Affordances and Feedback in Nuance-Oriented Interfaces

Chadwick A. Wingrave, Doug A. Bowman, Naren Ramakrishnan
Department of Computer Science, Virginia Tech

660 McBryde Hall
Blacksburg, VA 24061

{cwingrav,bowman,naren}@vt.edu

ABSTRACT

Virtual Environments (VEs) and perceptive user interfaces must deal with complex users and their modes of interaction. One way to approach this problem is to recognize users' nuances (subtle conscious or unconscious actions). In exploring nuance-oriented interfaces, we attempted to let users work as they preferred without being biased by feedback or affordances in the system. The hope was that we would discover the users' innate models of interaction. The results of two user studies were that users are guided not by any innate model but by affordances and feedback in the interface. So, without this guidance, even the most obvious and useful components of an interface will be ignored.

Categories and Subject Descriptors

Computing Methodologies -Computer Graphics - Three-Dimensional Graphics and Realism (I.3.7): **Virtual reality**; Computing Methodologies -Computer Graphics - Methodology and Techniques (I.3.6): **Interaction techniques**; Information Systems -Models and Principles - User/Machine Systems (H.1.2): **Human factors**;

Keywords

nuance, virtual environments, perceptive, machine learning

1. INTRODUCTION

Nuance-oriented interfaces[1] work under the hypothesis that the user has an innate model of interaction; we only have to perceive it. This innate model will be their first intuition for performing any action and will dictate the user's methods of increasing performance by performing *nuances*, which can be found in input device data. A *nuance* will be defined for our purposes as a repeatable action the user makes in their interactions in an environment, intentional or not, that are highly correlated with an intended action but not implicit in the interaction metaphor[1]. We have recognized four categories of nuances; object, environment, refinable and supplemental. If these nuances could be identified and framed as a part of the interaction technique through machine learning (ML), users would have a more responsive environment to their actions. For virtual environments (VEs), this could lead to improved efficiency and presence, possibly even new uses for VE technology [2]. In this paper, we discuss the partial results of from an ongoing study to create a demonstrable VE nuance-oriented interface.

The aspects of what a nuance-oriented interface hope to capture can best be described through scenarios of interaction:

> *Dave the architect is trying to place walls in a model of a building he is working on. His interface uses a non-linear arm mapping function to move his hand when he manipulates objects in the environment. This allows him to move objects precisely up close but still reach way off in the distance, due to the non-linearity. The problem he is having now is trying to place a wall on the other side of the building. He tries to put it into place but even the jitter of the tracking system is enough to make his wall bounce out of position.*

The non-linear arm extension mapping [3] is helping him reach the other side but it is not allowing him the ability to work with enough precision once his arm is there. The interface should build a nuance that recognizes when the user is getting frustrated and change the mapping function to something more appropriate.

> *Dave has built his building and now wants to see which type of lighting looks best. As he is standing under the hanging lamp in the center of the room, he realizes that there is not enough illumination. To remedy the situation, he points at the lamp and a ray extends from his finger. As he points, the ray keeps flipping back and forth between the light bulb and the lamp since the bulb is so small and hard to point at. Dave is frustrated.*

In this situation, it is probable that Dave is quite accurate in pointing at the bulb but his precision is varying, making his pointing overflow onto the lamp. The nuance could be that the object to be selected should have some dependency upon where the ray has been pointing over time. In a sense, this nuance would damp the ray's motion.

This paper discusses two experiments whose results showed that users have an innate model of how to respond to affordances and feedback rather than an innate model of interaction in the environment. An affordance is, "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used" [4]. Feedback is information provided to the user by the interface during or after a user action that assists them in their understanding of the current action. These two concepts are critical to interface design and our ongoing work is framing them in nuances.

2. ADVANTAGES OF NUANCES

We should be able to recognize the smallest pieces of information in an environment, the nuances, by mining interaction logs of users and use those nuances to build interfaces. These nuances should be transferable to new environments so time spent by researchers identifiing nuances can be amortized across the construction of similar interfaces. This will lead to robust, quickly developed, interfaces using nuances as building blocks. There are two other main advantages of such a representation.

