

Real-Time Sonification of Movement for an Immersive Stroke Rehabilitation Environment

Isaac Wallis, Todd Ingalls, Thanassis Rikakis, Loren Olsen, Yinpeng Chen, Weiwei Xu,
Hari Sundaram

Arizona State University
Arts and Media Engineering Department
Tempe, Arizona
riwallis@asu.edu

entire feedback is dynamically generated, in real-time using the analysis of movement [2].

ABSTRACT

This paper describes a novel and functional application of data sonification as an element in an immersive stroke rehabilitation system. For two years, we have been developing a task-based experiential media biofeedback system that incorporates musical feedback as a means to maintain patient interest and impart movement information to the patient. This paper delivers project background, system goals, a description of our system including an in-depth look at our audio engine, and lastly an overview of proof of concept experiments with both unimpaired subjects and actual stroke patients suffering from right-arm impairment.

[Keywords: Biofeedback, Sonification]

1. INTRODUCTION

Recent scientific advances show that exercise contributes to brain reorganization and behavioral recovery in stroke-related hemiparesis [1]. We are working to develop a comprehensive biofeedback system that monitors task performance, stress and engagement and uses the monitoring data to drive structurally coherent, multimodal, sensory and cognitive stimuli that promote performance improvement and enhance motivation, attention and active engagement. This system offers a potentially powerful solution to the problems of rehabilitation treatment – delivering an intensive and efficacious dosage that is accessible, and that promotes long-term carryover.

Our work is the development of a digital media based system that integrates task specific motor training and cognitive stimuli within an interactive, multimodal environment. The environment provides purposeful, engaging audiovisual scenes in which patients can practice functional reaching and grasping tasks, while receiving multi-modal feedback indicating measures of performance and results and reducing stress.

The unique and advanced features of the system include the development of context aware, multimodal, experiential feedback environments: adaptable 2D scenes representing 3D movement provide intuitive, corrective feedback on spatial movement parameters and accommodate depth perception issues in certain stroke survivors; generative music composition frameworks present structurally rich auditory feedback and feed-forward assistance that corresponds to hierarchical movement parameters and encourages performance accuracy and improvement. The

2. BACKGROUND

Recent research indicates that the use of interactive feedback in conjunction with movement therapy has a significant impact on the functional recovery of stroke patients with sensorimotor deficits [1,3,4]. However, there are some significant limitations affecting the outcome and validation of these studies: (a) Most of the systems relied on unimodal feedback (such as using only visual or sound feedback) or very simple combinations of modes; (b) feedback was informational and superficial rather than experiential and structural (for example a system would alert the patient that their movement speed was fast but would not provide feedback that slowed the patients movement in an intuitive, subconscious manner); (c) these systems were not adaptive (e.g. it was not possible to change the therapy, based on an specific individual's abilities and rate of progress), and (d) the test population was small.

In a previous study we designed an interactive multimodal environment (IME) based biofeedback system for repetitive reaching and grasping retraining [5]. In this environment the patient's arm was animated to create a virtual arm that could move through different visual scenes and approximate "grabbing" objects. There were two primary environments: (1) a virtual table with objects like teapots and cups and (2) a fish-tank with fish that the patient could reach out and "grab." Cues about spatial accuracy were shown visually through a semi-transparent cone and line indicating how far off from the correct trajectory the hand was. Audio feedback was provided by the progression through a chord sequence, with register changes mapped to opening of elbow. This helped indicate smoothness of movement and provided incentive for further opening of the elbow to extend reach. The shoulder of the tracked arm was also monitored and if moved beyond a pre-determined threshold, as would be seen with compensatory trunk movements during reaching activities in persons with stroke, a dissonant collection of notes in a high register played by winds was gradually presented to indicate to the user the undesired movement. It provided incentive to the user to contain that movement so the main musical phrase could be heard unimpeded.

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 [5].

