Experiential Media Systems – The Biofeedback Project

Yinpeng Chen[1], Hari Sundaram, Thanassis Rikakis, Todd Ingalls, Loren Olson and Jiping He[2]

Abstract. Experiential media systems refer to real time, physically grounded multimedia systems in which the user is both the producer and consumer of meaning. These systems require embodied interaction on part of the user to gain new knowledge. In this chapter we have presented our efforts to develop a real-time, multimodal biofeedback system for stroke patients. It is a highly specialized experiential media system where the knowledge that is imparted refers to a functional task – the ability to reach and grasp an object. There are several key ideas in this chapter: we show how to derive critical motion features using a biomechanical model for the reaching functional task. Then we determine the formal progression of the feedback and its relationship to action. We show how to map movement parameters into auditory and visual parameters in real-time. We develop novel validation metrics for spatial accuracy, opening, flow and consistency. Our real-world experiments with unimpaired subjects show that we are able to communicate key aspects of motion through feedback. Importantly they demonstrate the messages encoded in the feedback can be parsed by the unimpaired subjects.

Keywords: Biofeedback, Analysis, Action-feedback coupling, Validation

1 Introduction to experiential media

Our interaction with our physical surroundings provides us with a complex multi-modal experience. Indeed, the reader may be viewing this chapter as part of a physical book. As part of this simple, everyday experience, the reader experiences the weight of the book, the texture of the paper, the temperature of the book, executes coordinated hand-eye movements to turn the page, and decodes the visual symbols imprinted on a page. This simple task involves sensing, perception, cognition, as well as action that alter the physical world (turning the page). It is a task that requires manipulation of data in a control loop, and this can occur at multiple time-scales. In addition to requiring real-time sensing, perception and decision making, the interaction with the physical world serves as the primary mechanism for us to acquire new knowledge.

[1] Yinpeng Chen, Hari Sundaram, Thanassis Rikakis, Loren Olson, Todd Ingalls

Arts Media and Engineering, Arizona State University, email: {first.last}@asu.edu

[2] Jiping He

Harrington Department of Bioengineering, Arizona State University, email: jiping.he@asu.edu

In this chapter, we examine *experiential media systems* – real-time, physically grounded multimedia systems that enable human beings to acquire knowledge through interaction with a physical system. An important measure of success of such systems is the ability of the participants to transfer (or generalize) the learned knowledge to physical interactions outside of the system, in the physical world.

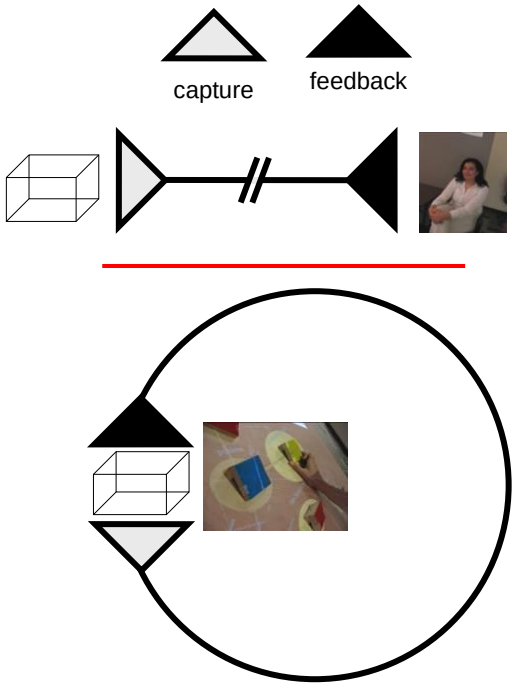

Figure 1: In a situated system (bottom), sensing, analysis and display are co-located and occur in real-time, facilitating a context aware, real-time coupling between human activity and feedback within in the environment. There is a break between capture and consumption in traditional multimedia systems (top).

The successful development of such systems requires us to integrate knowledge from multiple disciplines – computational elements from science and engineering, ability to develop meaningful immersions from the arts and understanding cognition and learning from psychology and education.

The development of the computational elements of this framework is challenging and requires knowledge from several disciplines – artificial intelligence (robotics [3,4,5]), human-computer interaction (tangible interfaces [25,26,31,38], embodied interaction [11]), distributed cognition [23,24,29] and ubiquitous computing [1,40]. Each of these fields have lead us to re-imagine computation integrated with the physical world. Each focuses on a complementary aspect of the interaction with the physical world, however this knowledge is not

integrated with research in multimedia, which focuses on analysis of multimodal data.

Multimedia analysis primarily focuses on difficult problems in the offline analysis and labeling of *captured* data e.g. labeling of a digital photo as "outdoor" / "Jane", summarizing produced media from television or films etc. [19,30,34,35]. Experiential multimedia systems involve *participating* in our multi-sensory environment through the use of real-time, context aware computational elements – sensors and presentation devices that are part of the physical environment. An important feature of experiential media systems is the bi-directional transfer of semantics. The system, by observing human activity, begins to parse meaningful patterns, while through audio-visual feedback, the system *embeds semantics* to be acquired by the human participant. **Figure 2** highlights this difference between traditional multimedia systems and experiential media systems.

A key insight is that with a rapid decline in cost of sensing, storage [15], computing and display, sensors (audio, video, pressure), computing and feedback (displays, sound) can now be co-located in the same physical environment, creating a real-time feedback loop. This allows us to develop a rich contextual understanding of human activity, at different time scales. A physical environment with tangible interfaces, and other sensors (pressure / movement) provides a very detailed, real-time contextual information for each person, helping us to interpret / understand as well as affect human activity in a radically new way. Multimedia analysis does not yet deal with interpreting and summarizing human activity in such environments. We believe that experiential media systems research is a first step towards establishing a computational framework for understanding embodied interaction in our multi-sensory world.

In this chapter, we shall discuss our efforts to rehabilitate stroke patients, via a prototype experiential media system. The goal of this system is to enable the transfer of highly specific knowledge relating to a specific functional task – to be able to reach out and grasp a cup. This simple task is enormously challenging for stroke patients, who have lost the ability to control their arm, and hence must be re-trained (i.e. they need to relearn the necessary motor control) to perform the task.

The rest of the chapter is organized as follows: In section 2, we introduce the biofeedback problem. In section 3, we discuss the framework used to analyze the action of the stroke patient. In section 4 and 5, we present our ideas on coupling action to media feedback in real-time. In section 6, we introduce the performance metrics. In section 7, we discuss system validation. In section 8, we discuss challenges and opportunities for experiential media system design. Finally in section 9, we present our conclusions.

2 The Biofeedback Problem

In this section we present key challenges in the biofeedback problem and introduce our current biofeedback environment. The problem is important – every

45 seconds, someone in the United States suffers a stroke [16]. It results in functional deficits of neuropsychological and physical functions in post-stroke survivors. Up to 85% of patients have a sensorimotor deficit in the arm, such as muscle weakness, abnormal muscle tone, abnormal movement synergies, and lack of coordination during voluntary movement [9].

2.1 A Review of the Main Ideas

The primary goal of this project is the development of a real-time multimedia system that integrates task dependent physical therapy and cognitive stimuli within an interactive, multimodal environment. The environment provides a purposeful, engaging, visual and auditory scene in which patients can practice functional therapeutic reaching tasks, while receiving different types of simultaneous feedback indicating measures of both performance and results.

Biofeedback can be defined as the use of instrumentation to make covert physiological processes more overt while including electronic options for shaping appropriate responses [2,12,32]. The use of biofeedback allows the patient who has sensorimotor impairment to regain the ability to better discriminate a physiological response thereby better learning self-control of that response [20].

We now discuss related work on repetitive therapy for task training that involves multimodal processes to facilitate motor function recovery (e.g. reaching for a cup). Virtual reality (VR) is an emerging and promising technology for task-oriented biofeedback therapy [13,21]. The approach is ecologically valid [33]. Furthermore, it has been shown that task learning in a VR can be transferred into real world task performance [27]. VR can offer complex, highly detailed and engaging multimodal feedback in response to physical action. This has significant potential in augmenting traditional task-oriented therapy training. In VR, visual feedback is easily accomplished via computer graphics. Different devices have been used to generate various immersion effects. The 3D stereo vision can be generated by using special glasses with a disparity between right and left eye and binocular fusion [21,28]. The head-mounted display (HMD) or 3D monitor generates stereo vision by presenting the separate images with different perspectives to each eye [28,37].

There has been related work on virtual reality based techniques to perform the biofeedback intervention for functional task retraining [21,22,27,43]. Holden et al. [22] utilized VR to train reaching and hand orientation of stroke patients. A virtual mailbox with different slot height and orientation was presented to the patient. To put the “mail” into the slot, the patient has to reach the slot with correct hand orientation. A virtual “teacher mail” demonstrated the “desired” motion for patients to imitate. The design and implementation of a virtual kitchen used to practice common daily-living activities is proposed in [41]. Nine participants were recruited for testing [21]. Comparing before and after VR-based training, these subjects showed significant improvement in the Fugl-Meyer (FM) score, the Wolf Motor Function (WMF) score, and selected strength tests. However, no control

group was assigned in this study to compare the effect of VR based training with that of other therapy.

There are three key contributions of this project which builds upon our recent work [6]:

- *Analysis*: We develop domain specific, highly detailed analysis of hand movement analysis to promote therapy of the reaching functional task. We capture of hand dynamics using 12 labeled reflective markers on the arm. These markers are tracked via six high-speed infra-red cameras, and used to build an arm biomechanical model. The model allows us to compute arm and torso joint angles, as well as hand movement dynamics, including movement segmentation.
- *Feedback*: Development of three multimodal feedback environments, with increasing levels of complexity, and closely coupled to the three semantic action goals of *reach*, *open* and *flow*. The structure of the feedback environment and its relationship to the achievement of the goals are based on well established principles regarding the role and function of art [17]. The feedback images used are all well known paintings or photographs and the music played is based on well established rules of western classical music. The overall idea driving the mappings is that spatial and target information is better communicated through visuals and complex time series data is better communicated through audio [14]. *Reaching* is encouraged through the implied existence of a visual target, an image completion/reassembly task, a visual centrifuge effect pulling forward towards the target, and an accompanying musical progression that requires completion and encourages movement towards the implied target. *Opening* is encouraged through the control of a rich, resonant musical accompaniment. *Flow* is encouraged by pointillist sound clouds in the main musical line, flowing particles in the visuals, a smoothly swelling and dipping, wave-shaped musical accompaniment, promotion of synchrony of the involved musical lines and an overall selection of relaxing sound timbres and images.
- *Validation* metrics: We developed novel validation metrics for the reaching task, to determine if our semantic messages have been communicated well. We also developed a measure of stylistic consistency. First, we segment the whole trial offline into five parts: (a) reaction, (b) accelerating reaching, (c) decelerating reaching, (d) adjustment for grasping and (e) returning. We compute two spatial errors at the end of decelerating reaching: (a) distance from hand to target, and (b) hand orientation. Since our goal is to encourage subjects to grasp the target by nearly full arm stretching without torso compensation, arm opening is a key metric. In this chapter, we only focus errors with respect to two arm joint angles for evaluating arm opening: (a) *shoulder flexion* and (b) *elbow extension*. Reaching duration, the time between the beginning of reaching and onset of grasping, is easily obtained via the segmentation results. The flow error is related to the smoothness of speed curve of the hand marker. The flow error is computed by measuring the

smoothness of three speed curves: (a) hand marker speed, (b) shoulder flexion speed and (c) elbow extension speed. We represent the movement consistency by speed variance over several consecutive target reaching trials. The smaller the variance of speed, the higher is the consistency of the subject for reaching the target. In order to compute the speed curve variance, we first align the speed with the spatial coordinates. Then, we compute the speed variance over consecutive trials.

