

# INTERACTIVE MEDIA DESIGN METHODS FOR REAL-TIME, REACTIVE ANIMATION

Yangmi Lim<sup>1</sup>, Jaewon Lee<sup>2</sup>, Jinwan Park<sup>3</sup>

<sup>1,3</sup>Graduate School of Advanced Imaging Science, Multimedia & Film, Chung-Ang University, Seoul, Korea,  
<sup>2</sup>School of Computer Science and Engineering, Sungshin Women's University, Seoul, Korea,  
yosimi@chol.com<sup>1</sup>, jwlee@sungshin.ac.kr<sup>2</sup>, jinpark@cau.ac.kr<sup>3</sup>

## ABSTRACT

Digital interactive contents require interface functions that are user-friendly and easily understood, prompting the need to control successive and free-form user motions during their creation. To promote extended user interaction with contents, it is, therefore, critical to design the various interaction functions with this goal in mind. In this paper, we introduce three different systems for promoting human motion and computer interaction. We implement and experiment with the interface understanding of users according to the serial system, parallel system, and story-based system. We compare these three approaches to know measure the degree of efficiency of the following: the number of interaction functions, reaction time, and the relation between user interface and content reaction. The research done in this paper paves the way for future interface design, and this data can be used for creating a media interaction.

Keywords: Interface, Interaction, reaction

## 1. INTRODUCTION

In this paper, we introduce the importance of an interface process that utilizes user motion for interactive digital contents. In the creation of interactive artwork, the understanding and subsequent implementation of the interface is dependent on the designers and programmers, on how much interface they choose to integrate into their work. In the creation of a work, there is a need to balance the requirements of the different parties involved, from the programmer, to the designer, to the user, and regulations must be implemented to address these. This paper explains three interface systems using user's motion, developed step by step: the serial system, parallel system, and story system based interface designs (Cavazza, Martin, Charles, Mead, and Marichal 2003). Interface development is not based upon just the designer's idea, it is made in collaboration with a programmer familiar with computer systems. Thus, in developing digital contents with interaction, the understanding and knowledge of both, a designer and a programmer, is the critical important factor.

Until recently, user interaction with digital contents has been limited to user actions with a mouse and keyboard. The navigation vehicles on the internet include the list box, tree views, large icon list view, small icon list view, detail list view, toolbar buttons, and the kinds of actions are defined by toolbar buttons, command buttons, menu items, drop-down menu items, popup menu items, double-clicks, and, drag and drop. This paradigm clearly delineated the responsibilities of the designer and programmer. However, current interface improvements such as the blue tooth virtual keyboard, operating machinery by remote control and enjoying games in virtual world, are able to recognize user motion based on the camera, requiring a new understanding of interface design. These can be broken down into these major areas of interest: the reaction of a character in a virtual reality environment (Lawn and Takeda 1998), and new interfaces using intuition and bodily sensation (Tae-il 2003).

In this research, we communicate with the computer through a camera-based interface instead of the traditional mouse and a keyboard (Wang, Hu and Tan 2003), and the digital content is controlled by user motion. Therefore, the three interface design experiments presented in this

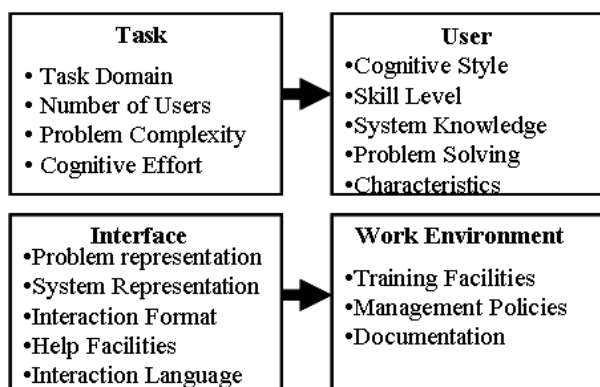
paper make a comparative study of the degree of user understanding between the different contents produced as a result of their action and reaction.

The aforementioned three works on user motion discusses how designers, programmers, and users, define and understand user motion interface. We describe and address the limitations of the camera- based interface, since a computer is unable to compute successive movements and unplanned motion. For our experiments, we set the following as benchmarks: interface function numbers, response time/accuracy and relation. Using these, we compare and differentiate three approaches to interaction by means of user motion: the serial system, parallel system, and story system-based interfaces; we then study user reaction to the content.