2.1 Nuances Support Mutual Disambiguation

The argument has been made that multiple input modalities support mutual disambiguation which increases accuracy [5]. In the same manner, multiple nuances and can be though of as multiple modalities providing information to the interface to increase accuracy. Hidden Markov Models (HMMs) have been employed to recognize co-occurrence of voice and gestures as well as sequences of action [6], the goal being methods to develop interfaces from co-occurrence. Our previous work has used a neural network (NN) to recognize user actions in a selection task among randomly placed and overlapping spheres [1]. The NN learned the trivial nuance that hand location was a determining factor as to which sphere was being selected. The NN went a step further and recognized that wrist orientation also could be used, especially in the case of overlapping spheres. The resulting interface was robust, a feat that would be difficult for interface designers to duplicate.

2.2 Inducing/Deducing Interfaces

Nuances help in the recognition of the actions a user takes (induced from user input data) and the optimization of the paths to complete a goal (the deduction of the path to reach a goal). VEs produce parallel, continuous, probabilistic, passive input data streams [7] which we can record. Learning methods such as neural networks, decision trees, HMMs or inductive logic programming can then treat the data as a programming by example problem and map the data to user actions. For the optimization of user goals, we can look upon the user's actions and choices as the result of a Markov decision process. A Markov decision process is where there exists a set of states S, a set of actions A, an agent (in this case the user) and where the state transition probabilities are stationary. At each step, the agent knows its current state s_t, and chooses an action a_t, receiving a reward r_t which is a function of the state and action chosen ($r_t = R(s_t, a_t)$ where $s_t \in S$, $a_t \in A$). The modeling assumption is that the user chooses actions whose associated rewards contribute to a value function. Inverse Reinforcement Learning [7] can then be used to deduce that value such that we can anticipate, and hopefully assist with, the user's next subgoal towards the completion of their goal.

3. LONG TERM PLAN

Our original plan was of four phases to generate a selection technique in VEs based upon the concepts of a nuance-oriented interface. The goal of this process was to identify the nuances of selection techniques based upon preference, regions of space, object affordances and properties of the environment. In each phase, using the nuances from the previous phase, selection trials were to be performed to collect data from the users. This paper discusses phase 1 and part of phase 2.

In each phase, we used the same group of eight users and all eight of the users were taken from a graduate level course on virtual environments, though not all were computer science students. All had some familiarity and interest in the field of virtual environments but not necessarily experience. There were five males and three females between the ages of 24 and 54, with most towards the lower end of the age-group. Their compensation was receiving extra credit in the graduate level course.

The equipment used for these phases was an SGI Indigo 2 with Max Impact graphics and the user inside a Virtual Reality V8

Head Mounted Display (HMD). They had their hands and head tracked using a Polhemus 3 Space Fastrak™ magnetic tracker and finger pinches were recorded using Fakespace PinchGloves™. A selection was considered to have taken place when the user pinched either their index and thumb or middle and thumb fingers together. The environment was programmed using DIVERSE [9] and JIVE [10].

3.1 Phase 1: Optimizing Selection Techniques

Phase 1 was to discover refinable nuances, nuances that alter existing behavior, for three VE selection techniques: arm extension, ray casting and occlusion. The concept of arm extension involves the user reaching their hand out to the object to be selected. When they feel that the hand is touching the object to select, they pinch. The second selection technique was ray casting [11] and involves the user pointing their hand at an object and pinching when they believe a line extends from their hand to the object. This allows users to select objects when their hand is in a non-fatiguing position down by their side. The third technique was occlusion selection [12]. It involves the user placing some part of their hand between their eye and the object, pinching when aligned. Most interface designers stop at these high-level techniques and do not try to tune them for the user. For example, ray casting is almost always implemented with a ray extending directly from the hand without ever trying to discover if this is the optimal configuration.