In looking at ways to improve the system we decided that we needed to increase the number of parameter being looked at and find the most accurate measures for them. We also needed a more sophisticated audio feedback engine that could effectively communicate multiple measures at the same time.

3. SYSTEM GOALS

Physical therapy is a repetitive process that requires the subject to be highly engaged and motivated. It is crucial to provide the subject with feedback that gives them intuitive, real time measurements and evaluation of their performance and promotes behaviors that produce improvement. This can only be achieved through generative frameworks that combine all aspects of feedback into a semantically and aesthetically coherent experience. Challenges we need to address for the development of these frameworks include: (a) Optimization of mappings: for example, we believe interactive graphics with moving elements can communicate spatial parameters of movement and musical structures are optimal for meaningful layering of multiple time series. (b) Integration of modes: we need to develop formal structures that integrate various types of visual, auditory and tangible feedback into coherent multimodal experiences. (c) Complex driving systems: feedback formation and adaptation must be driven by multiple, interrelated sensing and modeling streams of the system; and (d) Adaptation to User-context: feedback must adapt to user preferences, performance, cognitive capacity and physical ability for long-term therapy and training.

While our prime goal is the integration of a task dependent physical therapy and cognitive stimuli within an interactive, multimodal environment, we identified three concepts computable from the motion features and that we also wished to communicate back to the user. These were *reach*, *openness*, and *flow*.

- *Reach*: Hand moving to the target with minimum spatial error, appropriate hand orientation and speed.
- *Openness*: Enough extension of joints to place hand in vicinity of the target with minimal shoulder/torso movement compensation.
- *Flow*: Coordination and synchrony of joints while smoothly and consistently reaching for the target.

In accordance with these goals we identified five measures to use in determining effectiveness of system

- Spatial error: distance of hand to target AND hand orientation measured after deceleration;
- Trajectory optimization
- Arm opening: shoulder and elbow extension.
- Jerk cost – measure the smoothness of the speed curve
- Asymptotic reach time and velocity profile

These are described in greater detail in [2].

4. DESCRIPTION

Our system is comprised of five separate components (Figure 1). The first is a six camera, marker-based Motion Analysis Corporation motion capture system. The data from this is multicast over UDP. The second component is our custom motion analysis engine that extracts real-time movement features from the marker data and outputs this analysis over multicast UDP [1]. There are both a visual feedback engine and an audio feedback engine running separately that provide feedback based on the motion analysis features. Lastly, there is an annotation, storage, and retrieval system that provides a web based interface for annotation and display of data. It also includes a database for storage and indexing of all data being broadcast, the settings for both the visual and audio feedback engines, and audio and video captures of the environment [6].

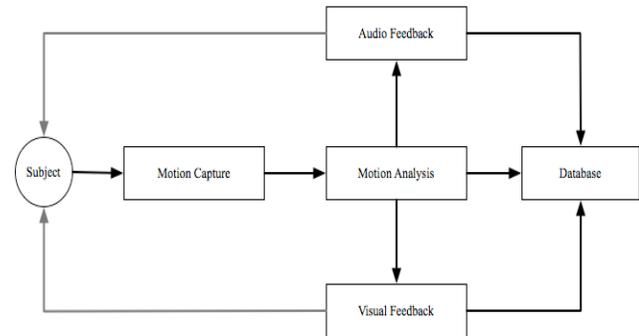


Figure 1. Diagram of biofeedback system information flow.

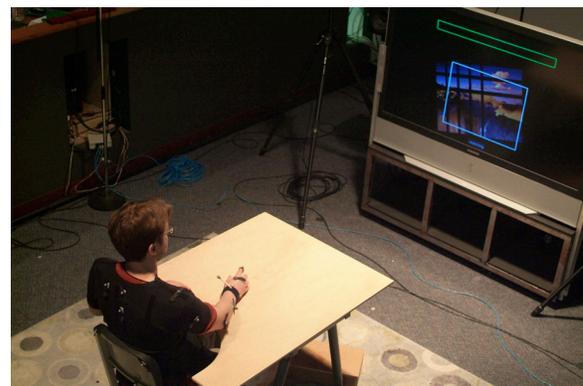


Figure 2. View of a subject using the system.