We conducted experiments with six unimpaired subjects to test if the multimodal environments communicate the semantics of action. All of the subjects were right handed adults. Each unimpaired subject did 85 trails over five environments. We have excellent experimental results that suggest that the biofeedback system has considerable therapeutic promise. For the unimpaired subjects the environment does not introduce any significant change in spatial error over the entire session. For all other error metrics there is a jump at the beginning of every set. However, there error *decreases* within each set, indicating that the subject is able to parse the messages encoded in the feedback.

2.2 Why is it difficult?

Activation of conscious sensorimotor integration during the therapy promotes neural plasticity for recovery of motor and cognitive function, especially in neural trauma patients, such as those with stroke and spinal cord injury [10,18]. The effectiveness of inducing neural plasticity for functional recovery from any therapeutic system is based upon the active participation of the patient. This consideration is critical for repetitive exercise type of therapy because it is a challenge for the subjects to remain attentive and motivated during a long and tedious session and they easily become physically and mentally tired. Furthermore, conscious sensorimotor integration requires participation and coordination of multitude sensory systems in addition to the motor systems and necessitates a system that holds attention through engagement of the subject.

In traditional neuromotor rehabilitation, biofeedback intervention has most often been associated with non-purposeful, single-joint movements. While isolated muscle activity may improve, functional improvements are rarely noted [42]. Therefore, it has been suggested that biofeedback therapy aimed at enhancing motor function should be task-oriented [39]. We now outline some key challenges:

- **Determination of feedback parameters:** In the conventional biofeedback intervention, the feedback parameter is localized on single muscle activity or joint movement. However, for multi-joint coordinated movement training, the feedback parameter that characterizes the dynamic movement may be a high-dimensional vector, with correlated dimensions.
- **Feedback relationships:** The second design challenge is that multiple biofeedback parameters may overwhelm the perception and cognition of neurologically injured patients who may also have psychological deficits.

2.3 Our Biofeedback System

We now present our biofeedback system. First, we discuss the functional task in our system. Second, we introduce the physical setup. Finally, we propose the system overview.

2.3.1 Functional Task

We shall first discuss the functional task in our biofeedback system. The functional task in our biofeedback system is *reaching out the right arm and grasping the target*. This function task includes three sub-goals: reaching, opening and flow.

- *Reaching*: reaching for and grasping the target successfully. We expect the subject to reach out for the target with minimum spatial error, correct hand orientation and speed while within the vicinity of the target.
- *Opening*: open the arm to reach for the target without shoulder/torso compensation. We expect the subjects to extend their joints appropriate to the target. While this is trivial for normal subjects, patients might not achieve this sub-goal by means of shoulder/torso movement compensation.
- *Flow*: reach for the target smoothly. We expect the subject to coordinate their arm movement while reaching and grasping for the target smoothly in a consistent way.

Therefore, our biofeedback environment encourages the subject to achieve these three sub-goals through beautiful audio/visual feedback.

2.3.2 Physical Setup

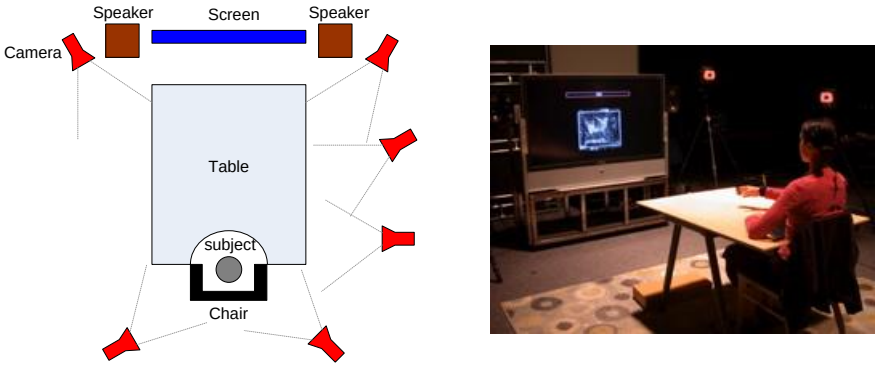

Figure 2. Physical setup of biofeedback system. Left: physical setup diagram. Right: snapshot of real environment used by a normal subject.

We now introduce the physical environment. Figure 2 shows the physical setup of our biofeedback system. The environment needs six motion capture cameras, two

speakers, one screen, one table and one chair. The subjects wear 12 markers on their arm, hand and torso and sits at one end of the table. At the other end of the table there are a big screen showing up the visual feedback and two speakers playing audio feedback. Six motion capture cameras are looking over the entire capture volume. Note that the six cameras are asymmetrically placed because we are focusing on rehabilitation of the right arm.

2.3.3 System Overview

In this section, we present an overview of the system. The Biofeedback system integrates seven computational subsystems: (a) Motion capture; (b) Task control; (c) Motion analysis; (d) Visual feedback; (e) Audio feedback; (f) Database for archival and annotation and (g) Visualization. All seven subsystems are synchronized with respect to a universal time clock. Figure 3 shows the system diagram.

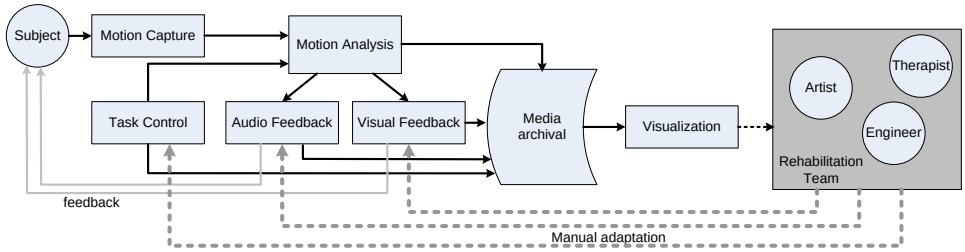

Figure 3. The biofeedback system diagram

The motion capture subsystem we are using is produced by Motion Analysis Corporation. We use six near-infrared cameras running at 100 frames per second to track the three-dimensional position of reflective markers that are placed on the subject. The task control subsystem provides a parameter panel where we can adjust the parameters related with the reaching and grasping task. The real-time motion analysis subsystem smoothes the raw sensing data, and derives an expanded set of task specific quantitative features. It multicasts the analyzed data to the audio, visual and archival subsystems at the same frame rate. The audio and visual subsystems adapt their auditory and visual response dynamically to selected motion features under different feedback environments. The archival subsystem continuously stores the motion analysis as well as the feedback data for the purpose of annotation and off-line analysis. Visualization subsystem visualizes the analysis results of subject's performance.

The rehabilitation team includes therapists, artists and engineers. They can fine-tune the system by adjusting the parameters in task control, audio feedback engine and visual feedback engine (e.g. moving target further in task control panel). The rehabilitation team adjusts the system by their observation, domain knowledge and help from the visualization subsystem that shows the subject's performance.

Our system situates participants in a multi-sensory engaging environment, where physical actions of the right arm are closely coupled with digital feedback. Participants are guided by our system to explore the novel environment. Through exploration, the participants begin to discover rules embedded in the environment. Those rules have been designed to couple action to feedback, consistent with the functional task. If the participants discover those embedded rules, the environment becomes stimulating, and aesthetically enjoyable.

3 Analysis of Action

In this section, we show how to analyze the subject's arm movement. We first discuss the raw feature extraction that includes arm representation and joint angle computation. Second, we transform the raw features into reaching/grasping task based features. Finally, we segment the trial into several phrases.

3.1 Raw Features of Arm Movement

We now show how to compute raw features of arm movement such as hand position and joint angles. We track the subject's arm movement by placing 12 three dimensional reflective markers on the subject's arm and torso. Through tracking the 3D positions of these 12 markers, we can obtain the subject's arm position and joint angles.

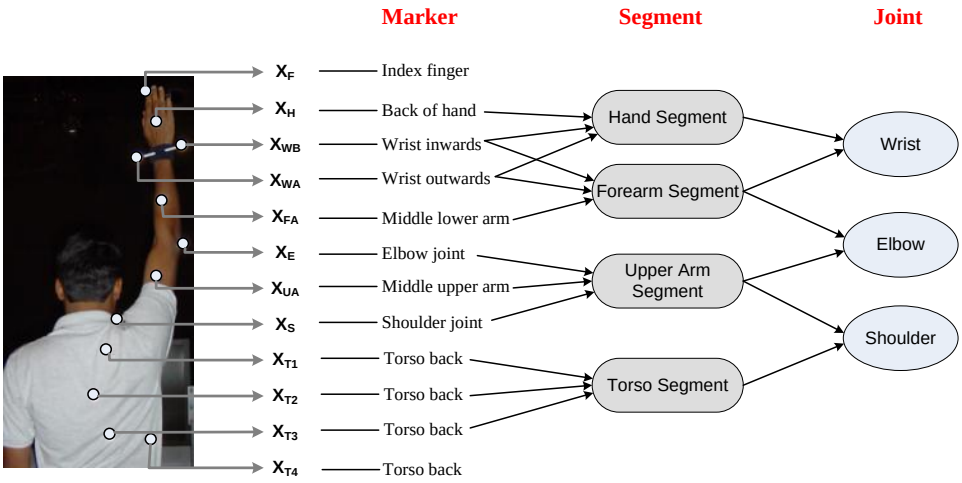

Figure 4. Arm representation by twelve 3D markers.