Researchers in computer science and engineering have devoted a great deal of effort in understanding how an interface ought to be designed. The data from this work can be used in the understanding, design and implementation of interfaces for the user, and may be used for involving other people aside from the designer in interface design (Santhanam and Batra 1998).

## 2. RELATED WORKS FOR INTERACTION MEDIA PLANNING

The interactive user component is composed of user characteristics, such as cognitive style, experience with computers, knowledge level, and problem solving skills, among others. The interface serves as the locus of interaction with the computer, and, to the user, the interface is “the system” (Norman 1986). The four major components of the human-computer interaction process are shown in Fig. 1.



(Fig. 1) Components of the Human-Computer Interface

Gerald et al classified visual information, and, made symbolic signs focusing on the visualization of common text information, before making the interface design. In their design process, they classified according to function, and proposed the method of expressing visual information in various media, from text tables to graphs (Gerald, Kevin, Neff, and Henry 1994).

According to McCormick et al, we can communicate with each other by using visual information after understanding the human-computer mechanisms. They improved the interface design, improving the interchange of visual information. For example, in a software product, they developed the interface design with the goal being easy user access: shortcut key(ctrl+q), icon(☒), menu bar(close or exit), close (McCormick and Brown 1987). However, this interface produced problems different from traditional interfaces, i.e navigation with a mouse and keyboard.

Communications with a computer should naturally recognize user intentions. Drawing standard symbols and icons is impossible because of the myriad array of user motions. Therefore, it is necessary to make the computer understand the user motions rather than have the user learn and adapt to the computer's requirements. At the MIT Media Lab, interface design is by the following: (1) to keep the focus on user action and interaction, (2) to permit multiple interface, and to simultaneously engage users in an interactive experience combining both real and virtual worlds, (3) to naturally recognize patterns/actions/motion, using computer-vision, (4) to use narratives to constrain the perceptual recognition of the user (Aaron, Stephen and James 1999).

In this interface development paper, we initially select a method, saved as a regular mode, the computer then calculates the user's arm motion. This process takes into account the benchmark factors-the same time access numbers, fast/accuracy, relation- during the human recognition process in interpreting the arm motions as the visual factors: position, size, color, figure, orientation, transparency, lighting, speed, and a visual model is created. This visualization is vital for designers in the creation of signs, symbols, icons and interfaces because it serves as the overall look the designer wants to achieve (Sungkon 2002). It is these important factors that form the visual language of the human-computer interrelation. Sungkon discussed an idea regarding the number of concurrent accesses, i.e., "Can users have an access to the computer at once?", "Can it be linked quickly and accurately by the click of users?", "Do they have a relation between current information and next information?", and we modify this concept in the design of our

storytelling-based system interface. We change the number of concurrent accesses to reflect the number of interactions.

### 3. INTERFACE EXPERIMENT FOR THE USER'S RECOGNITION

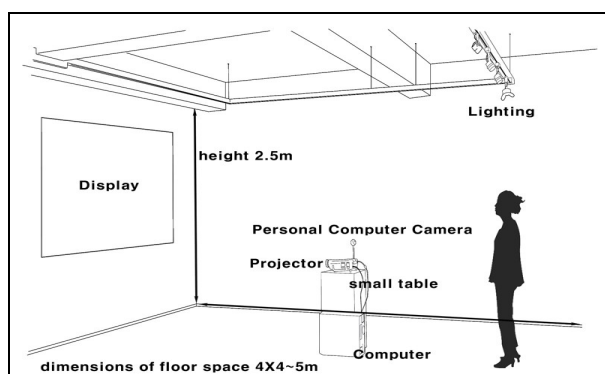
In this chapter, we explain the three experiments employed for verifying the degree of interaction recognition of the digital content. To express a quantitative difference of the function and efficiency of interrelation between human and computer, we divided the three experiments into three categories in the interaction media, based on the work of Sungkon in the web-media: the number of interface functions, response time accuracy, and relation between action and reaction in the structure. The recognition conditions we select are the position, form, orientation and speed of the visual factors. We performed three different experiments for each situation.

