The Design of a Real-Time, Multimodal Biofeedback System for Stroke Patient Rehabilitation

Yinpeng Chen, He Huang, Weiwei Xu, Richard Isaac Wallis, Hari Sundaram,

Thanassis Rikakis, Todd Ingalls, Loren Olson, Jiping He

Arts Media and Engineering Program

Arizona State University

E-mail: {yinpeng.chen, he.huang, wwxu, richard.wallis, hari.sundaram, thanassis.rikakis, todd.ingalls, loren.olson, jiping.he}@asu.edu

ABSTRACT

This paper presents a novel real-time, multi-modal biofeedback system for stroke patient therapy. The problem is important as traditional mechanisms of rehabilitation are monotonous, and do not incorporate detailed quantitative assessment of recovery in addition to traditional clinical schemes. We have been working on developing an experiential media 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. There are three contributions of this paper – (a) identification of features and goals for the functional task (b) The development of sophisticated feedback (auditory and visual) mechanisms that match the semantics of action of the task. We additionally develop novel action-feedback coupling mechanisms. (c) New metrics to validate the ability of the system to promote learnability, stylization and engagement. We have validated the system for nine subjects with excellent results.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Health; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *auditory (non-speech) feedback, screen design, interaction styles*; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis

General Terms

Performance, Design, Experimentation, Human Factors

Keywords

Biofeedback, Analysis, Action-feedback coupling, Validation

1. INTRODUCTION

The goal of this paper is to design a real time multimodal biofeedback system for stroke patient rehabilitation. The problem

is important – every 45 seconds, someone in the United States suffers a stroke [5]. 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 [2]. Biofeedback can be defined as the use of instrumentation to make covert physiological processes more overt while including electronic options for shaping appropriate responses. 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 [7].

We discuss related work on repetitive therapy for task training that involve 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 [8]. It 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. Holden et al. [9] 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. Nine participants were recruited for testing [8]. 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.

In studies to investigate the VR-based biofeedback for the benefit of hand function rehabilitation of stroke survivals, investigators used multisensing data from either the Cyberglove that sensed finger joint angles or the RMII glove that measured both the applied force under each finger and the position of fingertips [11]. Different scenarios were designed for exercises to improve joint range of motion, finger fractionation, and grasp strength on the impaired hand. The patients improved grasping force, finger joint range of motion, and movement speed after two weeks of VR-based biofeedback therapy.

We now present the key contributions of this paper:

- *Analysis*: We developed domain specific, highly detailed motional analysis to promote therapy of the reaching functional task.

- *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*.

- *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.

- *Experiments*: We conducted experiments with nine subjects (non-impaired), to test if the multimodal environments communicate the semantics of action. We have excellent experimental results.

The rest of the paper is organized as follows. In the next section we discuss the key challenges with the biofeedback problem. In Section 3 we present the analysis of action. In Sections 4,5 and 6 we present our feedback framework. In Section 7, we present our validation framework and in Section 8 we present our conclusions.

2. THE BIOFEEDBACK PROBLEM

In this section we present key challenges in the biofeedback problem, introduce our current biofeedback environment and finally summarize our prior work on stroke patient rehabilitation with a multimodal system.

2.1 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. 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. Therefore, it has been suggested that biofeedback therapy aimed at enhancing motor function should be task-oriented. 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.

2.2 An overview of the Environment

The Biofeedback system integrates five computational subsystems: (a) Motion capture; (b) Motion analysis; (c) Audio feedback; (d) Visual feedback; and (e) Database for archival and annotation. All five subsystems are synchronized with respect to a universal time clock. Figure 1 shows the 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 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 with universal timestamp, for the purpose of annotation and off-line analysis.

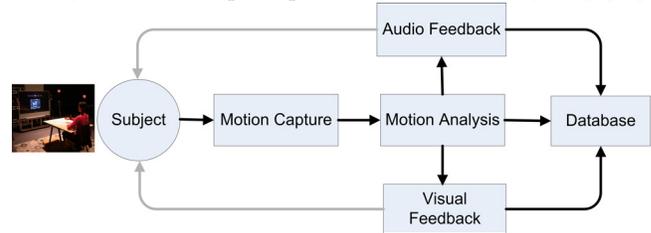Our system situates participants in a multi-sensory engaging

Figure 1: The biofeedback system diagram.

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.

