenty-one observers participated in the experiment, proceeding through eight blocks for each of the two cases: i) slipping with a stationary camera, and ii) slipping with a moving camera. This ideally produces n = 168 estimates of the perceptual threshold for each case. However, because of the static set of motion clips, the method sometimes failed to find an estimate of the perceptual threshold in a trial. For example, in a decreasing block with a stationary camera, an observer may have answered that the 0% error and 5% error clips were different. In this case, there was no motion clip available with 0% < error < 5%.

Surprisingly, in the case with the moving camera, observers also occasionally failed to detect a difference between 0% and 100% slipping error. However, both of these out-of-range problems combined occurred in only twelve of the 336 experimental blocks performed. Such blocks were logged as exceptions and omitted from statistical analysis.

Additionally, three observers with the highest and lowest average thresholds were omitted from analysis, for the sake of minimizing variance (see Appendix A for results by participant). Two of these participants were students with extensive experience with graphics environments, whose prior experience and conceptual understanding may have otherwise biased the results. Notably, these three participants also accounted for seven of the twelve out-of-range errors in the entire study. Since the data from two participants with exceptionally high thresholds were omitted and only one with low thresholds was omitted, if anything, this should bias the estimates downward (to a more conservative estimate of the threshold for each case).

With the guidelines noted, the case of a stationary camera produced a perceptual threshold for foot slipping of $\overline{y}$ = 17.1%. Applying a t-test, the 95% confidence interval for this threshold is ±1.4%, based on data from n= 143 experimental blocks. The data follows a slightly right-tailed normal distribution. This estimate for the perceptual threshold means that the average observer would note a difference (the slipping error) between regular walking and walking at $(1 - 0.171)v_0 = 0.829\ v_0$ fifty percent of the time.

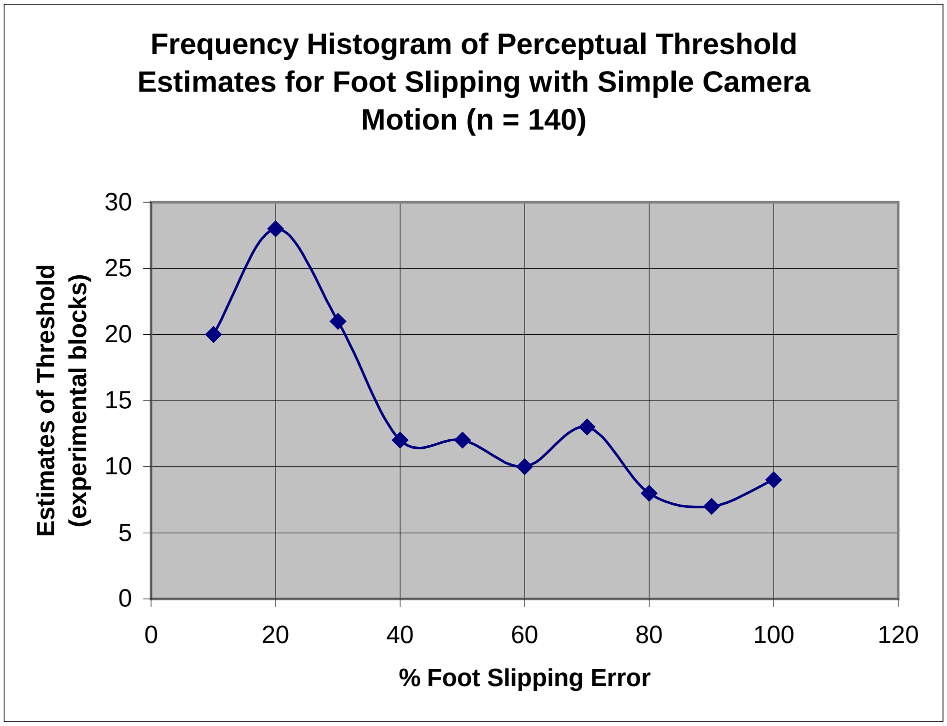
FIGURE 8. Frequency Histogram of Perceptual Threshold Estimates for Foot Slipping with a Stationary Camera.