The Three Experiments are the following:

Experiment 1: The candlelight flicker rate is controlled-stop, weak and strong, according to arm speed.

Experiment 2: The facial expression of the animation is controlled by the position, form, and orientation of the arm.

Experiment 3: The PuppyMate character in the digital content-based story structure system is controlled when a user walks, runs, or does arm gestures, changing the character's distance covered and the calorie consumption.



(Fig. 2) Experiment set-up for each work.

### 3.1 INTERFACE FUNCTION NUMBERS

A design method leading to the construction of a single interface design instance is inappropriate, as it cannot compensate for the diversity of the resulting reaction. Hence, there is a need for a systematic process in which alternative design decisions for different design parameters may be supported.

Table 1 represents the interaction structure system according to interaction function numbers.

types	Number of Functions	action & Re-action	interaction system structure
Experiment 1	3 (no motion, weak movement, strong movement)	Waving user's arm (action). The candlelight flickers weakly, strongly, and stops.(response)	serial system
Experiment 2	5 (No emotion, Sadness, Anger, Joy, Surprise)	User performing arm gestures (action) The face expresses no emotion, surprise, joy, anger and sadness.(response)	parallel system
Experiment 3	10 user Interface Types and distances ( table 5)	Performing body and arm motions. (action) The distance covered and the calorie consumption changes. (action) PuppyMate stops, walks, sniff, its ears pricked up, etc. (response)	Story based system

(Table 1) interaction structure system according to number of interaction functions

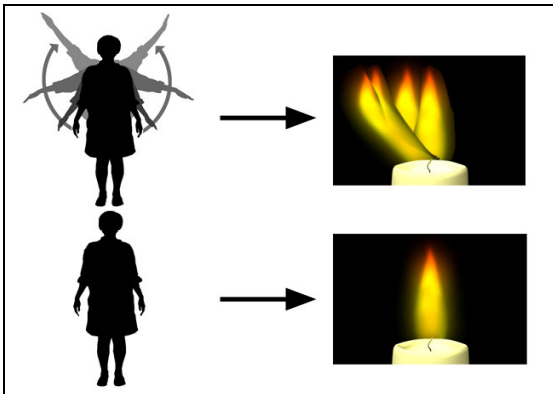
In the experiment 1, the serial system structure gives a one-on-one response, the interrelation changes depending only on the speed of each pose action. This first experiment that reacts based on arm speed has three interface functions.

The value of the speed interface is calculated in arm motions per second. If the arms move once from bottom to top every two seconds, the candlelight flickers weakly. If the arms move twice, the flicker becomes stronger, more evident. If no movement of the arms is detected every two seconds by the web camera, the candlelight flicker stops.

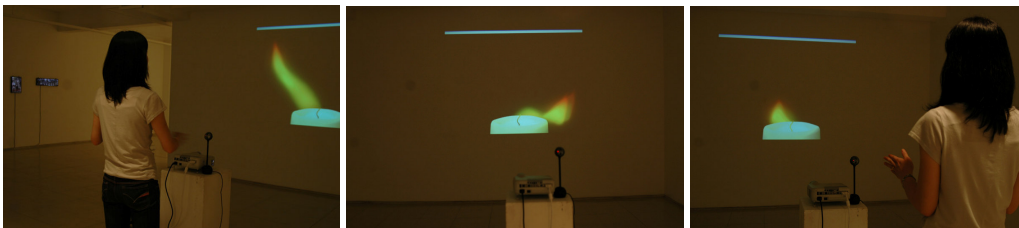
For the experiments, we employed thirty (30) volunteers. Over 83% of the users in our sample of 30 volunteers easily understood the experiment's interface function. Majority of the users understood the interface immediately, and related the experiment to the user blowing wind with every action, affecting the candlelight, increasing speed for stronger wind and not moving for any

wind. However, due to the simple interaction structure, no users continued their motion for more than 20 seconds.

Fig. 3 and 4 shows the relationship between the stop-weak-strong candlelight flicker and the user arm motion in first experiment.



(Fig. 3) The resulting stop→weak→strong candlelight flicker according to arm motion.