4.1. Motion Analysis

Real-time motion analysis is done using a six-camera motion capture system. This means that an initial calibration phase is required before any system usage, but gives us a high degree of flexibility in all three dimensions, as well as allowing us to track as many points of motion as are required for our needs.

In order to analyze the motion capture effectively, the effective space is divided into zones (Figure 3). The locations and areas of these zones are unique to each patient. They are initially decided during calibration before each session, but can be changed later as rehabilitation demands. These zones are:

- Resting zone: An area on the table near the patient where his or her right arm tends to rest naturally.
- Grasping zone: A target area that the patient must attempt to reach.
- Hull: A three-dimensional pathway between resting zone and grasping zone.

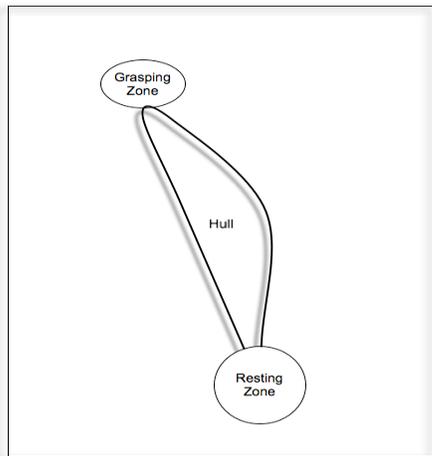


Figure 3. Diagram representing an overhead view of spatial reaching zones.

In addition, the reaching motion is broken into modes of movement.

- Resting mode: The patient's hand lies static in the resting zone before or after a reach.
- Reaching mode: The patient's arm is in transit between resting and grasping zones.
- Grasping mode: The patient's hand has reached the grasping zone and remained there for 0.5 seconds.
- Returning mode: The patient's arm is in transit from the grasping zone back to the resting zone.

One hundred times per second, the motion analysis system multi-casts a data frame to the audio and visual feedback system, and the annotation system. This data frame holds such information as the mode, the elbow and shoulder openness angles, the wrist supination angle, and velocity of motion.

4.2. Audio Engine

Our audio feedback is designed to help achieve the goals of improved patient openness and flow. The temporal nature of musical feedback uniquely qualifies it to indicate qualities such as smoothness/jerkiness and speed of reach. In addition, indicator sounds and audio alerts are used to ensure minimal movement compensation, thereby improving joint usage.

The audio feedback system is made of two programs: a third-party MIDI-controlled sampler (Native Instruments' Kontakt) which acts as our sound engine, and a proprietary gesture analyzer/music generator which receives the data message and controls the sound engine via MIDI. This controlling software is written in Max/MSP. A Max object written in C gets the data frame from the analysis system.

There are two types of instrumentation: foreground and background. There is only one foreground instrument, which plays throughout the reach. It may be a marimba, guitar, or piano sound. There may be as many as four background instruments at any one time, but they may be inaudible at times during the reaching motion. The instruments used for background are flute, trumpet, guitar, piano, marimba, and lastly, orchestra. The foreground instrument choices were used for their somewhat percussive timbres: as the primary purveyor of information, the foreground needed to be easily heard above the background ensemble. The background instrument timbres were chosen for their ability to deliver harmonic structure while filling up the sonic space.

Certain instruments are linked to aspects of the motion. The 'orchestra' instrument, for example, which is usually the first background instrument to be included, is linked to elbow openness. Initially inaudible, the volume of the orchestra swells as the elbow angle increases. This instrument mapping is motivated by those stroke patients whose opening and flow problems stem from reduced ability to extend their elbow.