We use 12 labeled three dimensional markers to represent the arm and torso. Figure 4 shows the positions of these 12 markers placed on the subject's arm and torso. There are four markers on the back of torso, three markers on the upper arm

(shoulder, middle upper arm and elbow), three markers on the forearm (middle low arm, wrist inwards and wrist outwards), one marker on the hand and one marker on the finger. Each marker coordinate can be captured by the motion capture subsystem. A calibrated three dimensional capture system provides labeled data, specifying the location on the arm for each marker.

We can track hand position and all joint angles based on this marker setup. We use the maker placed on the hand (shown in Figure 4) to represent hand position. In order to achieve computational accuracy of joint angles with minimum complexity, each segment including the hand, lower arm, upper arm and torso has three non co-linearly positioned markers to construct a plane. The marker selection for segment plane construction is shown in Figure 4. The details of the joint angle calculations can be found in [7].

3.2 Task Based Features

We now show how to compute task based features. We divide the task based features into three groups which are associated with three sub-goals (ref. Section 2.3.1): *Reaching*, *opening* and *flow*.

3.2.1 Reaching

We use hand-target distance and hand orientation to represent subject's reaching and grasping sub-goal. We compute the hand-target distance with respect to three directions (X′, Y′, Z′) in local coordinate system. The origin of the local coordinate system is the 3D coordinate of subject's hand marker before subject starts a reaching trial (rest position). The Y′ axis, which is parallel to Y axis in global system, is perpendicular to the table plane and towards up. The X′-Z′ plane is parallel to the table plane. The Z′ axis is from the rest position to the target. The X′ axis is perpendicular to both Y′ and Z′. Figure 5 shows the relationship between global coordinate system (XYZ) and the local coordinate system (X′Y′Z′). At time slice *t*, we denote the hand-target distance along the three directions (X′Y′Z′) as $\Delta x(t)$, $\Delta y(t)$ and $\Delta z(t)$.

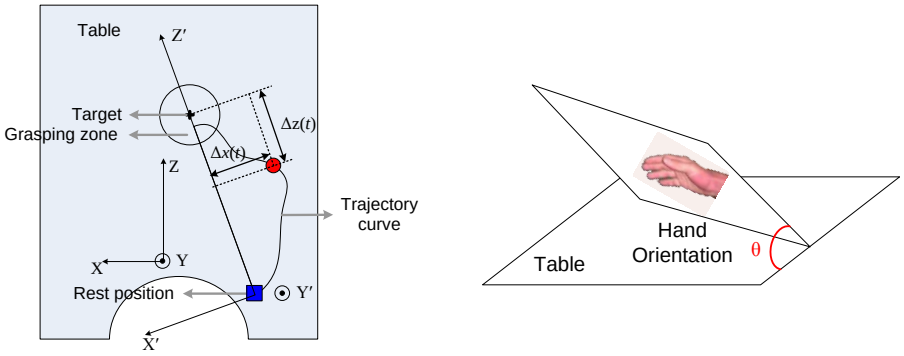

Figure 5. Reaching goal based features. Left: diagram for computing hand-target distance in local coordinate system (X′ Y′ Z′). Right: Hand orientation.

Another feature associated with reaching sub-goal is hand orientation. Hand orientation is defined as the angle between the table plane and hand plane (shown in Figure 5 (right)). It indicates if the subject rotates his/her hand to a proper angle with respect to grasping the target.

3.2.2 Opening

There are two kinds of features related to opening sub-goal: (a) opening of shoulder and elbow and (b) torso compensation. We expect the subject open his/her shoulder and elbow to reach for the target without compensation. The shoulder opening is represented by shoulder flexion angle that is the shoulder joint angle around X-axis. The X-axis is shown in Figure 5. The elbow opening is represented by elbow extension angle that is the angle between upper arm and forearm. The torso compensation is defined as the maximum of three torso rotation angles. The three torso rotation angles are rotation angles of torso segment from global coordinate system around X-Y-Z axis respectively.

3.2.3 Flow

The features related to flow sub-goal includes: (a) flow of hand movement and (b) flow of shoulder/elbow angular opening. The flow of hand movement is represented by hand velocity and hand acceleration. The flow of shoulder/elbow angular opening is represented by velocity and acceleration of shoulder flexion angle and elbow extension angle. We compute the velocity and acceleration using the difference equation on the smoothed raw data. Note that all task based features in this sub-section are computed for every frame (i.e. every 10 ms). In the following section, we shall discuss how to segment one reaching trial into several phases.

3.3 Real Time Trial Segmentation

We now present the trial segmentation in real time. We sequentially segment the reaching trial into six phases: (a) ready check, (b) reaction, (c) reaching, (d) grasping, (e) returning and (f) stop. Each phase has a unique ID which is sent to feedback generation engine. The feedback engine has different feedback design for different phase. We explain the segmentation in detail as follows:

- *Ready check*: Each trial starts with the ready check phase. In this phase, we check if the subject is ready for a reaching trial at the beginning.
- *Reaction*: In reaction phase, we keep checking if subject starts reaching for the target. If subject starts, the phase jumps to *reaching*.
- *Reaching*: The subject's phase stays in reaching phase until the subject achieves grasping successfully. We set three conditions for successful grasping: (a) subject's hand is in grasping zone (ref. Figure 6), (b) subject has proper hand orientation, and (c) hand speed is small. If all of these three conditions are satisfied for continuous 250 mille-seconds, the subject is considered to be successful in grasping and the system jumps to the grasping phase.
- *Grasping*: The grasping phase is kept for subject until the subject's hand crosses the return line (shown in Figure 6) and moves toward the rest position (i.e. crossing the return line from dark region to light region in Figure 6). After grasping phase, the next phase is returning.
- *Returning*: The subject's phase stays in returning phase until the subject stops hand movement and goes to stop phase.
- *Stop*: In stop phase, the system idles for a short duration. This allows the subject to take a rest and learn from the current reaching trial. The length of stop phase is random variable between one second and three seconds based on uniform distribution.

In Figure 6 (left), we illustrate the hand position on the table for the six phases (red dots) in one reaching trial where the red dot with label 1 is in the ready check phase and the dot with label 6 indicates the stop phase. Note that the subject can manually stop the trial at any time by hiding the hand marker with another hand.

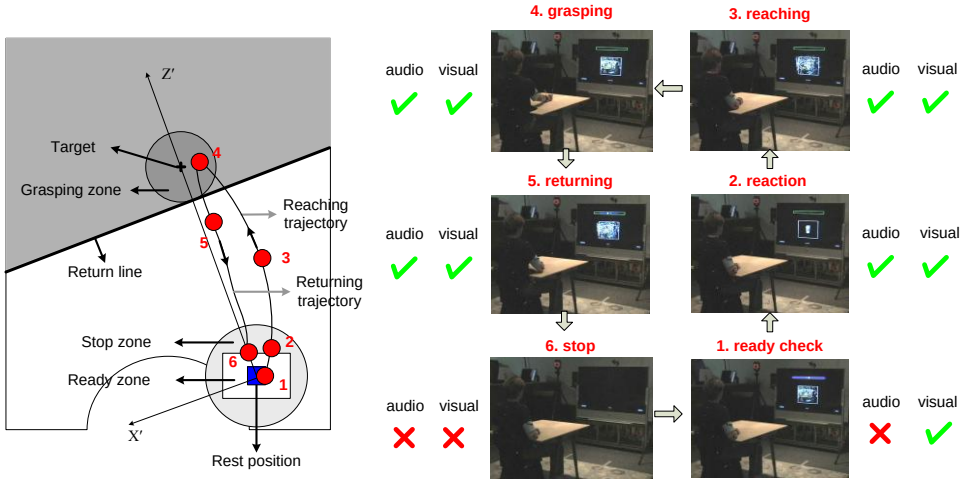

Figure 6 phase segmentation in a reaching trial. Left: several zones used for phase segmentation. The six red dots are hand positions for six phases (1–ready check, 2–reaction, 3–reaching, 4–grasping, 5–returning, 6–stop). Right: snapshots of subject's reaching trial for six phases. For each phase, we indicate the status of audio and visual feedback engine. "√" means that the feedback engine is turned on and "X" indicates off.

4 Coupling Action to Feedback

In this section, we present how to couple arm movement action to feedback generation. First, we shall discuss our intuitions under the design. Second, we shall explain how to trigger the audio and visual feedback in different phases in a reaching trial. Finally, we propose two abstract feedback environments.

4.1 Intuition

We structure the feedback so as to encourage the accomplishment of the movement goals (i.e. reach, opening and flow). The structure of the feedback environment and its relationship to the achievement of the goals are based on well established principles regarding the role and function of art [17].

To achieve initial engagement, the environment must be aesthetically attractive, easy to use and intuitive. Having attracted the attention of the patient the environment must maintain their attention through evolution of form and content. At the highest level of its structure the environment must communicate to the patient the messages that can encourage the accomplishment of the movement goals. These messages are: *reach*, *opening*, *flow*.

The feedback images used are all well known paintings or photographs and the music played is based on well established rules of western classical music. Thus

the content has a high probability of attracting and engaging the subjects and deepening their immersion in the experience.

The overall idea driving the mappings is that spatial and target information is better communicated through visuals and complex time series data is better communicated through audio [14]. Furthermore, a number of parallel streams of music can be handled simultaneously if assignment and mappings are psychoacousticaly correct and musically valid. The movement parameters allowing successful manipulation of the environment are the key parameters of an everyday reaching and grasping movement thus the environment can be easily connected in terms of action to its goal and does not require unintuitive movement learning that is an artifact of the interaction.

The abstract nature of the feedback environment helps patients practice their movement in an environment that can partially removed them or at least distance them from the everyday struggles they associate with movement of their arm. The attraction and engagement of the interactive environment becomes a dominant element and characterizes the experience. Above all, it reduces the tedious nature of repetitive physical therapy. At the same time, learning achieved in that environment can be directly transferred to every day movement tasks. The control of the interactivity, requires continuous participation by the patient which again raises their interest and engagement, reduces tediousness and promotes neural plasticity for sensorimotor integration.

The mappings and content follow a similar structural hierarchy as the movement parameters and goals with sub-message levels supporting the communication of each larger message. As is the case of movement parameters, there are feedback parameters that the subject can quickly understand and control, parameters that require practice to control and subconscious parameters supporting the achievement of the consciously controlled goals.