(Fig. 4) Experiment 1: Candlelight Image (Yangmi, Kueun and Sezin's gallery, 2006)

In Experiment 2, we observed the length of time common users interacted with the shadow playing system of the parallel system. The parallel system is designed to give a one-on-one response whenever a user acts. We created a separate interface, unrelated with the previous behavior, in implementing experiment 2. The resulting user understanding of the interface was lower compared to experiment 1- all 30 users needed over the 30 minutes required to understand the interface. The elapsed time spent by the user interacting and playing with the system was significantly better than experiment 1 because of the increased number of functions numbers, 25 out of 30 users played with the experiment 2, using their arms to create facial expressions whereas the other 5 users failed to understand the interface. The relationship between arm position and facial expression is more thoroughly explained in the section 3.2.

Experiment 3 recognized and interpreted the action of the body and arms within a narrative system. Therefore, the following interface functions were controlled by user arm gestures: stopping, walking, running, shaking hands, upward hands, shaking head, and stretching, etc. Most users quickly understood the interface because they were able to interact with a dog simulation through the PuppyMate, character utilized in experiment 3, which exhibited real-world canine behavior. This experiment yielded a high understanding among users, 28 out of 30. Furthermore, by the addition of more interface functions, the elapsed playing time was higher than the other 2 experiments. The elements of user gestures and resulting system actions in this story-based system are explained in section 3.3.

### 3.2 TIME AND ACCURACY OF THE RESPONSE

Though experiments 1 and 3 enable the user to draw analogies with real-world situations, giving them high time and accuracy responses, experiment 2 does not have a real-world relation between action and reaction. Expectedly, experiment 2 did not yield good response times and accuracy, prompting the introduction of another performance metric: the creation of symbols from human motion.

The main goal of experiment 2 is to express changes in facial expression as a reaction to the arm gestures. Fig. 5, in chapter 3, shows some examples: anger, joy, surprise and sadness are expressed according to arm gestures. In Fig. 9 in section 3.3, A shows pictures of an actual test, and B shows the captured shadows before animation mapping. C illustrates the resulting images from animation mapping, expressing the emotions of anger, joy, surprise and sadness through arm gestures. However, users can have difficulty drawing analogies between their experiences and expressing facial emotion through their arm gestures.

To address this, we introduce a method of recognizing actions that are usually experienced in real-world situations as a symbolic control method.

Experiment 2: Creating gesture symbols for measuring time and accuracy of the response



For the computer to understand user arm gestures, these need to be divided into clear motions. In the earlier interface design process, we defined five standard poses in experiment 2. A character in the digital content was controlled to include the other motions.

The rationale behind grouping other motions according to the standard poses, defined in Fig. 5, is that the computer has difficulty understanding minute and continuous motions such as arm gestures. To create a standard for identifying individual arm gestures, we separate motions according to the position of the arm, shown in Fig. 5, according to a user's perspective, not a viewer's.

Fig. 5 shows the standard poses, and Fig. 6 groups similar motions into the categories in Fig. 5.



- (A) talking with arms down.      (B) talking with right arms up.  
 (C) talking with left arm up.      (D) talking with both arms up.  
 (E) talking with arms stretched sideways.

(Fig. 5) separation of standard motion



<1. (A) based on various user motions>

<2. (B) based on various user motions>



<3. (C) based on various user motions>



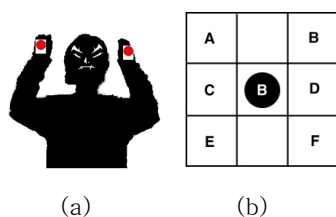
<4. (D) based on various user motions>



<5. (E) based on various user motions>

(Fig. 6) the separation of the various motions

Fig. 6 shows the classification of arm motion gestures, these forms are the most common motions that a user can perform. We define a region Fig. 7 for identifying arm position based on the captured picture data. Fig. 7 shows that a computer equipped with a camera can perform region recognition according to the arm position when the head of the user is at the center of the grid.



(Fig. 7) the tracking method of the standard motion

Thirty (30) volunteers were used in this experiment and 80% of the 100 types of arm motions conducted in the experiment were included into the standard poses of Fig. 5. The remaining 20% were difficult to detect due to two factors. The first factor is that one of the two body parts, (e.g. arms) converges at a single point, making it impossible for the algorithm to distinguish the left arm from right. The second is when the two detected regions are covered by the overlapping of body parts (e.g., an arm overlapping onto the torso), making it appear as a single region.

