Mikhail Gorman,[*] Amir Lahav,[†] Elliot Saltzman,[†] and Margrit Betke[*]**
[*]Image and Video Computing Group
Department of Computer Science
Boston University
111 Cummington Street
Boston, MA 02215 USA
{mishaz,betke}@cs.bu.edu
[†]Music Mind & Motion Laboratory
Sargent College for Health and Rehabilitation Science
Boston University
635 Commonwealth Avenue
Boston, MA 02215 USA
esaltz@bu.edu
[**]The Music & Neuroimaging Laboratory
Beth Israel Deaconess Medical Center
Harvard Medical School
330 Brookline Avenue
Boston, MA 02215 USA
alahav@bidmc.harvard.edu

A Camera-Based Music-Making Tool for Physical Rehabilitation

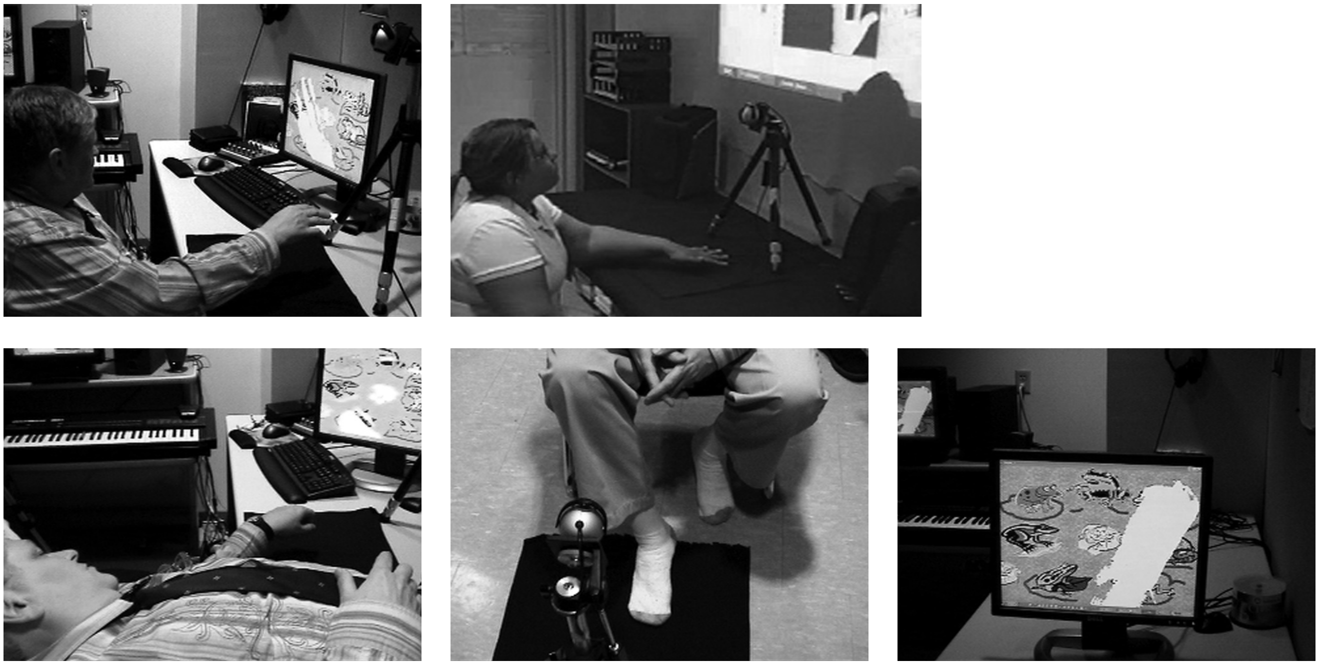
The universality of music among human cultures as well as our common experience of naturally responding to music with motion, seem to be widely recognized (Tramo 2001). Recent brain-imaging studies (Lahav et al. 2005; Lahav, Saltzman, and Schlaug 2007) show that humans, given appropriate auditory inputs, seem to be "tuned" to produce corresponding motor outputs. This unique auditory-motor interplay provides the conceptual basis for the use of music therapy, in particular, *active* music therapy, where a patient is physically involved in producing music rather than simply reacting to or accompanying music (Pacchetti et al. 2000; Paul and Ramsey 2000; Lahav 2005).