2.3 Our Initial attempt at Biofeedback

Our previous study has designed an interactive multimodal environment (IME) based biofeedback system for repetitive reaching and grasping retraining [10]. A virtual living room was presented through 2D screen. While a patient tries to reach a virtual teapot within the scene, the virtual arm animates the real arm movement in real time based upon the inputs of angular and position sensors. The design assigns feedback from different sensory modality (visual or auditory) so as to take full advantage of the affordance (strengths) of each sense. For example, the visual perception is sensitive for spatial information [3]. Hence, visually, a 3D cone shape is displayed to guide the patient to reduce the spatial error between the patient's hand position and the "ideal" trajectory, a line from the start position to the target. If the spatial error is large and the hand moves out the boundary of the guiding cone (spatial limits), the transparency of the cone becomes reduced, i.e., the cone is more visible, to produce the knowledge of performance that informs the patient to correct the error. An innovative feature of this system was the design of an auditory feedback module that gave subjects knowledge of their performance.

Five hemiparesis patients secondary to stroke were tested using the designed IME biofeedback system. The results show that patients could perceive assigned biofeedback parameters. The visual augmented feedback improved the spatial consistency of the endpoint position during reaching. The auditory augmented feedback contributed improvement of the smoothness of endpoint trajectory, and the spatiotemporal consistency of reaching performance. After 3-5 training sessions, patients indicated faster, smoother, and more applied joint range of motion while reaching.

The results were encouraging. The further considerations to improve the system design were as follows:

- Increase the number of parameters mapped beyond just the endpoint trajectory.
- Variety of design environments and training programs is required to keep patients engaged. We did not consider complex formal properties of the environment.

3. ANALYSIS OF ACTION
In this section, we discuss our arm representation framework, extraction of joint angles, and our multi-goal approach.

3.1 Arm Representation by 3D Markers
We use 12 labeled three dimensional markers to represent the arm and torso. There is one marker on the finger and another marker on the hand to capture hand movement and palm joint angle; two markers on the wrist to capture wrist joint angles; four makers on the arm as shown below to capture elbow and shoulder joint angles; another four markers on the back of torso to capture torso joint angles. 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. 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.

3.2 Feature Extraction
In this section we will discuss our hierarchical feature extraction framework. The calibrated motion capture system has a global coordinate system. Before feature extraction, we convert the global system to subject's local coordinate system (XYZ). We also construct another local coordinate system (X'Y'Z') centered with respect to the start point of the hand and along to the target.

Table 1: Hierarchical Motion Features.

Level	Derived Motion Features
1	3D hand trajectory / 3D hand trajectory relative to the predefined straight line
2	Shoulder/elbow extension and hand orientation
3	Shoulder rotation/abduction/elevation
4	Trunk flexion/rotation/lean and shoulder trajectory
5	Forearm supination
6	Wrist extension

Based on domain knowledge of the reaching and grasping arm action, we derive and group motion features from the raw motion sensing data into six levels, with level 1 being the most important (Table 1 shows the hierarchy.). Within each level we also compute both the first and second time derivatives of features (velocity and acceleration).

To compute arm and torso joint angles, we apply the following general algorithm. First of all, a 7-DOF (degree of freedom) arm biomechanical model is built (2-DOF on wrist joint, 2-DOF on elbow joint and 3-DOF on shoulder joint). Each DOF is represented by a joint angle which can be obtained from the rotation matrix between two segments connected by the joint. The rotation matrix of two connected segments can be obtained by following steps using the 3D marker positions.

First, we compute the rotation matrix from the normal coordinate system to the marker-based local coordinate system as follows:

$$^{N}R^{MA} = [M_x, M_Y, M_Z] ,$$ <1>

where $^{N}R^{MA}$ is rotation matrix from the normal coordinate system N to the marker-based local coordinate system on segment A, M_x, M_y and M_z are orthogonal unit vectors of the marker-based local coordinate system. They can easily be computed from the positions of 3 markers which are located nonlinearly on the segment A.