Introducing camera motion produced a higher average threshold for foot slipping, but it also dramatically increased the variability of the data (Figure 9). The data presents a strongly right-tailed distribution, with outlying values well above the estimated threshold. In this case, the same walking motion produced an average perceptual threshold of $\overline{y}$ = 43.9%, with a 95% confidence interval of ± 4.7%, assuming use of the t-test is valid. Data from n = 140 experimental blocks was used in this case. Based on this analysis, with a moving camera, an ideal observer would only note a difference fifty percent of the time between regular walking and walking at 0.561 times the original velocity. That is, a character's velocity can be nearly halved without consistent detection of slipping, if the camera is moving at 2.1 m/sec.

However, the distribution in Figure 9 features local maximums at 20% and 70%, which implies unforeseen factors may have contributed to a problem with the sample. As displayed in the histogram, some observers may frequently notice the slipping error at a magnitude of around 20%. This phenomenon of highly "skilled subjects" has been documented in a related perception study [Hodgins et al. 1998]. One participant even produced a slightly lower perceptual threshold with the moving camera. On the other hand, many participants noted at the conclusion of the trials, that the clips with a moving camera are much harder to differentiate. The number of experimental blocks which produced a threshold of 50% or higher reinforce this testimonial evidence.

FIGURE 9. Frequency Histogram of Perceptual Threshold Estimates for Foot Slipping with Simple Camera Motion.



Perhaps the best explanation for these two local maxima (at 20% and 70%) is the effect of learning. From the first eight blocks of trials with a stationary camera, one group of observers may have become familiar with the foot slipping phenomenon, and thus knew what to look for in subsequent blocks. Others may not have determined the exact difference between the pairs of motion clips, and thus had greater difficulty in distinguishing between slipping and normal walking once camera motion was introduced. If these two local maxima are considered to be two distinct samples, the third local maximum at 50% can be explained as an additive effect of the intersection of the two samples. To test this hypothesis, we could easily run naïve observers through only the second portion of the study (with camera motion), without the chance to first become familiar with the type of error studied with a stationary camera. If these results established an average threshold around 70%, the data would provide strong evidence for learning in some observers.

A second possible cause for the variability of the data with a moving camera is added ambiguity. At the outset of data collection, participants were instructed not to look for changes in the floor. Aside from potential random rendering artifacts, the floor does not change in the presentations with a stationary camera. Therefore, if the observer notes a difference in the motion of the character's feet, it must be due to a change in character motion.

However, in blocks with a moving camera, the instructions could be ambiguous, because the orientation of the floor appears to be changing. In fact, several participants classified these trials as "the ones with the moving floor," although the camera, or equivalently, the entire scene, was in fact moving. Foot slipping is a result of the relative position of the character's feet and the floor. Therefore, perception of slipping could have been confounded by a belief that the floor motion, not the character's motion, varied between motion clips. If this were the case, we would expect observers to repeatedly fail to produce a threshold at 100% slipping with a moving camera. This was seemingly the case with two participants, whose data was withheld from analysis.

It is also possible that fatigue played a factor in the results for trials with a moving camera. Some observers may have lost focus or interest, or suffered from visual exhaustion, after concentrating on many trials of similar motion clips. Many observers demonstrated higher variability for threshold estimates in this case, which may indicate fatigue.

In summary, the estimate for the perceptual threshold of foot slipping with a stationary camera is well-supported by the distribution of the sample data. The more complex case of slipping with a moving camera produced inconclusive results, due to abnormalities in the resulting data. However, although the data is statistically inconclusive, it does provide evidence that camera motion can affect perceptual thresholds. Further data collection would be useful in this case, to be able to characterize the effects of camera motion more confidently.