Music therapy has benefited people with physical disabilities, mental health needs, developmental and learning disabilities, Alzheimer's and Parkinson's disease, autism, substance abuse problems, brain injuries, and acute and chronic pain (Pacchetti et al. 2000; Harris and Summa-Chadwick 2005; see also the American Music Therapy Association's Web site, www.musictherapy.org). Playing an instrument, such as piano, guitar, or drums, may be very difficult or even infeasible for patients with motor dysfunctions. As an alternative, easy-to-use tool for active music therapy, we designed Music Maker, a human–computer interface that converts body movements into musical and visual feedback in real time using the open software platform EyesWeb (Camurri et al. 2004). Music Maker is a non-obtrusive camera-based tool that allows physically impaired patients to naturally create music with no prior musical training. Patients simply move their hand or foot in space or on a surface, and Music Maker detects and interprets these movements. Detection is accomplished without any sensors or markers that must be attached to the patients' bodies. Music Maker relies only on the video input from a camera to observe patient motion and on computer-vision techniques to analyze the motion.

Music Maker is an adaptive interface that can be adjusted to provide auditory and visual feedback based on the patient's needs and interests. Auditory feedback could range from a single piano note to a recording of the patient's favorite piece of music. Visual feedback is provided by a graphical display on a computer monitor or wall-mounted screen. Music Maker uses "fun" displays, for example, cartoon drawings or pictures of musical instruments. Its hardware setup can be adjusted according to patients' levels of impairment, their particular therapeutic goals, and the equipment available in hospitals or patients' homes. (See Figure 1 for sample setups.)

Computer Music Journal, 31:2, pp. 39–53, Summer 2007

Figure 1. Music Maker in use. It observes the movements of the users' hands (top row and bottom left) or a foot (bottom middle and right) with a downward-facing camera, interprets them, and provides auditory and visual feedback to the user. Visual feedback is provided with a monitor display (top left and bottom row) or a wall-mounted screen display (top right).



Work related to our cross-disciplinary effort can be found in the literature of the fields of computer vision, human–computer interaction, multimedia, and rehabilitation. In designing the Music Maker, we made use of the EyesWeb work from the Laboratorio di Informatica Musicale in Italy (e.g., Camurri et al. 2000, 2003, 2004) and our own experiences in developing a number of human–computer interfaces for people with severe disabilities who use cameras to access the computer (Betke, Gips, and Fleming 2002; Grauman et al. 2003; Magee et al. 2004; Chau and Betke 2005; Akram, Tiberii, and Betke 2006). EyesWeb is an open, multimedia software platform that provides software modules in the form of visual-language blocks. By connecting these blocks, a software developer can analyze video input using the rich functionality of Intel's computer vision library OpenCV (www.intel.com/technology/computing/opencv), create graphical displays, and provide visual and audio output. Music Maker is implemented using EyesWeb software modules and OpenCV functions.

EyesWeb has primarily been used to create tools for computer music interactions in large performance spaces (e.g., to facilitate interactive dance performances; Camurri et al. 2000), but its potential as a design tool for therapeutic exercises has also been explored previously (Camurri et al. 2003; Lewis-Brooks and Hasselblad 2004). Camurri et al. (2003) employed camera-based full-body tracking to create pilot exercises such as "Stand and Sit." Our focus has instead been to develop tools for detection and tracking of smaller body parts, such as hands or feet, during physical exercises (see Figure 1). Our goal has been to design exercises that have the potential to improve measures of motor function and hand–eye, foot–eye, or bi-manual coordination. Music Maker provides quantitative tools for analyzing and monitoring these movement measures, for example, the range of motion of hands or feet, the frequency and amplitude of finger tapping, or the shape of the trajectory of the hand during a reach-to-grasp movement. The therapist may use these analysis tools for (1) initial diagnosis, (2) development of safe and effective therapeutic exercises, and (3) subsequent evaluation of the patient's recovery process.

In addition to the interfaces developed with the EyesWeb platform (e.g., Camurri et al. 2000, 2003; Lewis-Brooks and Hasselblad 2004; Burns and Wanderley 2006), there are several other camera-based human–computer interfaces that produce music from body movements. These systems detect movements within pre-defined performance areas and trigger sounds in real time. DanceSpace uses a computer-vision gesture-recognition system (Sparacino et al. 2000; Wren et al. 1997a, 1997b) to track a performer's body parts, and it can map different instruments to these parts while a melody plays in the background. The performer mimes, for example, playing a virtual cello with one hand and a drum with a foot. The spatial extents of movements are matched to pitches of notes. With STEIM's BigEye system (www.steim.org/steim/bigeye.html), a user can define objects and spatial zones that are of interest. Configured in this manner, the system then extracts the objects of interest from the input video and compares their positions to the user-defined zones. It generates sound messages each time an object appears or disappears in a zone or moves within a zone. Sound messages can also be generated using additional object parameters, such as position, size, and speed of object motion.