Secondly, we compute the rotation matrix from the marker-based local coordinate system to the anatomic local coordinate system on segment A - $^{MA}R^{A}$:

$$^{MA}R^{A} = [^{N}R^{MA}_{baseline}]^{-1} \cdot {}^{N}R^{A}_{baseline} ,$$ <2>

where $^{N}R^{MA}_{baseline}$ is the baseline rotation matrix from normal coordinate system to marker-based local coordinate system and can be obtained using eq.<1>, and $^{N}R^{A}_{baseline}$ is the baseline rotation matrix from normal coordinate system to the anatomic local coordinate system. Baseline means that the subject is asked to fully straighten his arm to align each local coordinate system fixed on the segment with normal coordinate system. In the ideal case (perfect alignment), $^{N}R^{A}_{baseline}$ is an identity matrix. However, based on the anatomy of the body, this may not hold true as there may be a small offset angle between normal coordinate system and body segment. The offset could be obtained from still image when the subject is asked to fully straighten his arm. Therefore, we can construct $^{N}R^{A}_{baseline}$ using the offset.

Thirdly, we compute the rotation matrix between normal coordinate system and the anatomic local coordinate system using eqs. <1> and <2>:

$$^{N}R^{A} = {}^{N}R^{MA} \cdot {}^{MA}R^{A} .$$ <3>

Finally, the rotation matrix between two segments A and B ($^{A}R^{B}$) can be computed as follows:

$$^{A}R^{B} = [^{A}R^{N}] \cdot {}^{N}R^{B} = [^{N}R^{A}]^{-1} \cdot {}^{N}R^{B} .$$ <4>

Each joint movement could be described by a sequence of rotations. All angle features shown in Table 1 can be computed using these equations.

3.3 Multi-Goal Framework
In this section we will discuss our multi-goal framework in the action space. Reaching and grasping task includes three sub-goals in the action space – *reaching*, *opening* and *flow*. All sub-goals are computable from motion features.

- *Reaching*: 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*: 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*: We expect the subject to coordinate their arm movement while reaching and grasping for the target smoothly in a consistent way.

4. COUPLING ACTION TO FEEDBACK

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 [6].

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 [4]. 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 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 and 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.

4.1 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 of a 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.

4.1.1 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. Arm supination is mapped to movement 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.

4.1.2 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.

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 extension 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 2: 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 hand trajectory velocity in the Z direction 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 3: 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 4: 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 and Elbow Extensions

There were three control parameters that the shoulder 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 extension between the starting angle and the expected angle at the target. $\sqrt{x}$ is used to interpolate between the following ranges:

Table 5: Midi velocity and duration range of 3 instruments connected to shoulder extension.

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.
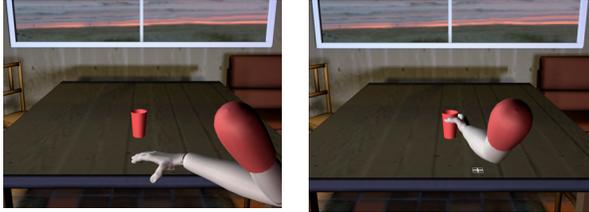

Figure 2: Transition Environment

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. See our joint angle calculation algorithm in section 3.2.

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 3).

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:

$$\vec{P'}_x = \vec{P}_x - \frac{n_{col}}{2}, \quad \vec{P'}_y = \vec{P}_y - \frac{n_{row}}{2}, \qquad <5>$$

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:

$$\vec{T_{\vec{P}}} = R_{(\theta,Z)}[\vec{T_E} + \vec{T_T} + \vec{T_{HP}} + \vec{T_{VP}}], \qquad <6>$$

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

Forearm supination controls the rotation angle. The difference between expected supination and subject supination (θ_s) is mapped to image rotation (θ), and a scaling factor (α_s) and is applied:

$$\theta = \alpha_s(\widehat{\theta}_s - \theta_s) . \qquad <7>$$

5.2.3 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
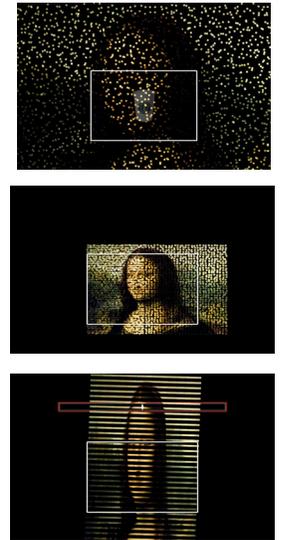

Figure 3: Visual feedback in the abstract environments. Top: particles begin to form the image as the hand approaches the target. Middle: Image pulled to the right when subject is off target. Bottom: Vertical bands appear when the subject has wrong target height.

position, the explosion effect is modified to collapse the image.

$$\vec{T_E} = \begin{cases} \beta_1 (1 - Z_N)^4 \vec{P'}, & 0 \le Z_N \le 1 \\ \beta_2 (Z_N - 1)^2 \vec{P'}, & Z_N \ge 1 \end{cases}, \qquad <8>$$

where β_1 is a explosion scale and β_2 is a collapse scale, P' is the relative position of the particle.