The foreground instrument plays notes at equal intervals during resting mode and grasping mode. During reaching and returning, however, that interval diminishes, speeding up the rate of note generation (Table 1). This acceleration is controlled by the velocity of the motion. If the motion is very slow, the acceleration is very mild or nonexistent. However, if the motion is very quick, the rate of note generation can be very quick. This feature helps patients control their movement speed and smoothness: use of speed and momentum to reach a target is a common tactic of stroke patients with little arm strength, one that is detrimental to full rehabilitation.

Velocity	Note Value
0.00 - 0.20	quarter note
0.20 - 0.40	dotted eighth
0.40 - 0.65	eighth
0.65 - 0.80	dotted sixteenth
0.80 - 1.00	dotted thirty-second

Table 1: Mapping of normalized reaching velocity to note generation interval

The rhythms of the background instruments remain steady throughout the reach. These instruments are used to impart rhythmic variety and rhythmic consistency to the music during the accelerated musical gestures. At one early stage, we experimented with accelerating foreground and background equally during the reach. The result was aesthetically unpleasing, making the reaches sound like interruptions to the musical continuity. This method, where the background remains steady and the foreground accelerates over it, improves the feeling of homogeneity between reaching and static sonifications.

Interestingly, the interruption effect does not pose a problem when the instrumentation consists of a single solo foreground instrument.

During resting mode, the harmony remains on one chord. When the patient performs the reaching motion, a harmonic progression is cycled through (Table 2). If the patient attempts to reach and returns, but does not reach the grasping zone, the progression is not completed. Instead, the harmonic progression cycles backward to the beginning, with no chordal resolution. Essentially, harmonic resolution is used as incentive for reaching the target on each attempt.

Activity	Zn	Harmony
Resting	0.00 - 0.20	I
Reaching	0.20 - 0.50	ii
	0.50 - 0.85	III7
Grasping	0.85 - 1.00	vi
Returning	1.00 - 0.63	V7
	0.63 - 0.30	V7
Resting	0.30 - 0.00	I

Table 2: Mapping of normalized distance to grasping zone in Zn direction to a sample harmonic progression

Harmonic progressions change upon every reach. This, along with the addition and removal of instruments, is a main source of musical variety. Thought of artistically, the generative material of the biofeedback therapy session can be said to have been 'composed' by the audio feedback team. All note selection is either arpeggiated or random within the harmonic structure.

Our system allows for the creation of any monophonic cyclical arpeggio pattern for any of our instruments. Each pattern is transposable, and only selects note values that are in the current chord. Furthermore, patterns can be constructed for note velocities, durations, and onset intervals as well. In this way, rhythms and rhythmic effects can be imparted to individual instruments, allowing for greater compositional control.

When an instrument's note selection is random rather than arpeggiated, that instrument selects notes from the current chord and in ranged boundaries. The ranged boundaries are important for ensuring no notes are selected outside of an instrument's natural range. The foreground instrument often begins in arpeggiated mode, but switches to random mode during the accelerated note generation of reaching, switching back after deceleration has occurred. This was done for aesthetic reasons: a cyclical pattern which sounds good at the slower tempo of patient resting is unlikely to sound equally good when the tempo is accelerated by greater than a factor of four.

The audio feedback system makes use of three audio alerts. The first is an audio alert indicating successful reach; it is a single strike of the triangle. We use harmonic resolution as a motivator for reach completion on the subconscious level, but the success alert is far more explicit.

The other two audio alerts each indicate poor movement technique. Rather than discrete events like the success indicator, these are sustained indicators that increase or decrease in volume in relation to the severity of compensatory movement.

For example, a rain-stick sound tells the patient when he is leaning forward or slouching past a predetermined threshold. Some patients have learned to lean forward in order to bring their bodies closer to any objects they need to reach. This type of compensation, while easing the reaching process, is less helpful for rehabilitation and could facilitate other health problems in the future.