- *Reaching* is encouraged through the implied existence of a visual target, an image completion/reassembly task, a visual centrifuge effect pulling forward towards the target, and an accompanying musical progression that requires completion and encourages movement towards the implied target.
- *Opening* is encouraged through the control of a rich, resonant musical accompaniment.
- *Flow* is encouraged by pointillist sound clouds in the main musical line, flowing particles in the visuals, a smoothly swelling and dipping, wave-shaped musical accompaniment, promotion of synchrony of the involved musical lines and an overall selection of relaxing sound timbres and images.

A good balance of repetition and variation must be achieved to allow for learning while maintaining interest and reducing boredom. Although the overall task and feedback mapping structures, remains the same for a set of trial, the images and sounds used vary with each trial based on an algorithm that promotes gradual variation and avoids sudden changes that can produce tension. The complexity also gradually increases. As the subject concurs the obvious aspects of an environment they simultaneously become aware of lower levels of structures that they can attempt to master. When learning stabilizes in one of the feedback

environments a new set of trial is started where the complexity of the environment is increased.

4.2 Coupling Trial Phase to Feedback Engine Trigger

We now discuss the feedback trigger for six trial phase (ref. Section 3.3). The audio/visual feedback trigger for these six trial phases are shown in Figure 6 (right).

- In the stop phase, both audio and visual feedback engines are disabled (speaker is mute and screen shows black screen). Because the feedback engines are shut down, the subject can take a rest and rethink about the previous reaching trial without distraction.
- In the ready check phase of a reaching trial, the visual feedback is turned *on* but the audio feedback is turned *off*. The visual feedback in ready check transfers two things to subject: (a) subject's hand position with respect to the rest position and (b) subject's slouching. We design the visual feedback like this because we hope the subject starts from the rest position without slouching in the chair. Figure 7 shows the visual feedback diagram for ready check phase. The top bar indicates the subject's hand position with respects to the rest position. The bottom image indicates if the subject slouches. If the subject slouches, the image gets blurred.

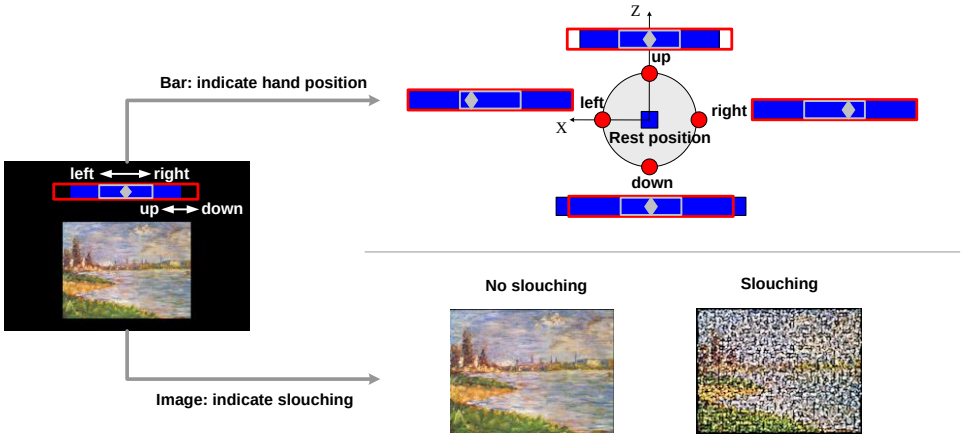

Figure 7. Visual feedback for ready check phase. The bar indicates the hand position with respect to rest position. The central diamond moves left and right from the center when subject's hand moves left and right with respect to the rest position. The blue bar shrinks or expends from red bar when subject's hand moves up and down the table, with respect to the rest position. The image indicates subject's slouching. If the subject slouches in the chair, the image gets blurred.

- In other four phases (i.e. reaction, reaching, grasping and returning), both audio and visual feedback engines are turned on. We shall discuss the feedback

environment design in terms of the coupling of action to audio/visual feedback for these four phases and the cooperation between audio and visual feedback in the following section.

4.3 The Feedback Environments

The transition from an actual reaching and grasping in the real world to reaching and grasping actions controlling an abstract multimodal environment is done gradually. We use a transition interactive environment that allows the necessary semantic and action transference and the necessary gradual acclimation.

The subjects start by performing reaching and grasping a physical cup that is on a table in front of them. After several trials, a representation of their arm, the table and the cup appears in front of them on the screen. The subjects are asked to reach and grasp as they are reaching for a real cup in front of them on the table. They quickly realize that their actions in the physical world are being duplicated in the virtual world and they acclimate to the mappings. After they have successfully reached and grasped the virtual cup a number of times they are ready to move to the abstract feedback environments.

4.3.1 Abstract Environment 1

The use of the abstract environment starts with the subject in the rest position. An image appears in the center of the screen and then explodes into particles that spread all over the screen. The frame of the image remains and a coffee cup appears in the center of the frame. Pointillistic sound clouds played on the marimba begin to sound. The subject is asked to move as if reaching for a regular cup on an actual table in front of them. By moving they realize that their movement is controlling all aspects of what they are hearing and seeing. An increase in engagement is achieved here simply though the realization of their level of control of this abstract, attractive experience.

The movement of the subject's hand outwards from their body allows the subject to collect the image particles into the frame, reassemble the image and make the cup disappear. Movement of the hand to the left, right, or up sways the particles in that direction. Hand orientation is mapped to image orientation. Movement of the hand forwards also controls the playing of the musical phrase. The sequence of notes being chosen follows a traditional, forward-moving musical progression that requires it be completed for the subject to hear a resolution, for the music to sound as if it has reached a resting point. If the subjects hand reaches the three dimensional target position and the arm is correctly supinated for grasping the image and the chord progression is completed telling the subject that the reaching and grasping has been successfully achieved. The percentage of particle reassembly and progression completion gives the subject a sense and measure of depth of movement. That is a key parameter for measuring completion of the reaching task and without the appropriate mappings it is a parameter that is

lost when moving from the real world to a two dimensional screen representation. Because music patterns provides a great tool for organization of time, the playing of the musical progression, also allows the subject to organize the timing of their forward movement. The velocity of the hand is mapped to density of notes of the musical cloud being played. This mapping promotes a memory of velocity at the level of a continuous contour rather than a sequence of individual musical events. Thus the subject can develop an integrated speed, time, space plan of action for their movement towards the target. Figure 8 show the coupling diagram from action to feedback.

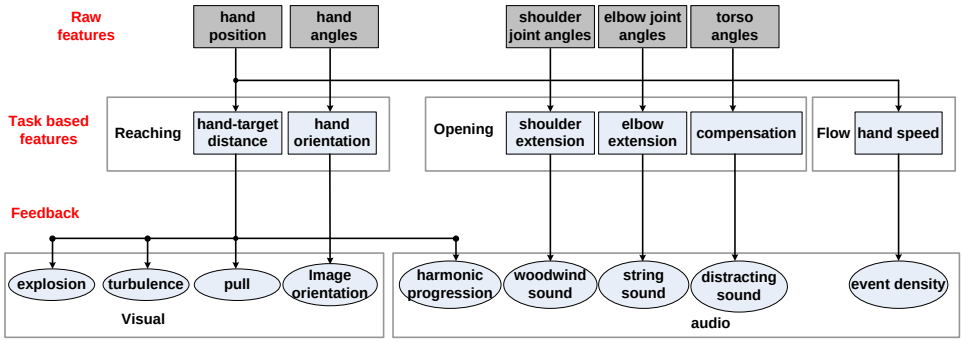

Figure 8. Diagram of coupling arm movement to feedback .

4.3.2 Abstract Environment 2

In the second test environment a richer musical accompaniment is added. The opening of the elbow introduces an accompaniment played by string instruments and the opening of the shoulder introduces an accompaniment played by winds. When the arm reaches full natural extension for grasping then the richest possible accompaniment is achieved. The synchrony of joints and hand movement controls the synchrony of music lines and music harmony. If their movement of the hand and joints are in phase the note content of the corresponding musical phrases is similar. If they are out of phase the selection of notes controlled by each joint and the hand are from different parts of the musical progression. This synchrony of chords, like joint synchrony, can only be controlled subconsciously especially by musically naïve subjects. (That is especially true during the reaching action that is evolving fast). When synchrony is achieved and the chords are in harmony the subjects knows it. However, when that is not the case conscious analysis in real time will offer little to the subject. Synchrony and the resulting harmony need to be achieved through experimentation.

Once the subject begins to experience the environment he/she begins to discover the mappings and decide on the control strategies. Further engagement is achieved by this active investigation of mappings.

As the subject begins to learn to perform in this interactive environment, his/her accuracy of end point trajectory, speed of completion of reaching task, smoothness of reaching, openness of joints increases. At this point, a feed forward effect begins to appear. The subject begins to develop a memory of the music and particle movement that are associated with a correct and effortless performance. The subject can use this memory to create a movement plan before starting the next reaching movement. Additional trials help further define that movement plan. The strong, intuitive, memory that people have for musical content strengthens further the creation of this feed forward plan.

5 Creating the Feedback

We now present our audio visual mapping frameworks.

5.1 Audio

In this section we discuss how normalized distance, velocity, synchrony and shoulder flexion is mapped to audio feedback.

5.1.1 Distance to target and Harmonic Progression

We now present the dynamic mapping of the normalized distance to target along the Z' coordinate to harmonic progression. Underlying both test environments is the same harmonic progression (in musical terms - I^{ma7} vi V^7/IV IV ii^7 V^7 I).

There are three key states of hand movement activity: reaching, grasping and returning. Through empirical testing we developed the following ranges of Percentage Z' to correspond to specific chords. These ranges are variable, as are the number of chords.

Table 1: Mapping of normalized distance to target in Z′ direction to harmonic progression

Activity	Z_N	Harmony
Reaching	0.00 - 0.19	I^{ma7}
	0.19 – 0.50	vi
	0.50 – 0.85	V^7/IV
Grasping	N/A	IV
Returning	0.63 – 1.00	ii^7
	0.29 – 0.63	V^7

	0.00 – 0.29	I

The pitches of each harmony are constrained to be between midi note values 44 – 80 (Ab2 – Ab5). These notes comprise a set which is randomly selected from at each event point.

It was observed that it was necessary to weight the selection of the root note of the chord in a lower octave so that the harmonic movement could be clearly perceived. Therefore, note selection was weighted so that probability of the root note occurring in the octave C2 – C3 was 5% more likely than any other note.