5.2.4 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.

$$\vec{T_T} = (1 - Z_N)(\lambda_1 N_P(\vec{P} + \vec{O}) + \lambda_2 \, t \, \delta N_P(\vec{P})) , \qquad <9>$$

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.

5.2.5 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.

$$\vec{T_{HP}} = [\eta(\mu X_H)^2 \quad 1] \vec{P'} , \qquad <10>$$

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.

$$\vec{T_{VP}} = [1 \quad \chi C(\frac{Y_H}{\max(Y_H)}, 0, 1)] \vec{P'} , \qquad <11>$$

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.

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 4 shows an offline segmentation result based on the speed curve.
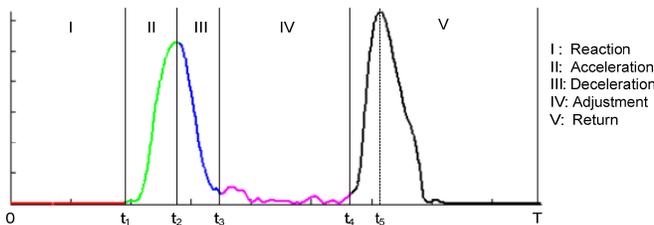

| | I: Reaction |
| II: Acceleration |
| III: Deceleration |
| IV: Adjustment |
| V: Return |

Figure 4: 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:

$$t_1 = \min\{t^* | t^* \in [0,T], \forall t \in [t^*, t^*+\varepsilon], v(t) > \alpha\} , \qquad <12>$$

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:

$$t_2 = \min\{t^* | t^* > t_1, v(t^*) > \beta, v(t^*) = \max_{t^*-w \le t \le t^*+w} [v(t)]\} , \qquad <13>$$

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:

$$t_3 = \min\{t^* | t^* > t_2, v(t^*) < \gamma, v(t^*) = \min_{t^*-w \le t \le t^*+w} [v(t)]\} , \qquad <14>$$

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 4). 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:

$$d_1 = \frac{\| X_h(t_3) - X_T \|_2}{\| X_h(0) - X_T \|_2} , \qquad <15>$$

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:

$$d_2 = \frac{| \theta_h(t_3) - \theta_T |}{\Theta_h} , \qquad <16>$$

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:

$$s = w_1^s \cdot d_1 + w_2^s \cdot d_2 , \qquad <17>$$

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

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

$$p_s = \frac{|\theta_s(t_3) - \theta_s^T|}{|\theta_s(0) - \theta_s^T|}, \; p_e = \frac{|\theta_e(t_3) - \theta_e^T|}{|\theta_e(0) - \theta_e^T|}, \qquad <18>$$

where p_s and p_e are the shoulder opening and elbow opening respectively, θ_s and θ_e are shoulder extension angle and elbow extension angle respectively, t_3 is decelerating ending time, θ_s^T and θ_e^T are the desired shoulder extension 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:

$$p = w_1^p \cdot p_s + w_2^p \cdot p_e, \qquad <19>$$

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:

$$r = t_3 - t_1, \qquad <20>$$

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 extension angle and (c) speed of elbow extension angle.

6.5.1 Smoothness Metric

Let us denote the speed curve during reaching as $v(t)$, $t_1 \leq t \leq t_3$. 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.

$$v_N(t) = \frac{v(t)}{\max[v(t)]}, t \in [t_1, t_3]. \qquad <21>$$

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:

$$e_f = \int_{t_1}^{t_2} [v_N(t) - f(v_N(t))]^2 dt + \int_{t_2}^{t_3} [v_N(t) - f(v_N(t))]^2 dt, \quad <22>$$

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 paper, 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:

$$M = [k, e_f]^T. \qquad <23>$$

6.5.2 Overall Flow Error

The overall flow error incorporates the smoothness of three speed curves: (a) hand marker speed, (b) shoulder extension speed and (c) elbow extension speed. Let us denote the smooth vector of hand marker speed, shoulder extension speed and elbow extension speed as M_h, M_s and M_e respectively. The overall flow error F is represented as the linear combination of these three smooth vectors:

$$F = w_1^f \cdot M_h + w_2^f \cdot M_s + w_3^f \cdot M_e, \qquad <24>$$

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 extension speed and elbow extension speed together to obtain the overall consistency.