Similarly, a *piatti* sound alerts the patients when they raise their shoulder in a compensatory movement past a threshold. Patients who have dealt with reduced arm strength sometimes use their shoulder to achieve height that they would have difficulty with otherwise. Again, this is poor movement practice and could lead to future health problems.

Figure 4 is a notated score from part of a good reach. The foreground instrument is marimba, and the background instrument is orchestra. This shows a chord change and a rhythmic change as the movement velocity slows because the patient is approaching the resting zone after a reach.



Figure 4. Sample score of returning movement

Figures 5-7 are movement parameter graphs showing the differences between the reaching movement of stroke victims and the reaching movement of unimpaired subjects.

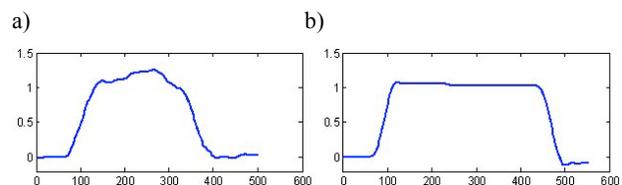


Figure 5: Comparison of percentage Z during a full reach-grasp-return cycle for (a) a stroke patient and (b) an unimpaired subject.

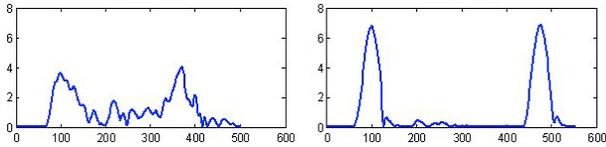


Figure 6: Comparison of velocity during a full reach-grasp-return cycle for (a) a stroke patient and (b) an unimpaired subject.

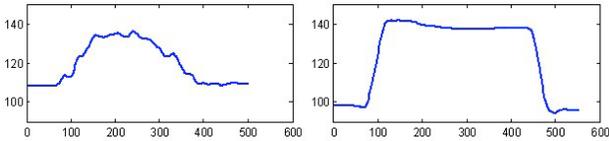


Figure 7: Comparison of elbow openness during a full reach-grasp-return cycle for (a) a stroke patient and (b) an unimpaired subject. The y axis represents angle calculated between upper arm and fore arm.

In figure 5 one can see that, although the stroke patient was able to reach the grasping zone, he did not have the same accuracy and control as the unimpaired subject. His approach slowed earlier and he overshot. His returning plot was more rounded as well. These two subjects' audio feedback would differ harmonically. The normal subject's harmonies would cycle more quickly and evenly than the impaired persons' would.

The stroke patient in figure 6 does not achieve the peak velocity of the unimpaired subject. He also has a very difficult time maintaining a velocity. His velocity oscillates even during grasping when no movement is required. These two people would hear very different musical gestures during their reaches. The unimpaired subject would hear a smooth acceleration and deceleration during reaching, and another acceleration/deceleration during returning. The impaired patient would hear very jerky music, with a burst of notes for every jerk of his arm.

Figure 7 shows a stroke patient's elbow openness during reaching versus a normal subject's elbow openness. Because of the more rounded, less defined curve of the stroke patient's elbow openness, the orchestra would enter later than it would with the unimpaired person's reach, and the orchestra would never attain the volume that the normal person would hear.

4.3. Visual Engine

In contrast to the audio engine, which targets openness and flow, the visual engine is designed to help achieve the system goal of improved reaching. Visual feedback has the ability to impart very specific location and orientation cues to our patients, helping them improve spatial accuracy during reaching, as well as allowing them to formulate a better plan of action on future reaches.

The visuals engine is created in Dash, an in-house animation software written in objective C and Python. There are two environments: an explicit environment that is used to introduce the patient to the system, and an abstract environment that is used once the patient has been introduced.