4.2 *Implications*

It is noteworthy that perception is not a static process. By the 19th century, Fechner recognized that thresholds are not absolute; perception is subject to alertness and the nature of the test situation [Plotnik 2002]. Similarly, Gombrich states that perception "is always an active process, conditioned by our expectations and adapted to situations" [1960 in Berger 1998]. Therefore, our findings for foot slipping with a stationary camera are statistically significant for the test environment, but the calculated threshold can best be applied in general as a basic guideline, rather than a hard, fast limit. A more complete study would investigate attentional factors on perceptual thresholds for foot slipping, since many objects and characters simultaneously vie for attention in a typical game environment. The numbers we present are therefore “best-case” thresholds, for situations where attention is fixed on the slipping character; it is likely thresholds would be even higher if distractors such as unrelated tasks or additional characters were introduced.

Although it seems reasonable that simple camera motion would impede visual perception, we can only speculate on the grounds for the higher average threshold produced. Camera motion is common in film and in games (with strafing and in cut-scenes, for example), so it is not entirely unfamiliar, and we chose a simple motion that would minimize disorientation. The camera motion of moving right while looking forward is somewhat unnatural, compared to typical human motion. Despite the explicit instructions to look at the character’s feet, it is possible that the introduced motion of the floor grid is a confounding factor in perception. The visual activity of the moving floor may add more noise for the limited visual system to wade through, in focusing on the character.

The camera motion also introduces uncertainty about the distance traveled by the character in each clip, since the observer must now factor in the relative velocities of the character and camera. The checkerboard floor was used to aid the observer in determining the character’s position, such that walking in place (100% slipping error) should be easily distinguishable from normal walking (0% error). Although the floor grid appears to the observer to be moving, the character’s feet step on the same grid squares over and over. However, it is surprising to note that with a moving camera, observers occasionally failed to distinguish between the two cases (in 9 of 140 blocks). This may be due to the fact that the camera motion is in the same direction as the character’s walking, so that as the squares of the floor grid progress from right to left across the foreground of the display, the observer assumes that the character is progressing as well. In other words, camera motion may introduce ambiguity to determination of character velocity, which diminishes the prominence of slipping.

Whatever the psychophysical cause, the collected data suggests that at least one type of camera motion may obscure foot slipping. Regardless of the distribution of the results with a moving camera, the results of the portion of the study with a stationary camera can be applied to perceptually-based rendering techniques as practical, numeric guidelines. In a character which is likely to receive visual attention, root velocity change should not exceed 17%, as it may produce perceptible foot slipping error. With camera motion introduced the slipping may go unnoticed with higher root velocity changes (even up to 70%), although this likely varies with the nature of the character and camera motion. Re-application of the second portion of the experiment with refined instructions and naïve observers would be required for a better estimate of this threshold. Cases where character velocity changes approach these empirical thresholds merit application of a footskate cleanup technique such as that suggested by Kovar et al. [2002].

5. Future Work

Due to time constraints, and the goal of limiting the variables involved for the sake of reliable data, we only considered two variables in this study: magnitude of slipping error and the presence or lack of camera motion. The data presented here suggest a possible correlation between the presence of camera motion and an increased perceptual threshold for foot slipping. However, further data collection with a moving camera is recommended for better support of this correlation. We may consider decreasing the camera velocity in future trials, in an attempt to produce more consistent thresholds. Once the effects of one camera motion can be demonstrated with statistical confidence, and the departure from the normal distribution is eliminated, we could then increase camera velocity and study correlation between velocity and perceptual thresholds for slipping.

Many other variables could be introduced, in order to investigate this phenomenon further and support a generalization about the effects of camera motion on perception of errors. That is, camera motion parallel to character motion seems to mask foot slipping, but data from related, future experiments could help to construct an overall theory relating camera motion and visual perception. Such studies would serve to advance both the fields of psychology and computer graphics.