Handel (Tarabella 2004) and Camera Musicale (Rémus 2006) are human–computer interfaces that produce music from movements of hands and fingers in the field of view of a video camera. The Handel interface has been used to develop an imaginary piano that the performer plays by reaching into different predefined spatial zones with various speeds. The Camera Musicale uses an infrared video camera to facilitate hand detection. Thousands of people reportedly interacted with various installations of Camera Musicale at concerts, festivals, and expositions.

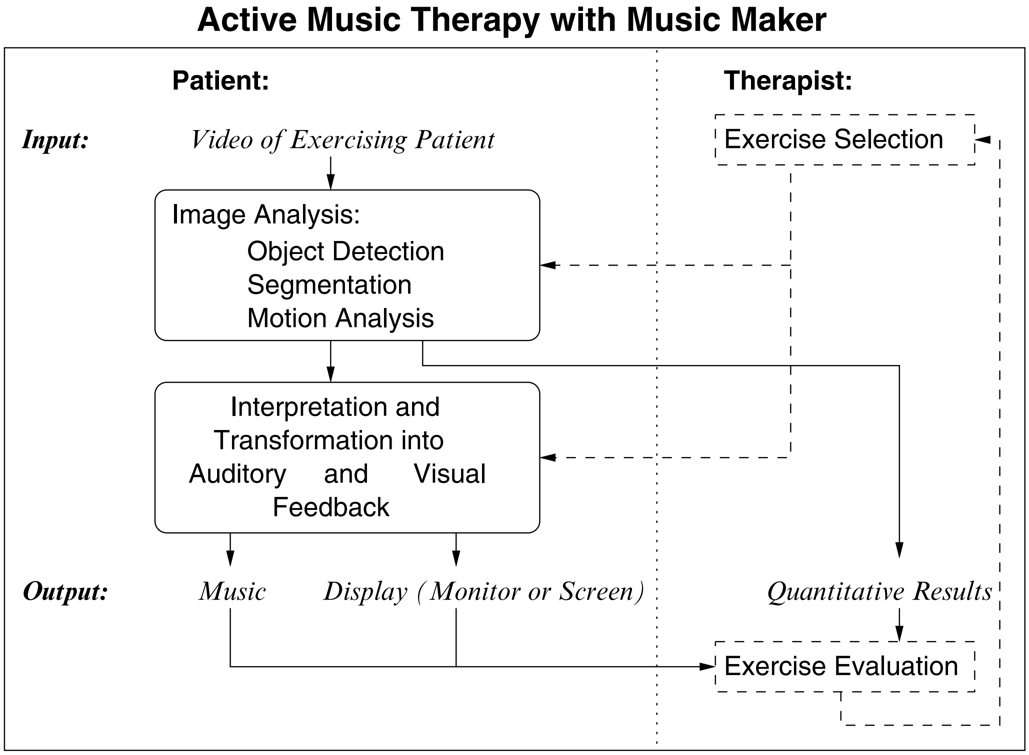
The Vicon 8 Motion-Capture System (Dobrian and Bevilacqua 2003) uses eight calibrated and synchronized video cameras to musically interpret the performer's body movements—in particular, the motion of 30 body markers. The performer may be in a remote location and seen by the audience as an animated musical avatar. The system was developed using Motion Capture Music (MCM) software. An alternative software platform is the Musical Gestures Toolbox (Jensenius, Godoy, and Wanderley 2005). Using modules in this toolbox, a user can view, analyze, and annotate musical gestures in multiple video streams, for example, of a performer who mimics playing the piano while listening to piano music. Another programming toolbox is the pureCMusic library (Tarabella 2004), which has been used to facilitate interpretation of musical gestures, for example, to control compositions with the Handel interface.

Systems without cameras that produce music from body movements have been surveyed by Miranda and Wanderley (2006), Morales-Manzanares et al. (2001), and Winkler (1995). These include interfaces with touch sensors placed on the floor (e.g., Johnstone 1991; Griffith and Fernstrom 1998) or attached to the body, clothes, or shoes of a user (Paradiso et al. 2000). The Kontrolldress (www.notam02.no/notam02/prod-maskin-kontrolldress.html) is a body suit equipped with sensors that a user can tap with the fingertips to produce sounds. In other systems, such as BodySynth by Van Raalte (1999), electromyographic (EMG) sensors attached to the body detect electrical signals generated by muscle contractions. Muscle contractions that trigger sounds can be very subtle, and the same sonic result can be achieved by a variety of movements (Nakra 1999). The MidiDancer (Coniglio and Stoppielo 2007) and the system by Siegel and Jacobsen (1998) use flex sensors that measure how much a dancer's joints (e.g., elbows, wrists, hips, and knees) are bent, and it converts the angle measurements into music. With both systems, the dancer must wear a wireless transmitter that is connected to the sensors.