For each technique, users were informed how the technique worked and the instructions were vague enough to not guide their actions but informative enough to let them know how the technique works and is implemented. The concept was that users have an existing model of how they wish to interact with the environment and if we isolate that underlying model, then we can use that knowledge to recognize their actions. This information can then be used in further phases where we isolate other factors of the interface.

The difficulty we encountered was to make an interface where the user would act naturally and not adapt. To this end, we attempted to remove all forms of feedback and affordances from the environment relevant to each selection technique. For this reason we did not implement ray casting with a ray extending from the user's finger or even a hand for arm extension. We then assumed that since the user was operating according to their own definition of optimality, and we knew their goal to be the selection of an orange sphere, then each time they conclude a selection with a pinch, they were correct in their selection. Since the user can assume that the interface is 100% accurate in its recognition, they can then interact without adapting.

3.1.1 Environment

The environment (see Figure 1) in all three sets of trials (one for each selection technique) had the user standing on a platform overlooking a floor with one orange sphere that they were told to select using the selection technique that was currently being tested. Their head was tracked and a virtual hand, at the same position and orientation as the user's physical hand, was shown (except in the arm extension technique as it was felt that the hand, being shown, would alter the way in which the user acted since the hand is the major form of feedback in arm extension). To account for a lack of depth cues, the users were told the sphere was the same size throughout the experiment and that the floor

was a grid of one-meter squares. There was also a shadow, properly scaled for depth and approximately scaled for height, placed below the sphere on the ground. Each set had 30 trials where the sphere was moved through different locations with the first three trials being the sphere at its furthest distance, middle distance and closest distance to the user to help them get an idea of the environment's depth. The other 27 trials had the sphere randomly located at a position composed of near, mid or far; low, level or high; left, center or right. One side effect noticed in the pilot study was that users were able to cycle quickly through selections because of our assumption that each episode was correct. To counter, we added a three second pause between each selection episode and added an audible sound when the orange sphere reappeared.

Figure 1. Phase 1 with the ball at a distant position. Notice the shadow of the sphere and the gridded floor.
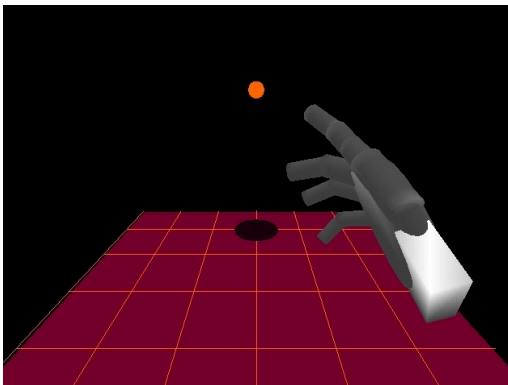
3.1.2   Results

With users free from the feedback of the environment, we expected them to revert to their most natural model of interaction built off of innate intuition. What occurred was an amazing display of adaptation on the part of the user; completely unnatural and inefficient but incredibly effective at aligning the scarce feedback that was left in the system with the user's belief in what the interface should be.

Figure 2. Two occlusion selections used most commonly in phase 1. Left is the palm occlude (with the sphere behind the palm) and right is the thumb knuckle occlude. Both are inaccurate and highly occlude the scene but for some reason users converged to them.
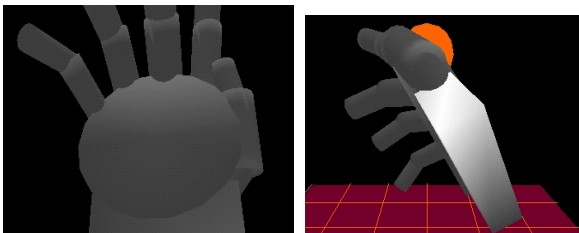
Each technique had interesting results. Users of arm extension were found to not have a concept of depth. We expected users to scale the extension of their arm to the objects being selected but found that users only divided space into "far" and "near" with far being a fully extended arm and near being a half-way extension. Occlusion selection contained the most interesting results. The