5.1 *Variation of Camera Motion*

In this study, we chose a camera motion which ideally minimizes disorientation and remains centered on the character of attention. This is a rational choice, both for game-play and perceptual research. However, it seems quite plausible that more exotic camera motion may produce higher perceptual thresholds. It would also be interesting to vary the direction of camera motion, in relation to that of the character. Just as looming and receding objects vary in their ability to capture attention [Franconeri and Simons 2003], perceptual thresholds for rendering error may vary as the camera approaches or recedes from a moving character as well. These problems are in some ways equivalent.

5.2 Variation of Motion Error

With the slipping error itself, we only considered cases where the character's root velocity (translational speed) is lower than that of the original (captured) motion. That is, 100% error means the character is walking in place. This seems to be the most common type of slipping in interactive games (or at least the most perceptible to us!). However, captured motion could certainly also be applied in cases where the character is translating faster than the actor's velocity during the capture session $\left(v' > v0\right)$, producing an "ice-skating" effect. In this case, 100% error would mean translating the character at two times the captured root velocity. It is not intuitive how perception of such slipping would compare to the type we studied.

Although the type of slipping that we measure is commonly found in games, rotational slipping is also common, and may be a more glaring error. This is where the character turns about its vertical axis, without stepping through the turn accordingly. We would expect that perceptual thresholds for this type of error would be quite low, measured perhaps by angular velocity over a neighborhood of frames.

Foot slipping is not the only common error in animated character motion. Motion discontinuities, typically arising from motion transitions, also still present themselves in state-of-the-art games. For example, as an outfielder transitions from running to throwing home, the legs or throwing arm may "snap" between successive frames, as the joint angle velocity is unrealistically high. Kovar et al. mention the need for numbers on the perception of this type of error as well [2002]. It should be possible to consider perceptual thresholds for discontinuities in a test framework similar to the one we constructed.

5.3 *The Effects of Attention*

Perhaps the most important and most logical next step, based on recent studies in this area, is to consider the effects of attention on perception of motion play-back error. Visual perception is directly related to attention. In 1907, Balint pointed out that when a person is absorbed in inspecting something, he does not notice anything in his surroundings, although their signals reach the brain [in Simons and Chabris 1999]. A series of studies in the 1970's and 1980's underscored this central role of attention

[Simons and Chabris 1999]. An observer actively carrying out a primary task often fails to notice major changes to other parts of a scene, such as a person in a gorilla suit leisurely passing by.

While a task or other high-level target tends to hold visual attention, certain low-level stimuli have also been proven to capture attention. These stimuli include certain changes in luminance, and the abrupt onset of new objects [Franconeri and Simons 2003]. In their published findings, Franconeri and Simons add moving and looming (approaching) stimuli to this list. The researchers hypothesize that certain visual stimuli are salient because they suggest "behavioral urgency" (for example, ducking or hiding to avoid approaching threats). Such stimuli may capture attention involuntarily, but low-level stimuli can sometimes be consciously ignored or overridden, for the sake of high-level goals.

Simons and Chabris cite multiple studies which suggest that people only perceive and remember objects which "receive focused attention" [1999]. In such studies, where observers fail to notice significant events, they are later quite surprised when the event is explicitly pointed out. Therefore, vision is deceptive, because it provides us with such a rich sensory experience [Simons and Chabris 1999]; in reality, the visual system is surprisingly limited.

Our experimental setup assumes visual attention is devoted to the feet of the walking character. Fatigue notwithstanding, focused attention is quite likely in this case, considering the explicit instructions to watch the feet. Also, the absence of distractors in the spartan environment of the experimental movies, and the salience of motion and human forms in the visual system would suggest a high level of attention on the character, as directed.

Perceptual thresholds have been shown to increase dramatically in the absence of attention, or when attention is “divided.” When an observer is instructed to carry out a measurable, unrelated task, even in the same portion of the visual field, he/she is less likely to notice the presented stimulus. As may be expected, thresholds are even higher when the gaze is directed to a task elsewhere [Harrison et al 2004].