Tables 2 and 3 show the accuracy results of the experiments. We observed the time required for users to comprehend the human-computer communication, and the accuracy of the user comprehension of the interface functions. The tests were conducted with 30 users who were unfamiliar with the interface function of experiment 2.

motions			interaction response	recognition time < 60 seconds		recognition time > 60seconds	
types	right hand	left hand	emotion	recognized users	rate	recognized users	rate
A	None	None	No emotion	27	0.9	30	1.0
B	Upward	None	Sadness	21	0.7	27	0.9
C	None	Upward	Anger	24	0.8	26	0.87
D	Upward	Upward	Joy	25	0.83	30	1.0
E	Side	Side	Surprise	27	0.9	30	1.0

(Table 2) Comprehension time for human-computer communication with grouping of gestures (Fig. 5 and 6)

Table 2 shows the time needed for users to comprehend the human-computer communication when actions are sub-grouped according to Fig. 5 and 6. 90% of the users understood the interface function after one minute, while only 41% of users accomplished this in under a minute.

Table 3 shows the time needed for users to comprehend the human-computer communication when gestures are not grouped together.

motions			interaction response	recognition time < 60 seconds		recognition time > 60seconds	
types	right hand	left hand	emotion	recognized users	rate	recognized users	rate
A	None	None	No emotion	20	0.67	24	0.8
B	Upward	None	Sadness	20	0.67	23	0.77
C	None	Upward	Anger	19	0.63	23	0.77
D	Upward	Upward	Joy	17	0.57	24	0.8
E	Side	Side	Surprise	15	0.5	18	0.6

(Table 3) Comprehension time for human-computer communication with no grouping of gestures

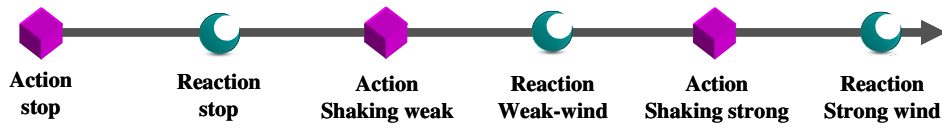
The ratio of users discovering an interface function in over one minute is under 37.4%. Despite numerous tries, the computer was unable to recognize correct value of the interface function. As a result, most users simply gave up. This illustrates that the mechanism for symbolic interaction by recognizing actions is critical to this endeavor. This necessitates the finding of a means to circumvent this limitation.

### 3.3 A COMPARISON OF SIMPLE, MOTION-CHANGE, AND STORY-BASED COMPREHENSION

In this chapter, we discuss the process of human recognition in our three systems, after we identify the factors affecting the interrelation between user behavior and the content character.

Experiment 1:

Experiment 1 uses knowledge-based experience (Andriole, Monsanto, and Ehrhart 1998), hence the candlelight flicker is weak when users shake their arms slowly. If users want to create strong wind, from their experience they can deduce that shaking their arms vigorously increases the flicker of the candle. Fig. 8 shows the flowchart for this simple understanding.



(Fig. 8) simple understanding by serial system

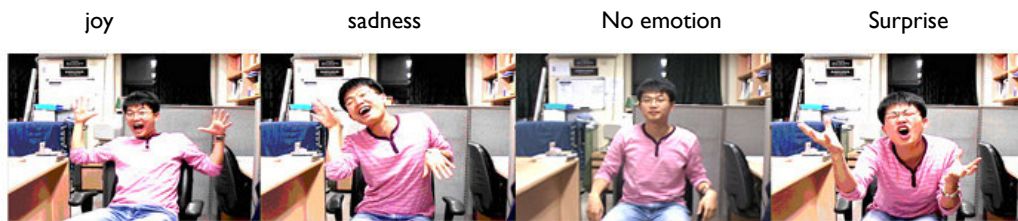
Experiment 2:

In experiment 2, using five arm motion types, the character expresses the user shadow as soon as the user moves his/her arms. The facial expression then changes according to position of the resulting shadow's arms created by the user behavior. Table 4 shows the mapping relations between arm motion and emotional expressions, according to the user's perspective. We reference the emotion motion model of Will Eisner for the minimum relation between user action and content reaction. Initially, the test subjects did not recognize that their behaviors were related to the resulting motion expression of Will Eisner, but, as the users interacted with the system, they produced motions A, D and E (in table 4) - similar to the behavior observed by Will Eisner and the five arm motions in Fig. 5 and 6 (Eisner 1996).