The explicit visual training environment depicts a 3-dimensional virtual room with a table (Figure 8). There is a

disembodied right arm hovering over the chair in the same place that the patient's right arm exists in relation to the physical table. The virtual arm follows all movements of the physical arm.

A virtual cup is placed in the center of the virtual table. This cup corresponds to the grasping zone. The task assigned to the patient is to reach for the cup. Once the virtual hand has touched the virtual cup, and the wrist has supinated in simulation of physical grasping, the cup disappears and it is time for the patient to return to the resting position. It is important to note that in this introductory environment, the audio feedback engine is not used.

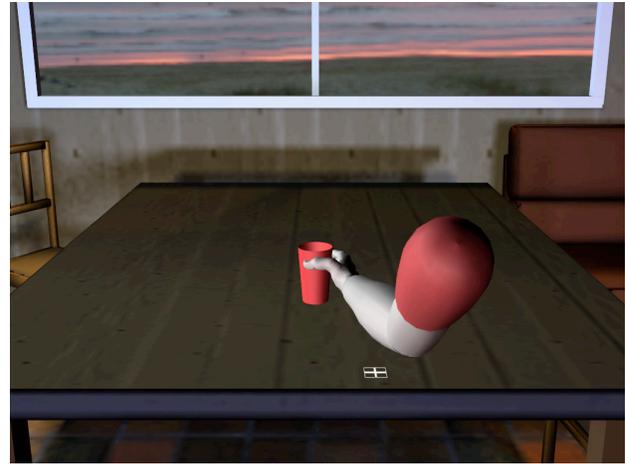


Figure 8. Explicit visual environment

After the patient has become familiar with the system through the use of the explicit environment, he or she is switched to the abstract environment. In the abstract visual environment, the patient is initially shown a picture centered on the screen. The picture then explodes into particles, leaving a target represented by a cup and a frame.

The particles move in conjunction with the patient's hand, and can be recombined with a reaching motion. As the patient advances his hand, the particles condense inward. If the patient supinates, the particle field rotates. If the patient movement travels outside of the hull, the particle field stretches in the direction of the movement error. Similarly, if the patient overshoots the grasping zone, the image will compress to be smaller than the target frame. (Figure 9)

After returning to the resting zone, the patient is given a few seconds, then restarts the process with a new picture. The pictures change upon every reach. These images may be pulled from different sources, depending on the needs of the patient. If we want to impart a sense of storyline, we may use frames from a well-known movie. If we want to give the patient a personal incentive, he or she may bring in digital family photographs. If neither is required, we might pull the image from a pool of art masterpieces.

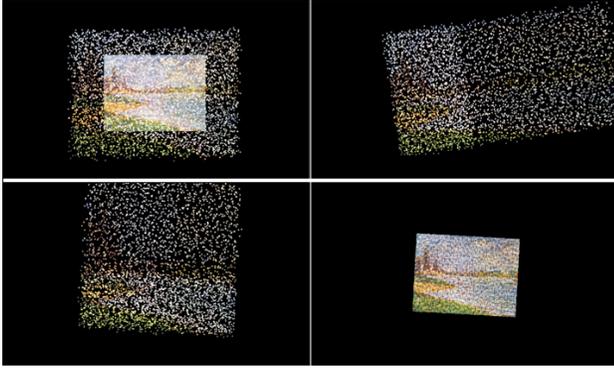


Figure 9. *Abstract visual environment*

4.4. Annotation, Storage, and Retrieval

Our system records system parameters, patient data, team and patient notes, and reaching data in an SQL-based database. In addition, each testing session is recorded by video cameras and microphones, and all generated audio and video is captured. This is helpful both for further research and development on our system and for rehabilitation.

Physical therapists are able to recall and analyze data from any patient down to the individual reach. One hundred frames per second of motion analysis data is saved by the storage and retrieval system, allowing the display of reaching data in graph form. This is particularly useful for rehabilitation purposes. Between each set of ten reaches the results are graphed and displayed for team members to analyze.