This is highly relevant to game development; it is rare that a single character is alone in a plain environment. Indeed, the question of perceptually-based rendering is somewhat irrelevant in such cases, as they would not be so computationally taxing. Granted, character motion is still limited by the richness of the motion capture data available, and foot slipping could thus still be an issue. However, visual perception and computational resources become more pressing considerations in complex games with dynamic action (and often multiple characters), such as sports games. Processor speed is more limiting in such instances, but so is the human visual system. It seems to be a fair assumption that a subject is more likely to notice a single character slipping in a one-character display, than in a 26-character display (for example in football, counting referees). On the other hand, the body of existing attention research suggests that if the subject is controlling that character, and thus devoting his attention to it, the threshold may decrease. Such a study would require careful design, to minimize interfering factors.

BIBLIOGRAPHY

ANGEL, E. 2003. *Interactive Computer Graphics*. (3rd ed.). San Francisco, CA: Addison-Wesley.

BERGER, A. A. 1998. *Seeing is Believing: An Introduction to Visual Communication*. (2nd ed.). Mountain View, CA: Mayfield Publishing Co.

BODENHEIMER, B., ROSE, C. F., ROSENTHAL, S., AND PELLA, J. 1997. The process of motion capture: dealing with the data. *Computer Animation and Simulation '97*, Eurographics Animation Workshop, D. Thalmann and M. van de Panne, eds. Wien, Österreich: Springer-Verlag, 3-18.

BRUDERLIN, A., AND WILLIAMS, L. 1995. Motion signal processing. In *Proceedings of ACM SIGGRAPH 95*, Annual Conference Series, ACM SIGGRAPH, 97-104.

FRANCONERI, S. L., AND SIMONS, D. J. 2003. Moving and looming stimuli capture attention. *Perception & Psychophysics 65*, 7, 999-1010.

HARRISON, J., RENSINK, R. A., AND VAN DE PANNE, M. 2004. Obscuring length changes during animated motion. *ACM Transactions on Graphics v. 23 n. 3*, 569-573.

HUTCHINGS, B. 2006. MotLib Toolkit. Graphics and Imaging Technologies Lab, Oregon State University.

HODGINS, J. K., O'BRIEN, J. F., AND TUMBLIN, J. 1998. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics 4*, 4, 17-25.

HORVITZ, E., AND LENGYEL, J. 1997. Perception, attention, and resources: a decision-theoretic approach to graphics rendering. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, Association for Uncertainty in AI, 238-249.

IKEMOTO, L., AND FORSYTH, D. A. 2004. Enriching a motion collection by transplanting limbs. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 99-108.

KOVAR, L., SCHREINER, J., AND GLEICHER, M. 2002. Footskate cleanup for motion capture editing. In *ACM SIGGRAPH Symposium on Computer Animation*, ACM SIGGRAPH, 97-104.

LINDSTROM, P., KOLLER, D., RIBARSKY, W., HODGES, L., FAUST, N., AND TURNER, G. 1996. Real-time, continuous level of detail rendering of height fields. In *Proceedings of ACM SIGGRAPH 96*, Annual Conference Series, ACM SIGGRAPH, 109-118.

MYSQL AB. 2006. Available from http://www.mysql.com.

OESKER, M., HECHT, H., AND JUNG, B. 2000. Psychological evidence for unconscious processing of detail in real-time animation of multiple characters. *J. Visualization and Computer Animation 11*, 105-112.

O'SULLIVAN, C., CASSELL, J., VILHJALMSSON, H., DOBBYN, S., PETERS, C., LEESON, W., GIANG,T., AND DINGLIANA, J. 2002. Crowd and group simulation with levels of detail for geometry, motion and behaviour. In *Proceedings of EG-Ireland 2002*, 15-20.

PARENT, R. 2002. *Computer Animation: Algorithms and Techniques*. San Francisco, CA: Morgan Kaufman, 177-186.

PHP GROUP, THE. 2006. *PHP: Hypertext Preprocessor*. Available from http://www.php.net. Accessed April 2006.

PLOTNIK, R. 2002. *Introduction to Psychology*. (6th ed.). Pacific Grove, CA: Wadsworth Group.