5.1.2 Hand Trajectory velocity to Event Density

We now show how to map the overall hand trajectory velocity to event density. The underlying pulse for the system was set at a constant rate of 92 beat per minute (bpm). It was decided that there would be 5 levels of event density, subdividing this pulse into 2,3,4,6 and 8.

The velocity of the hand in *Z* was first normalized to lie in [0, 1] and then mapped to these subdivisions as follows by identifying a velocity range to a pulse subdivision:

Table 2: Mapping of the hand trajectory velocity in Z direction to pulse subdivision. Pulse is 92 beats per minute.

Velocity Range	Pulse Subdivision
0.00 - .192	2
.192 - .410	3
.410 - .640	4
.640 - .780	6
.780 – 1.00	8

5.1.3 Joint Synchrony and Harmonic Progression

Before the start of each trial, a synchrony table was sent from the analysis engine that gave interpolated values for the shoulder angle and elbow angle aligned with percentage Z′ from the current starting point to the target position. The synchrony table is important because the precise relationships represent the coordination between the variables in the functional task of reaching. Only when the subjects can reproduce these variable relationships is the reference audio feedback reproduced. Once the trial started, the respective angles were used as an index into the table to find the corresponding value of percentage Z′.

The value for the shoulder angle was used to move woodwind sounds (flute, clarinet, bassoon) through the progression using the same method described above for the marimba. The elbow angle was similarly connected to string sounds (a violin section of tremolo, a violin section, and a pizzicato violincello section).

Each instrument was assigned a range in which it would randomly choose notes of the current chord. These ranges were as follows:

Table 3: Midi note range assigned for different instruments.

Instrument	Midi note range
Flute	72 – 86
Clarinet	58 – 72
Bassoon	36 – 60
Violin I (tremelo)	60 – 82
Violin II (sustained)	56 – 82
Violincello (pizzicato)	38 - 60

Event density, measured as subdivisions of the underlying pulse, was kept constant for each instrument, with all but Violin II using 2 subdivisions. Violin II used 4.

5.1.4 Mapping of Shoulder Flexion and Elbow Extension

There were three control parameters that the shoulder flexion and elbow extension were mapped to, midi velocity (M_v), duration (t_d) and the probability of an octave doubling (P_d) in the instrument occurring. The MIDI (Musical Instrument Digital Interface) protocol defines a specification for communicating musical events to and from hardware and software music applications. MIDI velocity is an indication of how loud a note should sound on a scale from 0 - 127.

Let x be the percentage of the current shoulder flexion between the starting angle and the expected angle at the target. $\sqrt{x}$ is used to interpolate between the following ranges:

Table 4: Midi velocity and duration range of 3 instruments connected to shoulder flexion.

Instrument	Reaching		Returning	
	M_v	t_d	M_v	t_d
Flute	0– 60	100 – 300	0 – 60	100 – 300
Clarinet	50–60*	200 – 600	50 – 60	200 – 600
Bassoon	0– 60	200 – 600	0 – 60	200 – 600

In all cases P_d range is [0,100]. In the case of the clarinet, if the value of x is 0, then midi velocity is set to 0, else the specified range is used. The elbow extension is mapped in a similar manner, we have omitted the details for the sake of brevity.

5.2 Visual

In this section we discuss how we create the visual feedback in the transition and the abstract environments.

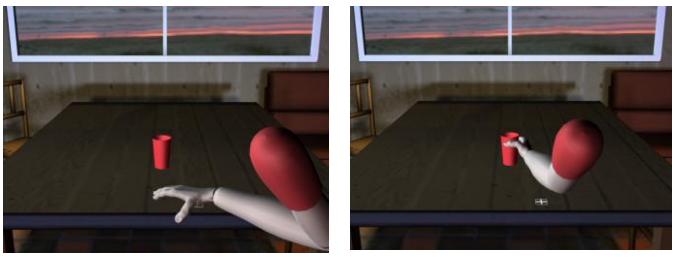

Figure 9. Transition Environment

5.2.1 Transition Environment

In the first environment we introduce the subject to the system, and the idea that their physical movement will control the virtual environment. A three dimensional arm model is transformed to the position of the optical motion capture markers. Fitting the predefined model to a subject in real time presents some challenge. We are using a limited marker set of just 12 optical motion capture markers. In addition, the markers are offset from the real bone joints that we are trying to use in our calculations of joint angles. If the markers are placed on clothing or muscle that moves, the relationship of the marker to joint can change, introducing error.

We provide a point of view that is similar to the subject's actual point of view in the motion capture volume. However, we move the camera slightly back and down, to give a clear view of the subject's arm. This viewpoint seems natural to the subject, while providing a better understanding of the arm movement than simply using the actual eye position.

5.2.2 Abstract Environment

In the abstract environment the subject is presented with a picture in a frame. The picture explodes into thousands of particles, and then the subject is able to reassemble the picture by completing the reaching and grasping movement (ref. **Figure 10**).

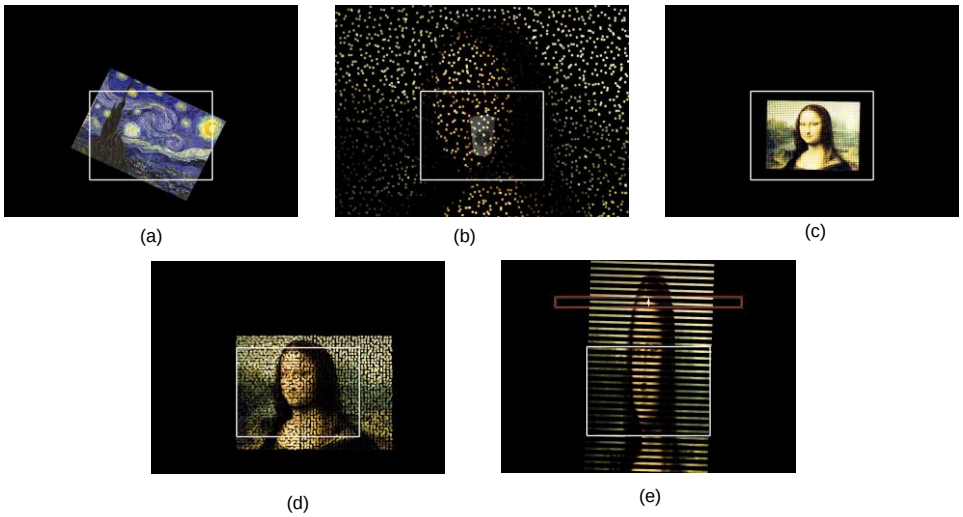

Figure 10. Visual feedback in the abstract environments. (a) rotation of image, (b) particles begin to form the image as the hand approaches the target, (c) the picture begins to collapse when the hand overshoots the target position, (d) image pulled to the right when subject is off target, (e) Vertical bands appear when the subject has wrong target height.

The image is broken into a 60x40 grid of particles. Each particle is a quad polygon with four vertices and four texture coordinates. The vertices locate the particle in three dimensional space, while the texture coordinates provide a two dimensional mapping to a color from the image. Each particle has an offsetIndex (P') that locates its relative original position in the picture:

$$\overrightarrow{P'}_x = \vec{P}_x - \frac{n_{col}}{2}, \quad \overrightarrow{P'}_y = \vec{P}_y - \frac{n_{row}}{2}, \qquad <1>$$

where P_x and P_y are the original position, P'_x and P'_y are the relative position of the particle in the image.

The motion of the particles (T_P) has five components: rotation angle (θ), and four motion vectors: explosion (T_E), turbulence (T_T), horizontal pull (T_{HP}) and vertical pull (T_{VP}). The position of a particle is calculated with a translation of the motion vectors followed by the rotation:

$$\overrightarrow{T_{\vec{P}}} = R_{(\theta,Z)}[\overrightarrow{T_E} + \overrightarrow{T_T} + \overrightarrow{T_{HP}} + \overrightarrow{T_{VP}}], \qquad <2>$$

where $R_{(\theta,z)}$ is the rotation along the z axis by θ.

5.2.2.1 Rotation Angles

Hand orientation controls the image rotation angle. The difference between expected hand orientation angles (i.e. θ_s^e) and subject hand orientation angles (i.e. θ_s) is mapped to image rotation (θ), and a scaling factor (α_s) and is applied:

$$\theta = \alpha_s(\theta_s^e - \theta_s). \qquad <3>$$

Figure 10 (a) shows the image rotations.

5.2.2.2 Explosion

The explosion of particles is controlled by movement towards the target position. As the z distance approaches 0, the particles return to their origin, thus reassembling the picture. A non-linear mapping function is used to control the explosion movement, so that the subject quickly begins to see the picture assemble. However, this means that the picture is mostly together before the z distance reaches 0.

Let us denote the normalized Z′ value as Z_N (Z_N=(Z_H-Z_R)/(Z_T-Z_R)). Z_H, Z_R and Z_T are Z′ coordinates of subject's hand, rest position (or hand starting position) of current trial and target position respectively. Z_N represents how far between the rest position (0.0) and the target position (1.0) the hand marker has traveled. If the subject reaches past the target position, the explosion effect is modified to collapse the image.

$$\overrightarrow{T_E} = \begin{cases} \beta_1 (1 - Z_N)^4 \overrightarrow{P'}, & 0 \le Z_N \le 1 \\ \beta_2 (Z_N - 1)^2 \overrightarrow{P'}, & Z_N \ge 1 \end{cases}, \qquad <4>$$

where β_1 is a explosion scale and β_2 is a collapse scale, *P'* is the relative position of the particle. Figure 10 (b) shows that particles begin to form the image as the hand approaches the target and Figure 10 (c) shows that the picture begins to collapse when the hand overshoots the target position

5.2.2.3 Turbulence

An additional, smaller turbulent motion is created with a Perlin noise function (N_P below). The turbulent motion is controlled by a linear mapping of the normalized Z′ distance (Z_N=(Z_H-Z_R)/(Z_T-Z_R)). This motivates the subject to complete the entire z movement.

$$\overrightarrow{T_T} = (1 - Z_N)(\lambda_1 N_P(\overrightarrow{P} + \overrightarrow{O}) + \lambda_2 \, t \, \delta N_P(\overrightarrow{P})), \qquad <5>$$

Where, λ_1 is the turbulence scale, λ_2 is the product of the noise scale and the octave scale, *t* is time, δ is noise speed. *O* is the phase offset.