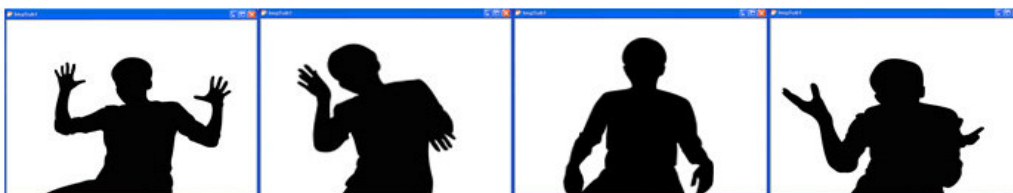
types		user action	motion expression of Will Eisner	content reaction
A	No emotion			
B	Sadness		No	
C	Anger		No	
D	Joy			
E	Surprise			

(Table 4) mapping of human action and expression reaction

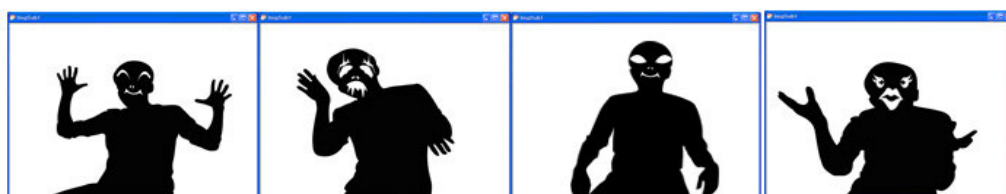
Fig. 9 shows the resulting image using the system employed in experiment 2.



A. pictures of an actual test



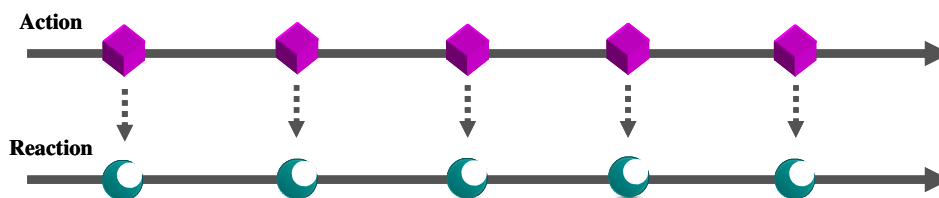
B. separated shadows captured before animation mapping.



C. the resulting images

(Fig. 9) Examples of motion in Experiment 2

Fig. 10 shows the flowchart of the comprehension process.



(Fig. 10) Comprehension by parallel system

### Experiment 3:

In experiment 3, users interact with the PuppyMate character- a performing virtual reality pet, in a storytelling scenario, analogous to a real-world situation. PuppyMate greets its user, accompanies him/her running on the treadmill, and checks the amount of exercise accomplished. PuppyMate promotes its user's fitness programs by encouraging daily exercise until the minimum amount is accomplished. This not only encourages fitness and promotes recreation by collaborative play,

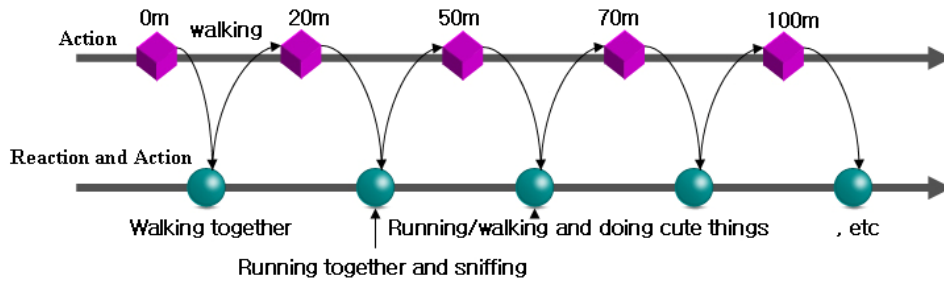
but PuppyMate may also become a companion to communicate with via actions. Table 5 demonstrates the story-based events created by user action, PuppyMate's resulting reaction, and finally, PuppyMate's own action.