PROCTOR, R. W., AND VAN ZANDT, T. 1994. Human factors in simple and complex systems. Boston, MA: Allyn and Bacon.

REITSMA, P. S. A., AND POLLARD, N. S. 2003. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. In *Proceedings of ACM SIGGRAPH 2003.* New York: ACM Press / ACM SIGGRAPH, 537-542.

SATTLER, M., SARLETTE, R., MÜCKEN, T., AND KLEIN, R. 2005. Exploitation of human shadow perception for fast shadow rendering. In *Proceedings of the Second Symposium on Applied Perception in Graphics and Visualization*. New York: ACM Press, 131-134.

SEKULER, R., WATAMANIUK, S. N. J., AND BLAKE, R. Perception of visual motion. In *Stevens Handbook of Experimental Psychology.* (3rd ed.). New York: Wiley, Ch. 4.

SIMONS, D. J., AND CHABRIS, C. F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception 28*, 1059-1074.

THINGVOLD, J. 1999. Biovision BVH. Lecture notes. University of Wisconsin. Available from http://www.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff/BVH.html. Accessed 1 March 2006.

WITKIN, A., AND POPOVIC, Z. 1995. Motion warping. In *Proceedings of ACM SIGGRAPH 95*, Annual Conference Series, ACM SIGGRAPH, 105-108.

APPENDIX A

SUMMARY OF RESULTS BY PARTICIPANT

This table presents the average perceptual threshold by participant, for each of the two cases studied: foot slipping error percentage with a stationary camera, and foot slipping error percentage with a moving camera.

If a participant failed to detect the difference between the 0% and MAX_ERROR% motion clips in an ascending block, or successfully detected the difference between the 0% and MIN_ERROR% motion clips in a descending block, this was logged as a “failure to find threshold” error.

ID	STAT. CAM	MOVING CAM	FAILURES TO FIND THRESHOLD
101	22.5	52.857	1
102*	10.625	18.333	2
103	11.875	33.75	0
104	18.75	70	0
105	11.875	17.143	1
106	10.714	22.5	1
107	20	60	0
108	16.875	38.75	0
109	11.875	57.5	0
110	20	21.25	0
111	11.875	17.5	0
112	16.875	56.25	0
113*	26.25	96.667	2
114	19.375	62.5	0
115	12.5	20	0
116	26.25	55	0
117*	22.5	80	3
118	18.75	75	2
119	21.875	47.5	0
120	18.75	17.5	0
121	16.25	71.25	0

* The data from these three participants was omitted from statistical analysis, due to irregularities, and in two of the cases, prior experience with graphics environments.

APPENDIX B

POST-EXPERIMENT QUESTIONNAIRE FORMAT AND RESULTS

The following questionnaire was applied to each participant at the completion of the experimental trials. The questions were presented as a web form, and the ID field was automatically filled.

Post-Experiment Questionnaire

Please take a few moments to fill out the following brief questionnaire, as accurately as possible.

1. ID: _______
2. Age: _______
3. Gender:
   M F
4. Do you require corrected vision (glasses or contacts) for computer work?
   Y N
5. Did you wear glasses or contacts during this experiment?
   Y N
6. How often do you play 3D computer or console (XBOX, Playstation, etc) games?
   a. Never
   b. Rarely (several times per year)
   c. Occasionally (several times per month)
   d. Frequently (several times per week)
   e. All the time (daily)
7. How many full-length computer-animated movies have you seen?
   a. None
   b. 1-3
   c. 4-6
   d. 7 or more
   e. Not sure

Summary of Responses

- 13 Males, 8 Females.
- Ages ranged from 21 to 34 (average 23.3).
- All 9 participants who reported they require corrected vision for computer work wore glasses or contacts during the trials.
- 10 of the 21 participants play computer or console games at least occasionally (several times per month); 5 reported that they never play.
- 9 of the 21 participants had seen seven or more computer-animated films; all participants had seen at least one.