5.2.2.4 Horizontal and Vertical Pull

Movement along the x axis away from the target causes a distortion in the particle movement on that side of the picture. If the subject strays to the right, the right size of the image will be spread out to the right.

$$\overrightarrow{T_{HP}} = [\eta(\mu X_H)^2 \quad 1] \overrightarrow{P'}, \qquad <6>$$

where μ is the *x* axis scale, η is the horizontal pull scale, *P'* is the relative position vector of the particle, X_H is the hand position along *x* axis. Extra y-axis movement is treated in a similar way. If the subject moves too high, the image will be spread upwards.

$$\overrightarrow{T_{VP}} = [1 \quad \chi C(\frac{Y_H}{\max(Y_H)}, 0, 1)] \overrightarrow{P'}, \qquad <7>$$

where C is a clamp function, Y_H is the hand position along y axis, χ is the vertical pull scale, *P'* is the relative position of the particle. Figure 10 (d) shows image pulled to the right when subject is off target and Figure 10 (e) shows that vertical bands appear when the subject has wrong target height

6 Validation Metrics

In this section, we shall discuss the validation metrics for evaluating the performance of the biofeedback system.

6.1 Offline Segmentation

First, we segment the whole trial offline into five parts: (a) reaction, (b) accelerating reaching, (c) decelerating reaching, (d) adjustment for grasping and (e) returning. Let us denote the whole trial duration as [0 T]. Because the target reaching trial is simple, we apply a simple segmentation algorithm based on the speed curve of the hand marker. Figure **11** shows an offline segmentation result based on the speed curve.

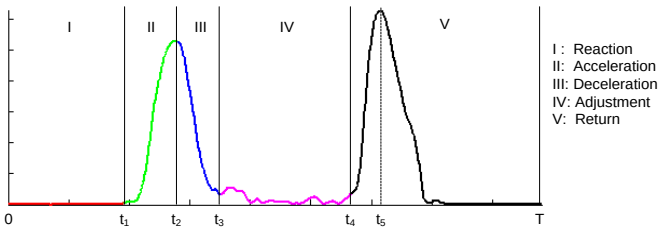

Figure 11: Trial segmentation based on speed curve

Reaction is the duration in which subjects prepare for reaching prior to moving the arm. The reaction time (t_1) is computed as the first time stamp such that the speed of next 700 ms (70 frames) is larger than a threshold:

$$t_1 = \min\{t^* \mid t^* \in [0,T], \forall t \in [t^*, t^* + \varepsilon], v(t) > \alpha\}, \quad <8>$$

where ε is 700 ms and α is the threshold (α=5mm/s).

In *accelerating reaching*, the subjects start the reaching trial with increasing speed. The accelerated reaching starts from reaction time t_1 and ends at the time t_2 with the first constrained local maximum speed. The accelerating end time t_2 is determined as:

$$t_2 = \min\{t^* \mid t^* > t_1, v(t^*) > \beta, v(t^*) = \max_{t^*-w \le t \le t^*+w} [v(t)]\}, \quad <9>$$

where β is a speed threshold and *w* is a local window size.

Decelerating reaching starts from t_2 until the time when a constrained local minimal speed is achieved. Constrained local minimum is just the local minimum

that is less than a predefined threshold. Thus the decelerating ending time t_3 is represented as:

$$t_3 = \min\{t^* \mid t^* > t_2, v(t^*) < \gamma, v(t^*) = \min_{t^*-w \le t \le t^*+w} [v(t)]\}, \quad <10>$$

where γ is speed threshold and w is local window size.

In *adjustment* duration, the subjects try to adjust their hand orientation to grasp the cup comfortably. The starting time of the adjustment duration is deceleration end time t_3. Before we obtain the end time of adjustment, we compute the last constrained local maximum speed and get the corresponding time stamp t_5 (see Figure **11**). Therefore, we can compute adjust ending time t_4. t_4 is corresponding to the nearest constrained local minimum before t_5. The last part, *returning*, starts from time t_4 until the end of trial.

6.2 Spatial Error

We compute two spatial errors at the end of decelerating reaching: (a) distance from hand to target, and (b) hand orientation. The normalized hand-target distance is computed as:

$$d_1 = \frac{\| X_h(t_3) - X_T \|_2}{\| X_h(0) - X_T \|_2}, \quad <11>$$

where X_h is the 3D position of the hand marker, X_T is the target position, t_3 is the decelerating ending time and $\|\cdot\|_2$ is L2 distance metric. The hand orientation error is defined as follows:

$$d_2 = \frac{|\theta_h(t_3) - \theta_T|}{\Theta_h}, \quad <12>$$

where $\theta_h(t_3)$ is the hand orientation angle at the decelerating ending time, θ_T is the desired hand orientation angle for grasping the target which is computed during calibration for every subject, Θ_h is predefined constant (Θ_h=75°). Here we use a predefined constant rather than the range of hand orientation because some subject's hand orientation angle at rest position is very close to the desired angle and hence the range of hand orientation during the trial is very small. The overall spatial accuracy of a target reaching trial is the linear combination of hand-target distance and hand orientation accuracy:

$$s = w_1^s \cdot d_1 + w_2^s \cdot d_2, \quad <13>$$

where w_1^s and w_2^s are two weights.

6.3 Arm Opening

Since our goal is to encourage subjects to grasp the target by nearly full arm stretching without torso compensation, arm opening is a key metric. In this

chapter, we only focus two arm joint angles for evaluating arm opening: (a) *shoulder flexion* and (b) *elbow extension*.

The shoulder opening and the elbow opening are defined as the relative error with respect to the desired shoulder flexion and elbow extension:

$$p_s = \frac{|\theta_s(t_3) - \theta_s^T|}{|\theta_s(0) - \theta_s^T|},\ p_e = \frac{|\theta_e(t_3) - \theta_e^T|}{|\theta_e(0) - \theta_e^T|}, \quad <14>$$

where p_s and p_e are the shoulder opening and elbow opening respectively, θ_s and θ_e are shoulder flexion angle and elbow extension angle respectively, t_3 is decelerating ending time, θ_s^T and θ_e^T are the desired shoulder flexion and elbow extension respectively. θ_s^T and θ_e^T are captured during the calibration. Both p_s and p_e are numbers between 0 and 1, zero meaning full opening and ones meaning no opening. Therefore the overall arm opening is defined as the linear combination of shoulder opening and elbow opening:

$$p = w_1^p \cdot p_s + w_2^p \cdot p_e, \quad <15>$$

where w_1^p and w_2^p are two weights.

6.4 Reaching Duration

Reaching duration, the time between the beginning of reaching and onset of grasping, is an important metric. As the subjects become more familiar with the system, their hesitation for reaching the target will decrease. Hence, the length of the reaching duration will decrease. Using the segmentation results, we can easily obtain the reaching duration by:

$$r = t_3 - t_1, \quad <16>$$

where t_3 is the decelerating ending time and t_1 is the reaction time.

6.5 Flow Error

In this section, we shall discuss the flow error of target reaching. Intuitively, the flow error is related to the smoothness of speed curve of the hand marker. The smoother the speed curve, the less the flow error. The organization of this section is as following: we first introduce two measurements of curve smoothness – (a) zero crossing number and (b) polynomial curve fitting error. Then we shall discuss the flow error measurement by combining three speed curves - (a) speed of hand marker moving, (b) speed of shoulder flexion angle and (c) speed of elbow extension angle.

6.5.1 Smoothness Metric

Let us denote the speed curve during reaching as $v(t)$, $t_1 \le t \le t_3$. The zero crossing number k is defined as the number of zero crossing of first order derivative of speed $v'(t)$. The smaller the zero crossing number, the smoother the speed curve. Another useful metric is the curve fitting error e_f which is defined as the square error between original curve and fitting curve. Before we compute the curve fitting error, we first normalize the curve by the maximum value.

$$v_N(t) = \frac{v(t)}{\max[v(t)]}, t \in [t_1, t_3]. \quad <17>$$

Then we divide the reaching duration into acceleration phrase and deceleration phrase due to the asymmetry of speed curve and fit the two phrases separately. Hence, the curve fitting error of speed curve is:

$$e_f = \int_{t_1}^{t_2} [v_N(t) - f(v_N(t)]^2 dt + \int_{t_2}^{t_3} [v_N(t) - f(v_N(t)]^2 dt, \quad <18>$$

where $v_N(t)$ is normalized speed curve, t_1, t_2 and t_3 are reaction time, acceleration ending time and deceleration time respectively, $f(\cdot)$ is curve fitting operator. In this chapter, we use polynomial curve with degree 3 to fit the speed curve. We combine the zero crossing number and curve fitting error as a smooth vector to represent the smoothness of reaching speed:

$$M = [k, e_f]^T. \quad <19>$$

6.5.2 Overall Flow Error

The overall flow error incorporates the smoothness of three speed curves: (a) hand marker speed, (b) shoulder flexion speed and (c) elbow extension speed. Let us denote the smooth vector of hand marker speed, shoulder flexion speed and elbow extension speed as $\boldsymbol{M}_h$, $\boldsymbol{M}_s$ and $\boldsymbol{M}_e$ respectively. The overall flow error $\boldsymbol{F}$ is represented as the linear combination of these three smooth vectors:

$$F = w_1^f \cdot M_h + w_2^f \cdot M_s + w_3^f \cdot M_e, \quad <20>$$

where w_1^f, w_2^f and w_3^f are constant weights.

6.6 Consistency

In this section, we shall discuss the movement consistency. We represent the movement consistency by speed variance over several consecutive target reaching trials. The smaller the speed variance, the higher the consistency of subject for reaching the target. In order to compute the speed curve variance, we first align the speed with the spatial coordinates. Then, we compute the speed variance over consecutive trials. Finally, we combine the hand marker speed, shoulder flexion speed and elbow extension speed together to obtain the overall consistency.

6.6.1 Spatial Alignment

We align the normalized speed of reaching phrase $v_N(t)$, $t_1 \le t \le t_3$ along the direction from rest position (starting position of the subject's hand) to the target position denoted as Z′ axis. First, we divide the space from rest position to target position along Z′ axis into *N* bins. For each bin, we can compute the mean of speed for each trial. For example, the mean speed of the i^{th} bin is:

$$\mu_i = \frac{\int_{t:z'(t)\in(z_{i-1},z_i]} v_N(t)dt}{\int_{t:z'(t)\in(z_{i-1},z_i]} 1dt}, \qquad <21>$$

where z_i is the upper bound of the i^{th} bin. Thus, the speed alignment can be represented by μ_i, i=1,…,N.