user Interface types		distance covered	story events	
No Motion		0m	-PuppyMate stops.	
Walking		20 m	- PuppyMate walks together. (reaction) - If PuppyMate walks a distance of 20m, PuppyMate stops. (action)	
Running		50 m	- PuppyMate runs together. (reaction) - If PuppyMate runs a distance of 50m, PuppyMate stops and sniffs around. (action)	
walking or running (action)	shaking arms (reaction)	right	70 m	- PuppyMate walks or runs (reaction) - If it moves a distance of 70m, PuppyMate does cute things. (action)
		left	100 m	- PuppyMate starts walking again. (reaction) - If it moves a distance of 100m, PuppyMate does more cute things. (action)
	feeding (reaction)	120 m	- PuppyMate starts walking or running again. (reaction) - If it moves a distance of 120m, PuppyMate want user to eats (action) - After PuppyMate eats, it starts walking again. (action)	
	stretching forward (reaction)	150 m	- PuppyMate starts walking or running again. (reaction) - If it moves a distance of 150m, PuppyMate waits with ears pricked and barks. (action)	
	throwing bones (reaction)	170 m	- PuppyMate starts walking or running again. (reaction) - If it moves a distance of 170m, PuppyMate give user to picks up bones somewhere. (action)	
	go to back (reaction)	190 m	- PuppyMate starts walking or running again. (reaction) - If it moves a distance of 190m, PuppyMate goes back home. (action)	

(Table 5) Story-Based Events created by user action, character reaction and character action

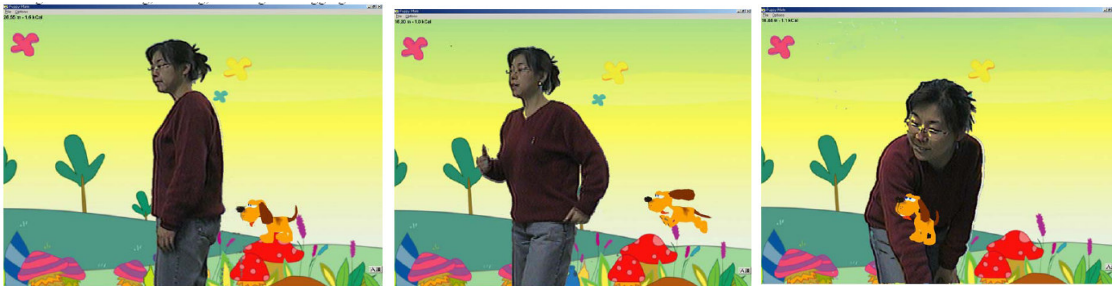
In table 5, if a user starts an action in experiment 3, a story based system, the user is encouraged by the reactions and actions of PuppyMate. Most users learn which actions to do because of the storytelling scenario and PuppyMate's behavior.

Fig. 11 shows the flowchart of this experience-based knowledge understanding and explains the story events in table 5 step by step.



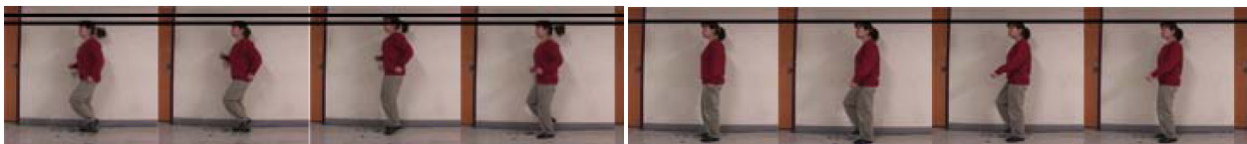
(Fig. 11) experience-based knowledge understanding through a story based system

Fig. 12 shows the resulting image of the system, for recognition of walking, running, standing and begging, in Experiment 3.



(Fig. 12) Experiment 3: Images of Interactions with the PuppyMate system

The identification method employed utilizes the change in the height of the user in the tracking and recognition of the user walking, running and standing. In the PuppyMate system, it was observed that a user often walks, runs and jumps in the same spot within the fixed area. Thus, rather than using temporal template (Cui and Weng 1997, Bobick and Davis 2001) data, which is the traditional method employed for similar situations, we discovered that utilizing the magnitude of height variation of the user while running and walking allows the faster calculations in our experiments. The following figure, fig. 13, shows the amount of change in the height of the user while running, and walking respectively.