users choose unusual points on the hand as the occluding points. The two most common were actually the palm of the hand and the knuckle where the thumb meets the hand (see Figure 2). The palm of the hand occlusion technique occluded most of the scene making the accuracy very low. The thumb knuckle technique is inaccurate and again occluding. It does however leave the hand in a natural and thus non-fatiguing state. One subject spent the entire occlusion selection trial making selections with their palm facing out. This is a very uncomfortable position, even for short periods of time, and completely occluded the environment. A few users did choose to use the more accurate and less occluding fingertips. For ray casting selection, only one user did true ray casting. All the other users occluded the object with the tip of their finger and considered that pointing at the object (see Figure 3). This completely voided the concept of "shooting-from-the-hip" to reduce fatigue, but with the lack of a ray extending from the fingertip, this provided the most feedback to the user.

Figure 3. All but one user considered ray casting to be a fingertip occlusion technique.
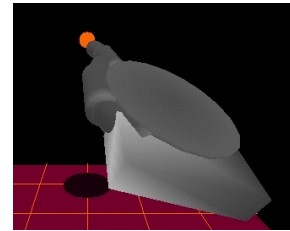
3.1.3   What We Learned

The results led to the following hypothesis:

> *Users largely do not have a model of interaction with the environment but a model of how to respond to feedback the environment provides.*

Stated another way, *users attempt to align their actions with feedback and affordances and not their innate model*. The effect of user experience with VEs may play an important role in this conclusion.

Our original intent was to build personalized selection techniques for the users. After reviewing the results, it was not considered possible to use the data since the users were so inefficient with their interaction in virtual environments without feedback to guide them. A k-Means clustering on logged data was performed to see if trends existed in user data but the trends just mimicked the observations.

3.2   Phase 2: Optimizing Selection Techniques

Because of the unexpected results, we reevaluated our assumptions. Instead of removing feedback and hoping that the user would act naturally, we added as much useful feedback as we could. To occlusion selection, a bullseye was added to guide users as to where they were to align the ray from their eye. When within a configuration-specific snap-to angle, the bullseye snapped to the closest object to provide feedback showing that the technique is aligned with an object.

A nearly identical environment was used in phase 2. The user was asked to do at least one set of 10 trials for each predefined configuration as shown in Table 1. The authors, through their intuition and experimentation, created these configurations and also included the two common occlusion configurations from

phase 1. After users experimented with the configurations, they were then asked to qualitatively rate them on a scale of 1 to 5. (this experiment was also performed on ray casting, but we limited the discussion here to occlusion selection).

Table 1. The predefined occlusion configurations.

Configuration 1: The bullseye is on the index finger and has a 10-degree snap-to angle.

Configuration 2: The bullseye is on the middle finger and has a 10-degree snap-to angle.

Configuration 3: The bullseye is on the thumb's knuckle with a 10-degree snap-to angle. This was a configuration that was used heavily in the first implementation.

Configuration 4: The bullseye is on the palm of the hand with a 10-degree snap-to angle. This was a configuration that was used heavily in the first implementation.

Configuration 5: The bullseye is placed a few centimeters off of the palm with a 10-degree snap-to angle.

Configuration 6: The bullseye is placed on the index finger and has a 45-degree snap-to angle.

Configuration 7: The bullseye is placed on the index finger and has a 3-degree snap-to angle.