6.6.2 Speed Variance

Let us denote the spatial alignment representation of the k^{th} reaching trial as $\mu_{i,k}$, i=1,…,N. The speed variance of K consecutive trials is the average variance of K trials over all N bins:

$$\sigma^2_{N,K} = \frac{1}{N}\sum_{i=1}^{N}[\frac{1}{K}\sum_{k=1}^{K}(\mu_{i,k} - \frac{1}{K}\sum_{k=1}^{K}\mu_{i,k})^2], \qquad <22>$$

where $\sigma^2_{N,K}$ is speed variance of K trials using N bins spatial alignment.

6.6.3 Overall Consistency

Combining the speed variance of the hand speed, shoulder flexion speed and elbow extension speed, we can obtain the overall consistency for K reaching trials:

$$C_{N,K} = w_1^c \cdot \sigma^2_{N,K,h} + w_2^c \cdot \sigma^2_{N,K,s} + w_3^c \cdot \sigma^2_{N,K,e}, \qquad <23>$$

where $C_{N,K}$ is overall consistency over K reaching trials based on N bins spatial alignment, $\sigma^2_{N,K,h}$,$\sigma^2_{N,K,s}$ and $\sigma^2_{N,K,e}$ are hand speed variance, shoulder flexion speed variance and elbow extension speed variance respectively.

7 Validating the System

We present results on validation of the system design. Specifically we wish to determine whether the mechanisms for encoding the semantics of the movement (reach, open and flow), were successful. Additional results can be found in [7,8].

7.1 Experiment Setup

In order to determine the effect of the semantic encoding in the biofeedback system with respect to the arm functional task, we tested our design scenarios on able-bodied subjects. We recruited six able-body subjects to test our system. All recruited subjects were right handed adults. They were all unfamiliar with the designed system prior to the test. Every subject was tested once. Each subject's visit is defined as a session. Each session includes five sets. Each set contains many reaching trials. A trial starts from the appearance of virtual target and ends when the subject finishes reaching, grasping, and arm withdrawing. At the beginning and at the end of session, a set of reaching to a physical cup was used to obtain the baseline performance of each subject prior to and after the test. There are 15 reaching trials to the physical cup, 15 to the transition environment (ref. Figure 9), 20 in each abstract environment (ref. Figure 10), followed by 15 reaching trials in the physical environment.

7.2 Test Results and Discussion

We now show the validation over 85 trials with 5 different environments. Each figure shows the average measure of six normal subjects. It is easy to find that the first trial in abstract environment I introduces large error for every metric. This is because the abstract environment is totally different with real cup reaching in physical world and transition environment. In the first trial, the subjects try to explore the space and understand the mapping between visual-audio feedback and their movements.

In Figure 12 (a), we can see the spatial errors of the abstract environment are at the same level with real world reaching. This suggests that our visual-audio feedback design can guide the normal subjects to do the reaching as accurately as they did in real world. We also see that the first trial of transition environment and abstract environment II do not introduce much error, this is because each of them does not introduce big difference compared with previous environment. However, the last real cup reaching which is so different with previous abstract environment does not bring errors. This is reasonable, since for the normal subjects, the everyday experience dominates short-time learning. We can also find that in the transition and abstract environment I, II, the spatial error keep decreasing slightly which reflects the subject's learning.

In Figure 12 (c), we find that the arm opening error keeps decreasing in the transition environment and abstract environment I and II and the arm opening error of abstract environment II is even less than the real cup reaching. This suggests that our chord design in audio feedback for the abstract environment II communicates the opening message to the subjects very well.

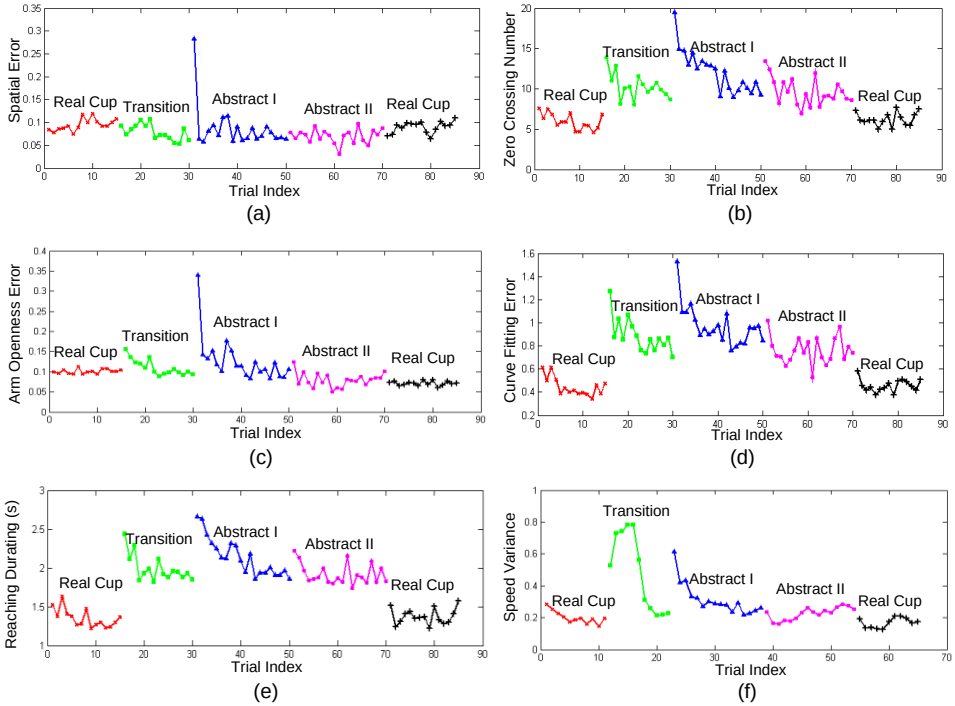

Figure 12. Average validation results of six normal subjects. (a) spatial error, (b) zero crossing number, (c) arm openness error, (d) curve fitting error, (e) reaching duration and (f) speed variance.

The length of reaching duration is show in Figure 12 (e). It is very clear that at the beginning of three biofeedback environments (transition environment and abstract environment I and II) the reaching duration length increases, this is because new information is introduced when changing the environment. Also, we can find that the reaching time keeps decreasing with each biofeedback environment. This reveals the subject's learning curve when playing with the system. This figure also shows us that at the end of each biofeedback environment, the reaching duration stays at about 2 seconds and there is a visible gap between biofeedback system and real world reaching. We conjecture that this gap is due to the tremendous unbalanced memory between real world reaching and our biofeedback system.

In Figure 12 (b) and Figure 12 (d), we show the flow error by two metrics: zero crossing number and curve fitting error. We can see that both zero crossing number and curve fitting error are decreasing in three feedback environments. This means the subject's velocity becomes smoother. In transition environment, the smoothness indicates that they find the mapping between their arm and virtual arm even if they have no information about the depth. In the abstract environment, the smooth speed curve implies three things: (a) the subjects are clear about the goal without hesitation. (b) the feedback cue is very clear for the subjects. Based on the feedback cue and their memory, they can easily find the way to reach the

target. (c) the mapping between the hand velocity and pulse subdivision in audio feedback works well in guiding the subjects to reach the target smoothly without looking at the target and their arms.

In Figure 12 (f), we can see the speed variance decreasing in the transition environment and abstract environment I. For the transition environment, the variance decreases because it is very similar with the real world. For abstract environment, reaching the target with consistent speed needs strong cue since the target and the arm are not present. This indicates that our feedback design enables the subjects to achieve a stylistic consistency of action.

8 Challenges and Opportunities for Experiential Media Design

Experiential media systems are challenging to build but provide significant opportunities for researchers to develop new theoretical and applied frameworks.

1. *Integration of knowledge*: A key aspect of developing experiential media systems lies in the recognition that the knowledge required in developing such systems, exists at the intersection of different disciplines invested in the human experience. These include engineering, the arts, psychology, education as well as architecture and design, to name a few. Working across disciplines requires patience, investment in the common problem, and the development of a common language in which to describe the research problems.

2. *New validation techniques*: Experiential media systems pose unique challenges in system validation. These are typically complex systems with many interconnected computational components, with the total number of systems parameters that can be tuned, running to the hundreds to a few thousand parameters; for example our biofeedback system has nearly 800 system parameters that can be set. Since the goal of the system is to transfer knowledge, there are many different combinations of parameters that can be legitimately set, making it a combinatorial problem. Typical strategies in engineering and psychology that call for changing one parameter at a time while measuring the effect on the transfer of semantics are not practical. Secondly, we need new strategies that move validation beyond the typical sum-of-parts scenario found in engineering wherein each part is optimized separately. Instead we need new metrics that measure the overall quality of the experience, instead of only measuring error.

   It is difficult to overemphasize the importance of experiencing the system, and the importance of that experience for measuring success. It is likely that the effect of such systems are best understood using both qualitative (e.g. observational - such as by the therapist in our system, more generally ethnographic) as well as quantitative measures.

3. *Feedback design*: The problem of encoding knowledge using media, from data collected from human activity in real-time, is a key and challenging problem. The encoding needs to be engaging, intuitively understandable and immersive. While the knowledge from the arts and design provides an important entry point to experiential feedback design, the existing formalisms were never developed keeping in mind the ability of the activity of the *audience* to be analyzed and tracked in real-time using computational means. Experiential media systems blur the distinction between the consumer and creator of the media artifact. This new paradigm can be addressed by understanding and tracking the user context, and allows for the development of new computational feedback formalisms that generalize beyond a specific system.

4. *Computational System design*: Experiential media systems can be understood that real-time feedback control systems. There are important and challenging research questions here on the overall design (e.g. what are effective information architectures that support the transfer of knowledge?) and system development (e.g. there are no application programming interfaces (API) for the system as a whole – instead the system is developed by a network of interconnected modules, each with its own set of API's). There is an interesting dual between experiential media systems and reinforcement learning (e.g. the work of Kaelbeling in robot navigation [36]). However, a significant challenge lies in the observation that we are interested in transferring semantics via feedback, not just measuring error. Hence re-evaluation of the objective function to ensure successful transfer of meaning is a challenge. Other questions include determining if the system is semantically stable, or divergent (i.e. is it possible that the overall system is on the wrong path, and transferring knowledge that is either not useful to the learning task, or which makes the overall task much harder?).

All these issues are complex, involve long term research questions, and provide exciting opportunities for researchers interested in the field.