(Fig. 13) variation of the height while running and walking

## 4. CONCLUSION

Human and computer interaction requires not only complex interfaces, but also these interfaces likewise need to be easy to use and intuitive. We experimented with three different systems for user interaction with digital media and compared the performance of these with our custom benchmarks.

Through the analysis of the number of interface functions, response time/accuracy and relations, designers and programmers can clearly communicate their expectations, allowing them to adhere to the overall design vision and clearly define a fixed scope for user motion.

In creating the most interactive design interface, Experiments 1 and 3 performed well, with experiment 3 performing best overall. We attribute this to the similarity of the digital story-based scenario to real-world situations, which allows the user to draw analogies from past experiences and interact well with the interface. However, all interfaces cannot be approached via the story-based method, hence, more research on symbolic interfaces is necessary.

Our results may serve as a design aid and reference for the creation of interactive media content. One critical component in designing interactive media is a system that can efficiently recognize human actions. Future research will involve studies into the various interface symbols using human motion, we shall also create a database system that adapts to changing environments in real-time.

## REFERENCES:

Aaron F.B, Stephen S. I et al. (1999), The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment, *Teleoperators and Virtual Environments*, Volume 8, No. 4, pp. 367-391.

Andriole S. J., Monsanto C. A. et al (1998), Knowledge-Based User-Computer Interface Design Prototyping and Evaluation - the Design Pro Advisory System, Air Force Research Laboratory, Information Directorate, Rome Research Site.

Bobick, A. F. and Davis, J. W. (2001), The Recognition of Human Movement Using Temporal Templates, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume.23, No.3, pp. 257-267.

Cavazza M., Martin O. et al (2003), Users Acting in Mixed Reality Interactive Storytelling, *International Conference on Virtual Storytelling (ICVS 2003)*, LNCS 2897, pp. 189–197.



Cui, Y. and Weng, J.J. (1997), Hand segmentation using learning-based prediction and verification for hand sign recognition, in *Computer Vision and Pattern Recognition*, Proc. of IEEE CS Conf. pp. 88-93.

Eisner, Will (1996), *Graphic Storytelling*, by Poorhouse Press.

Gerald L., Kevin B. et al (1994), Classification of Visual Representation, *Communication of the ACM*, Volume 37, No 12, December.

Lawn M. and Takeda T. (1998), Design of an action interface with networking ability for rehabilitation, *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp 25-30 Volume 20, No 3.

McCormick B. H., DeFanti T. A. et al (1987), Visualization in Scientific Computing-A Synopsis, *IEEE Computer Graphics & Applications* 7, Volume 7, pp. 61-70.

Norman, D. A. (1986), *Cognitive Engineering*, In D. A. Norman & S. W. Draper (Eds.), *User Centered System Design*(eds.), pp. 31-61, Hillsdale, New Jersey: Lawrence. Erlbaum Associates.

Santhanam R. and Batra D. (1998), The Human-Computer Interface in Information System Design, *The DATA BASE for Advances in Information Systems*, Volume 29, No. 1, pp. 25-30, Winter.

Sungkon. K. (2002), The Study on Information Visualization Methods Using 3D interactive Animation, *Journal of Korea Society of Design Science* 55, Volume 17, No 1.

Wang L, Hu W. and Tan T. (2003), Recent Developments in Human Motion Analysis, *Pattern Recognition*, Volume 36, No. 3, pp. 585-601, Publisher Elsevier, March.

Tae-il L. (2003), Development of Interactive Experience Scenarios for Interactive Media Design Applying Behavioral Prototyping Methods, 6th Asian Design International Conference, proceedings, Tsukuba, October 14-17.

Yangmi L., Kueun J. et al (2006), *Marriage of Hephaestus*, Kwanhoon Gallery, Seoul, June 28~July 4.

## **Acknowledgement**

This research was supported by Seoul Future Contents Convergence (SFCC) Cluster established by Seoul Industry-Academy-Research Cooperation Project.