Upon each trial, both feedback engines and the analysis engine transmit parameter data to the annotation system. This information is stored in a database. Using this database and saved motion analysis data, individual tests can be recreated.

A web-based form allows team members to annotate observations. All observations are saved to the database and are viewable in real-time by the other members of the team. In order to speed the annotation process and reduce cognitive load on team members, frequently repeated annotations are represented by check boxes. Otherwise observations are typed.

Written in Max/MSP, the audio/video capture system records the visuals and music from our system. In addition, video is recorded from a camera positioned to include both the patient and the video monitor (Figure 10). Audio from three microphones is recorded: one placed to pick up patient remarks, another placed to pick up team remarks, and a third clip-on microphone to pick up therapist remarks. This last microphone is important because, during testing, the therapist often does not have access to a computer in order to annotate observations.

For more information on the annotation, storage, and retrieval system, please refer to [6]



Figure 10. *Side view of the subject and system.*

5. EXPERIMENTS

5.1. Testing with Unimpaired Subjects

Testing on twelve unimpaired subjects was done for proof of concept and to determine strategies that are useful for rehabilitation. Because of the high number of interrelated scientific variables, we realized that it would be impossible to work with and eliminate one musical parameter at a time.

Instead, we worked with several at once, using our musical backgrounds to help us decide which variables had the greatest likelihood of achieving a certain goal. In our case the target goal was to affect the speed of reaching. With careful selection of musical parameters, we hoped to be able to hasten or slow the reach of unimpaired patients at will.

We started with three variables that we felt had a high likelihood of affecting the speed of the reaching movement: tempo, foreground instrumentation, and arpeggiated versus random note selection. Over the course of nineteen or twenty sets of ten reaches each, we tested carefully sequenced permutations of these variables.

The results of these experiments were that the speed of reaching invariably changed. The direction of change was not always consistent with our goals, but this was not surprising, as more than just audio feedback is involved in a subject's reaching speed decision. Fatigue and patience also play a large role.

Comments and criticisms by the test subjects were annotated in the database. Most of them dealt with instrument preference, but it was commonly noted that the audio feedback seemed tediously repetitive after a large number of reaches. This could be due in part to the unimpaired subjects' highly stylized methods of reaching. They are capable of reaching as they wish; therefore during a long-running repetitive reaching experiment, their reaches are likely to fall into a pattern of rote repetition.

5.2. Testing with Impaired Subjects

We recruited three stroke patients with right-arm impairment to test our system. They had differing types and levels of

impairment, but all were cognitively sound and had been healing from their latest stroke for six months or more. One patient had no supination ability, and reduced arm strength. Another had somewhat greater strength and supination, but a deep tremor in the right arm.

Because of the greater rate of fatigue in these patients, we only did ten sets of up to ten reaches. Unlike the experiments performed on unimpaired patients, we had no target goals other than the overall improvement of patient reaching ability.

In order to achieve this goal, we initially set the grasping zone for each patient in a location that was not difficult for them to achieve.

Once the patient was able to consistently perform grasping as judged by our physical therapist, we would make the task a little more difficult by reducing the grasping zone area and increasing distance to grasping zone. At the same time, we would reward the patient by including a new instrument into the musical ensemble, thus increasing musical variation.

As we performed these experiments, we began to happen upon useful rehabilitation techniques. For example, we noticed that one patient did not have good elbow opening during reaching. We succeeded in helping him by moving the grasping zone to the right a few inches. His elbow was forced to open wider in order to reach the new grasping zone, and later we were able to return the grasping zone to its original position with no loss in elbow openness.

Patient comments were interesting. One patient had a stylistic musical preference for jazz, and derived less benefit from the audio feedback as a result of tuning it out. Another patient requested a change of photographic imagery for the visual environment. She found our initial choice, which were stills from a popular animated movie, patronizing because of the childish images. Most comments, however, centered around the rehabilitative effects. For example, one patient remarked, "I forgot I could still do this [motion with this arm]."