9 Conclusion

Experiential media systems refer to real time, physically grounded multimedia systems in which the user is both the producer and consumer of meaning. These systems require embodied interaction on part of the user to gain new knowledge. A crucial test of the success of the experiential media system lies in its ability to scale beyond the system into the world itself – i.e. successful experiential media systems enables its users to learn, and use the learned knowledge as part of their real-world tasks.

The biofeedback system is a highly specialized experiential media system where the knowledge that is imparted refers to a functional task – the ability to reach and grasp an object. The narrow definition of the task provides us with an excellent example to study the design of experiential media systems in general.

In this chapter, we presented our efforts to develop a real-time, multimodal biofeedback system for stroke patients. There were several key contributions: we showed how to derive critical motion features using a biomechanical model for the reaching functional task. Then we determined the formal progression of the feedback and its relationship to action. We showed how to map movement parameters into auditory and visual parameters in real-time. We developed novel validation metrics for spatial accuracy, opening, flow and consistency. Our real-world experiments with normal subjects show we are able to communicate key aspects of motion through feedback, with excellent results. Importantly they demonstrate the messages encoded in the feedback can be parsed by the unimpaired subjects.

The design of experiential media systems is challenging, but provides significant and novel research opportunities in multimedia computing. New theoretical frameworks that deal with the system as a whole, as well as new validation metrics are needed.

10 References

[1] G. D. Abowd, E. D. Mynatt and T. Rodden (2002). *The human experience [of ubiquitous computing]*. IEEE Pervasive Computing **1**(1): 48-57.

[2] J. V. Basmajian (1989). Biofeedback : principles and practice for clinicians. Williams & Wilkins 0683003569 Baltimore.

[3] R. A. Brooks (1991). *Intelligence Without Reason*, International Joint Conference on Articial Intelligence, pp. 569-595, Aug. 1991, Sydney, Austrailia.

[4] R. A. Brooks (1991). *Intelligence without representation.* Artificial Intelligence **47**(1-3): 139-159.

[5] R. A. Brooks, M. Coen, D. Dang, J. Debonet, J. Kramer, T. Lozano-Perez, J. Mellor, P. Pook, C. Stauffer, L. Stein, M. Torrance and M. Wessler (1997). *The Intelligent Room Project*, Proceedings of the Second International Cognitive Technology Conference (CT'97), Aug. 1997, Aizu, Japan.

[6] Y. Chen, H. Huang, W. Xu, R. Wallis, H. Sundaram, T. Rikakis, J. He, T. Ingalls and L. Olson (2006). *The Design Of A Real-Time, Multimodal Biofeedback System For Stroke Patient Rehabilitation*, SIG ACM Multimedia, Oct. 2006, Santa Barbara, CA.

[7] Y. Chen, H. Huang, W. Xu, R. I. Wallis, H. Sundaram, T. Rikakis, T. Ingalls, L. Olson and J. He (2006). *The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation*, Proc. of the 14th annual ACM international conference on Multimedia, 763-772, Oct. 2006, Santa Barbara, CA, USA.

[8] Y. Chen, W. Xu, H. Sundaram, T. Rikakis and S.-M. Liu (2007). *Media Adaptation Framework in Biofeedback System for Stroke Patient Rehabilitation*, Proceedings of the 15th annual ACM international conference on Multimedia, ACM Press, Sep. 2007, Augsburg, Germany.

[9] M. C. Cirstea, A. B. Mitnitski, A. G. Feldman and M. F. Levin (2003). *Interjoint coordination dynamics during reaching in stroke.* Exp Brain Res **151**(3): 289-300.
[10] M. L. Dombovy (2004). *Understanding stroke recovery and rehabilitation: current and emerging approaches.* Curr Neurol Neurosci Rep 2004 **4**(1): 31-35.
[11] P. Dourish (2001). Where the action is : the foundations of embodied interaction. MIT Press 0262041960 (alk. paper) Cambridge, Mass. ; London.
[12] E. Dursun, N. Dursun and D. Alican (2004). *Effects of biofeedback treatment on gait in children with cerebral palsy*. Disabil Rehabil **26**(2): 116-120.
[13] J. Gallichio and P. Kluding (2004). *Virtual Reality in Stroke Rehabilitation: Review of the Emerging Research.* Physical Therapy Reviews **9**(4): 207-212.
[14] C. Ghez, T. Rikakis, R. L. Dubois and P. Cook (2000). *An Auditory display system for aiding interjoint coordination*, Proc. International Conference on Auditory Display, Apr. 2000, Atlanta, GA.
[15] J. Gray (2003). *What next?: A dozen information-technology research goals.* Journal of the ACM **50**(1): 41-57.
[16] G. E. Gresham, P. W. Duncan and W. B. E. A. Stason (1996). Post-Stroke Rehabilitation/Clinical Practive Guideline. Aspen Publishers, Inc. 30-010-00 Gaithersburg, Maryland.
[17] D. J. Grout and C. V. Palisca (2001). A history of western music. Norton 0393975274 New York.
[18] H. Woldag, G. Waldmann, G. Heuschkel and H. Hummelsheim (2003). *Is the repetitive training of complex hand and arm movements beneficial for motor recovery in stroke patients?* Clin Rehabil 2003 Nov **17**(7): 723-730.
[19] X. He, W.-Y. Ma, O. King, M. Li and H. Zhang (2003). *Learning and Inferring a Semantic Space from User's Relevance Feedback for Image Retrieval*. IEEE Transactions on Circuits and Systems for Video Technology.
[20] E. R. Hilgard and G. H. Bower (1975). *Recent developments*. Theories of learning(eds). Englewood Cliffs, N.J.,, Prentice-Hall**:** 550-605.
[21] M. Holden and T. Dyar (2002). *Virtual environment traing: a new tool for neurorehabilitation.* Neurology Report **26**(2): 62-72.
[22] M. Holden, E. Todorov, J. Callahan and E. Bizzi (1999). *Virtual environment training imporves motor performance in two patients with stroke: case report.* Neurology Report **23**(2): 57-67.
[23] J. Hollan, E. Hutchins, D. Kirsh and A. Sutcliffe (2000). *Distributed cognition: toward a new foundation for human-computer interaction research On the effective use and reuse of HCI knowledge.* ACM Trans. Comput.-Hum. Interact. **7**(2): 174-196.
[24] E. Hutchins (1995). Cognition in the wild. MIT Press 0262082314 Cambridge, Mass.
[25] H. Ishii and B. Ullmer (1997). *Tangible bits: towards seamless interfaces between people, bits and atoms*, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, 234--241,
[26] H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer and P. Yarin (1998). *ambientROOM: integrating ambient media with architectural space*, CHI 98 conference summary on Human factors in computing systems, ACM Press, 173--174,
[27] D. Jack, R. Boian, A. S. Merians, M. Tremaine, G. C. Burdea, S. V. Adamovich, M. Recce and H. Poizner (2001). *Virtual reality-enhanced stroke rehabilitation.* IEEE Trans. Neural Syst Rehabil Eng **9**: 308-318.

[28] R. V. Kenyon, J. Leigh and E. A. Keshner (2004). *Considerations for the future development of virtual technology as a rehabilitation tool.* J Neuroengineering Rehabil **1**(1): 13.

[29] D. Kirsh (1995). *The intelligent use of space.* Artificial Intelligence **73**(1-2): 31-68.

[30] Y.-F. Ma and H.-J. Zhang (2003). *Contrast-based image attention analysis by using fuzzy growing*, Proceedings of the eleventh ACM international conference on Multimedia, 1-58113-722-2, ACM Press, 374-381, Nov. 2003., Berkeley, CA, USA.

[31] A. Mazalek, G. Davenport and H. Ishii (2002). *Tangible viewpoints: a physical approach to multimedia stories*, Proceedings of the tenth ACM international conference on Multimedia, ACM Press, 153--160,

[32] J. Moreland and M. A. Thomson (1994). *Efficacy of electromyographic biofeedback compared with conventional physical therapy for upper-extremity function in patients following stroke: a research overview and meta-analysis.* Phys Ther **74**(6): 534-543; discussion 544-537.

[33] M. T. Schultheis and A. A. Rizzo (2001). *The application of virtual reality technology for rehabilitation.* Rehabilitation Psychology **46**: 296-311.

[34] Y. Sun, H. Zhang, L. Zhang and M. Li (2002). *A System for Home Photo Management and Processing*, Proceedings of the 10th ACM international conference on Multimedia, pp. 81-82, Dec. 2002, Juan Les-Pins, France.

[35] H. Sundaram and S.-F. Chang (2000). *Determining Computable Scenes in Films and their Structures using Audio-Visual Memory Models*, Proc. Of ACM International Conference on Multimedia 2000, pp. 95-104, Nov. 2000, Los Angeles, CA, USA.

[36] G. Theocharous, K. Murphy and L. P. Kaelbling (2003). *Representing hierarchical POMDPs as DBNs for multi-scale robot localization*, Workshop on Reasoning about Uncertainty in Robotics, International Joint Conference on Artificial Intelligence, Acapulco, Mexico.

[37] M. Tidwell, R. S. Johnston, D. Melville and T. A. Furness (1995). *The virtual retinal display-a retinal scanning imaging system*, Proceeding of Virtual Reality World' 95, 325-333, Heidelberg.

[38] B. Ullmer and H. Ishii (2000). *Emerging Frameworks for Tangible User Interfaces.* IBM Systems Journal **39**(3 & 4): pp. 915-931.

[39] J. P. Wann and J. D. Turnbull (1993). *Motor skill learning in cerebral palsy: movement, action and computer-enhanced therapy.* Baillieres Clin Neurol **2**(1): 15-28.

[40] M. Weiser (1993). *Some computer science issues in ubiquitous computing.* Commun. ACM **36**(7): 75--84.

[41] D. White, K. Burdick, G. Fulk, J. Searleman and J. Carroll (2005). *A virtual reality application for stroke patient rehabilitation*, IEEE International Conference on Mechatronics & Automation Niagara Falls, July 2005, Canada.

[42] S. L. Wolf, P. A. Catlin, S. Blanton, J. Edelman, N. Lehrer and D. Schroeder (1994). *Overcoming limitations in elbow movement in the presence of antagonist hyperactivity.* Phys Ther **74**(9): 826-835.

[43] S. H. You, S. H. Jang, Y. H. Kim, M. Hallett, S. H. Ahn, Y. H. Kwon, J. H. Kim and M. Y. Lee (2005). *Virtual reality-induced cortical reorganization and associated locomotor recovery in chronic stroke: an experimenter-blind randomized study.* Stroke **36**(6): 1166-1171.