6.6.1 Spatial Alignment

We align the normalized speed of reaching phrase $v_N(t)$, $t_1 \leq t \leq t_3$ 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 ith bin is:

$$\mu_i = \frac{\int_{t:z'(t) \in (z_{i-1}, z_i]} v_N(t) dt}{\int_{t:z'(t) \in (z_{i-1}, z_i]} 1 dt}, \qquad <25>$$

where z_i is the upper bound of the ith bin. Thus, the speed alignment can be represented by μ_i, $i=1,\ldots,N$.

6.6.2 Speed Variance

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

$$\sigma_{N,K}^2 = \frac{1}{N} \sum_{i=1}^{N} [\frac{1}{K} \sum_{k=1}^{K} (\mu_{i,k} - \frac{1}{K} \sum_{k=1}^{K} \mu_{i,k})^2], \qquad <26>$$

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 extension speed and elbow extension speed, we can obtain the overall consistency for K reaching trials:

$$C_{N,K} = w^c_1 \cdot \sigma^2_{N,K,h} + w^c_2 \cdot \sigma^2_{N,K,s} + w^c_3 \cdot \sigma^2_{N,K,e}, \qquad <27>$$

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 extension speed variance and elbow extension speed variance respectively.

7. EXPERIMENTS

We now discuss the experiment design, setup and results.

7.1 Experiment Design

In order to determine the effect of the biofeedback system on the arm functional task, six able-body subjects were recruited to test the design scenarios. Three subjects were used as the control group. The recruited subjects were right handed adults. They were all unfamiliar with the designed system prior to the test. Every subject was tested once.

The feedback condition is defined as a block. Each block 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 test, a block 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 2), 20 in each abstract environment (ref. Figure 3), followed by 15 reachings in the physical environment.

7.2 Setup and realization

In this section we will discuss setup and realization issues.

Calibration: To start calibrating the motion capture subsystem, we follow the standard calibration procedure provided by the Motion Analysis Corporation [1]. The normal 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 is a big screen showing up the visual feedback. Six motion capture cameras are looking over the entire capture volume. The subjects need to perform two baseline calibrations for our motion analysis subsystem to compute customized joint angle offset and synchronization table between hand positions and join angles.

Experiment process: Our experiment gradually switches from non-feedback environment, to transitional environment and to two feedback environments of different complexity. Each environment has exactly the same target position whether it is physical target or virtual target that has been put on the screen. Before each feedback or non-feedback environment test, the subjects will be given proper instructions of the next test.

7.3 Results and Discussion

We now show the validation over 85 trials with 5 different environments. Each figure shows the average measure of the 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 5, we can see the spatial errors of the abstract environment are at the same level with real world reaching. This proves 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.
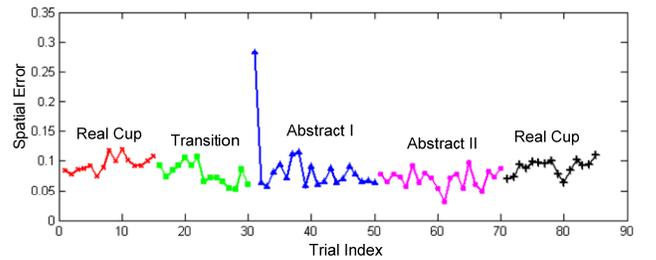

Figure 5: Spatial error

In Figure 6, 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 proves that our chord design in audio feedback for the abstract environment II communicates the opening message to the subjects very well.
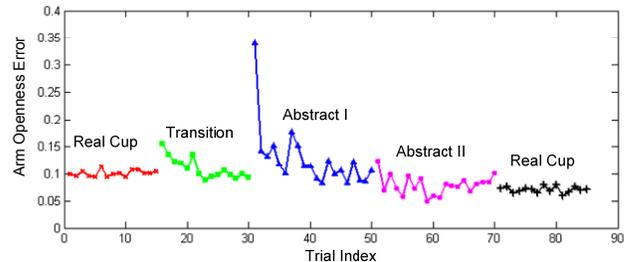