We are currently analyzing the data from these trials to evaluate the system and make determinations regarding. Some very preliminary observations are encouraging. For instance, Figure 11 shows the velocity graph for last 5 trials of the day for a patient on the first day of testing compared with the last five trials on the last day of testing. As can be seen both smoothness and quickness to target seem to improve. However, we need to develop further tests and trials before making any determination.

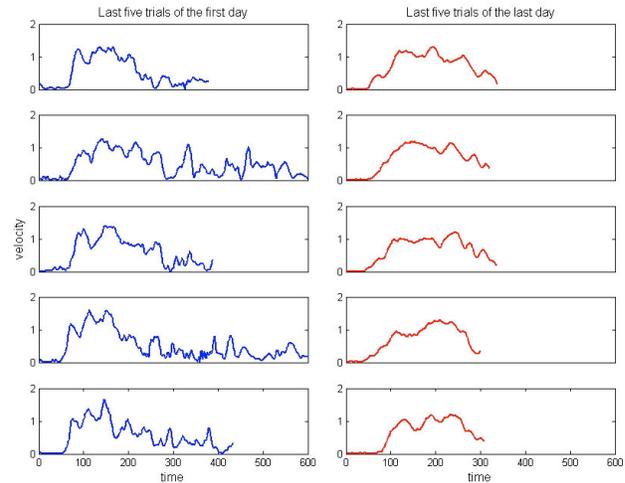


Figure 11. Patient velocity for first five trials versus velocity in last five trials of one session

6. CONCLUSION

This descriptive paper details the current state of development on our real-time multimodal environment for stroke rehabilitation, and the way in which musical sonification of the reaching gesture plays a part. We have shown how real-time mapping of musical gestures to movement is accomplished in our system. We also describe key experiments on both physically sound and stroke-impaired patients, with highly encouraging results in each. Some future goals of this research are: (a) to make the musical component of the sonification even more flexible, allowing for musically stylistic results. (b) to create a more game-like rehabilitation environment. After this latest round of testing, we are now entering our next development phase.

7. ADDITIONAL FILES

[Biofeedback_Sept06_02_H264.mov]

8. REFERENCES

- [1] J. Liepert, H. Bauder, H. R. Wolfgang, W. H. Miltner, E. Taub and C. Weiller, "Treatment-induced cortical reorganization after stroke in humans," in *Stroke*, vol. 31, pp. 1210-1216, 2000.
- [2] H. Huang, T. Ingalls, L. Olson, K. Ganley, T. Rikakis, J. He. "Interactive Multimodal Biofeedback for Task-Oriented Neural Rehabilitation." 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2005.
- [3] M. L. Dombrov, "Understanding stroke recovery and rehabilitation: current and emerging approaches." In *Curr Neurol Neurosci Rep*, vol. 4, no. 1, pp. 31-35, 2004.
- [4] H. Huang, S. Wolf and J. He, "Recent development in biofeedback for neuromotor rehabilitation," in *Journal of Neural Engineering and Rehabilitation*, vol. 3, no. 11, 2006.

- [5] Y. Chen, H.Huang, W. Xu, R.I. Wallis, H.Sunadaram, T.Rikakis, T. Ingalls, L. Olsen, J. He, "The Design of a Real-Time, Multimodal Biofeedback System for Stroke Patient Rehabilitation," in *Proceedings of the ACM Multimedia Conference*, Santa Barbara, CA, 2006.
- [6] W. Xu, Y. Chen, H. Sundaram, T. Rikakis, "Multimodal Archiving, Real-Time Anotation and Information Visualization in a Biofeedback System for Stroke Patient Rehabilitation," in *Proceedings of the ACM Continuous Archival and Retrieval of Personal Experiences (CARPE) Conference*, 2006.