Figure 4. User ratings of the predefined configurations showed that they did not like their configurations from the previous phase (3 and 4).
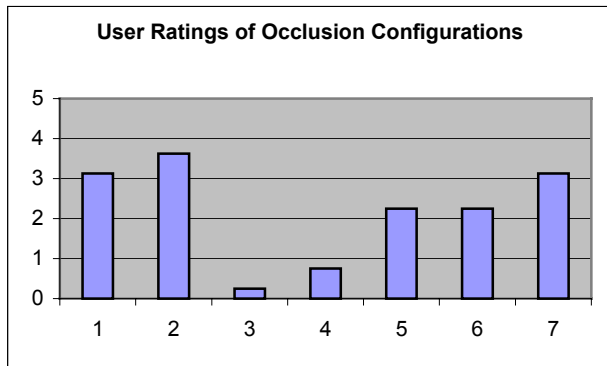
As can be seen in Figure 4, users did not like configurations 3 and 4, though they had performed them when they were not guided by any feedback in phase 1. The most liked configurations, as expected, were those with the bullseye on the fingers with an unexpected preference towards the middle finger. The significance was not calculated however, due to a small number of users.

The feedback of the system was the only change and it guided the user into performing selections requiring less fatigue. The users also preferred configurations that were more accurate and less occluding of the scene. Since the users without the feedback did not perform these more optimal configurations, which were available in the first phase, then the feedback plays an important role in guiding the users to better interaction.

4. CONCLUSIONS AND FUTURE WORK

The work is just beginning for nuance-oriented interfaces. The goal of this and continuing work is a system fully able to handle user nuances in complex interfaces, specifically VEs. In this attempt to perceive the true nature of the user's innate model of interaction, we observed that the model was not internal but built on the feedback and affordances inherent in the environment. Because of this, we recommend that nuance interfaces, being a perceptive interface, should not have their focus purely on being perceptive of the user but focus on how to perceive, such as to guide. This may lead to unnatural interfaces, but with users making better use of the interaction which they were guided to, hopefully it will also lead to higher user satisfaction. Without this guidance, the user will never take advantage of all the richness of an interface because they will not know it exists, with the overall effect being a perceptive, yet still useless, interface.

5. REFERENCES

[1] Wingrave, C., Bowman, D., and Ramakrishnan, N. A First Step Towards Nuance-Oriented Interfaces for Virtual Environments. Proceedings of the Virtual Reality International Conference '01, pg 181-188.

[2] Rosen, J., and Hannaford, B. Markov Modeling in Minimally Invasive Surgery… in IEEE Transactions on Biomedical Engineering, vol 48, no 5, May 2001, pg 579-591.

[3] Poupyrev, I. Billinghurst, M. Weghorst, S. Ichikawa, T. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. Proceedings of the ACM Symposium on User Interface Software and Technology, 1996, pg 78-80.

[4] Norman, D. The Design of Everyday Things. Doubleday, 1990.

[5] Oviatt, S. Mutual disambiguation of recognition errors in a multimodel architecture. Proceeding of the CHI 99 conference on Human factors in computing systems: the CHI is the limit, 1999, Pages 576 – 583.

[6] Sharma, R., Cai, J., Chakravarthy, S., Poddar, I., and Sethi, Y. Exploiting speech/gesture co-occurrence for improving continuous gesture recognition in weather narration. Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, 2000, pg 422 –427.

[7] Jacob, R. J. K., Deligiannidis, L., and Morrison, S. A Software Model and Specification Language for Non-WIMP User Interfaces. ACM Transactions on Computer-Human Interaction, 1999, vol 6, no 1, pg 1-46.

[8] Ng, A. Y. Russell, S. Algorithms for inverse reinforcement learning. International Conference on Machine Learning, 2000, Stanford, CA: Morgan Kaufmann.

[9] Arsenault, L., Kelso, J., and Kriz, R. DIVERSE. [Online] http://www.diverse.vt.edu.

[10] Wingrave, C. A. JIVE: Just In a Virtual Environment. [Online] http://csgrad.cs.vt.edu/~cwingrav/Jive/home.html.

[11] Mine, M. Virtual Environment Interaction Techniques. Technical Report TR95-018: UNC Chapel Hill CS.

[12] Pierce, J. S., Forsberg, A. S., Conway, M. J., Hong, S., Zeleznik, R. C., and Mine, M. R. Image plane interaction techniques in 3D immersive environments. Proceedings of the symposium on Interactive 3D graphics, 1997, pg 39.