Figure 6: Arm opening error

The length of reaching duration is show in Figure 7. 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 8 and Figure 9, we show the flow error by two metrics: zero crossing number and curve fitting error.
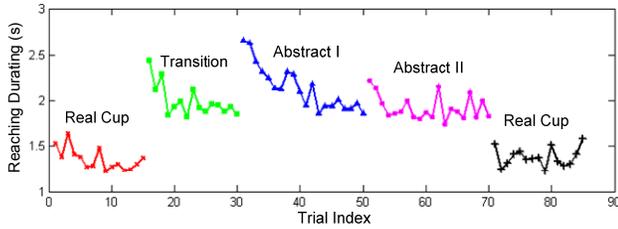
Figure 7: Reaching duration

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.
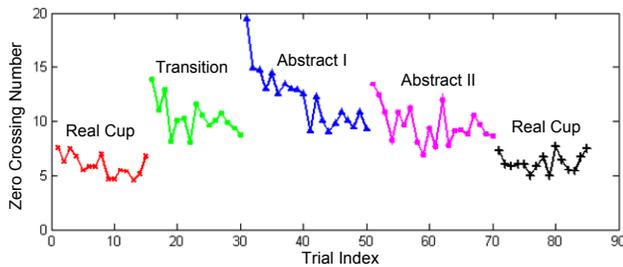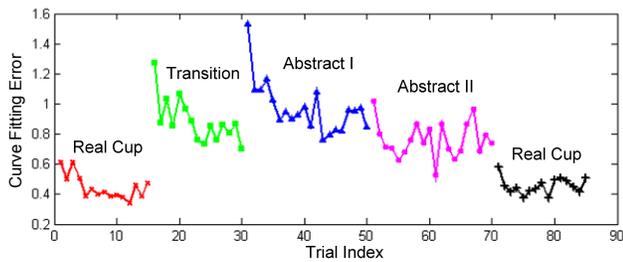

Figure 8: Zero crossing number

Figure 9: Curve fitting error

In Figure 10, we can see the speed variance decreasing in the transition environment and abstract environment I.
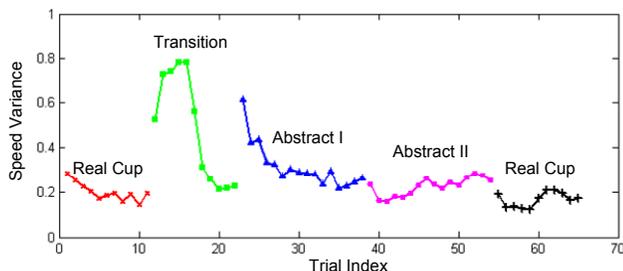

Figure 10: Speed variance

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 proves that our feedback design enables the subjects to achieve a stylistic consistency of action.

8. CONCLUSION

In this paper we have presented our efforts to develop a real-time, multimodal biofeedback system for stroke patients. There are several key contributions in this paper: we showed how to derive critical motion features 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. We plan to extend our research in several ways – (a) automated variation of feedback mapping through motion analysis, (b) user-level adaptation of feedback. We are planning to conduct extensive clinical trials in the next couple months.

9. REFERENCES

[1] *Motion Analysis Corporation* http://www.motionanalysis.com.

[2] 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.

[3] S. COREN, L. M. WARD and J. T. ENNS (2003). Sensation and perception. John Wiley Hoboken, NJ.

[4] 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.

[5] G. E. GRESHAM, P. W. DUNCAN and W. B. E. A. STASON (1996). Post-Stroke Rehabilitation/Clinical Practive Guideline. Aspen Publishers, Inc. Gaithersburg, Maryland.

[6] D. J. GROUT and C. V. PALISCA (2001). A history of western music. Norton New York.

[7] E. R. HILGARD and G. H. BOWER (1975). *Recent developments*. Theories of learning(eds). Englewood Cliffs, N.J.,, Prentice-Hall**:** 550-605.

[8] M. HOLDEN and T. DYAR (2002). *Virtual environment traing: a new tool for neurorehabilitation.* Neurology Report **26**(2): 62-72.

[9] 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.

[10] H. HUANG, T. INGALLS, L. OLSON, K. GANLEY, T. RIKAKIS and J. HE (2005). *Interactive Multimodal Biofeedback for Task-Oriented Neural Rehabilitation*, 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, Sep. 1st~4th, Shanghai, China.

[11] A. S. MERIANS, D. JACK, R. BOIAN, M. TREMAINE, G. C. BURDEA, S. V. ADAMOVICH, M. RECCE and H. POIZNER (2002). *Virtual reality-augmented rehabilitation for patients following stroke*. Phys Ther **82**(9): 898-915.