Other human–computer interfaces that produce music from body movements use joysticks and virtual batons (Borchers and Muhlheuser 1998), sliders and table rings (Bongers and Harris 2002), light-sensitive drum sticks (Leider 1997), and electromagnetic sensors that sample the magnetic fields emitted by an external transmitter (Morales-Manzanares et al. 2001). These systems have the advantage that they do not suffer from possible self-occlusion by a user's body part. Camera-based systems, in contrast, must have a relatively unobstructed line-of-sight between the camera and the user to allow reliable motion analysis. Camera-based interfaces, however, are more natural and comfortable, because sensors do not need to be attached to the user's body and are therefore particularly appropriate for therapy purposes.

Various acoustic instruments and electronic

Figure 2. Conceptual overview of the use of Music Maker (solid lines) in active music therapy. The therapist can select and monitor exercises and use system outputs to adjust exercises (dashed lines).

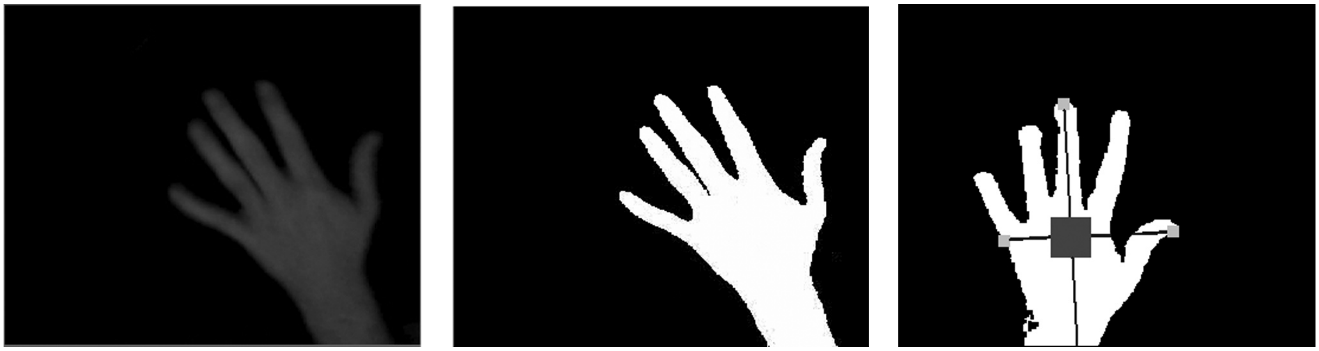

Active Music Therapy with Music Maker
Patient:
Therapist:
Input:
Video of Exercising Patient
Exercise Selection
Image Analysis:
Object Detection
Segmentation
Motion Analysis
Interpretation and Transformation into Auditory and Visual Feedback
Output:
Music
Display (Monitor or Screen)
Quantitative Results
Exercise Evaluation

music technologies are used for music therapy (Hunt, Kirk, and Neighbour 2004). Electronic interfaces are typically controlled by a computer mouse or a keyboard. The MidiGrid interface (Kirk et al. 1994), for example, plays notes or chords when the mouse pointer passes over certain regions on the computer screen. It has been used for music therapy for children with severe emotional problems and patients with physical disabilities due to cerebral palsy and traumatic brain injuries. Novel devices have also been used in therapy, including the "shell instrument," which contains various touch sensors that respond when tapped, scraped, hit, or wobbled (Hunt, Kirk, and Neighbour 2004) or instrumented footwear (Paradiso et al. 2004). Various assistive music technologies exist that may facilitate music therapy for patients with movement restrictions: these include the Drake Music Project (www.drakemusicproject.org), Soundbeam (www.soundbeam.co.uk), and the system proposed by Brooks et al. (2002). The Music Maker belongs to this category of technologies.

Method

Music Maker consists of two modules: (1) the image-analysis module, which processes the input video image of the patient, and (2) the interpretation module, which uses the video analysis to provide visual and auditory feedback for the patient. The computations performed by the image-analysis module are controlled by the type of exercise that the therapist selected. Exercises typically require the detection of the object of interest in the image, that is, the imaged body part of the patient, segmentation of the object from the background, and analysis of its motion. The therapist can monitor the patient's performance both qualitatively by observing music and visual output and quantitatively by reviewing the information about the patient's movement patterns that the image-analysis module provides. The therapist can then use this evaluation to select subsequent exercises and adjust the exercise parameters. An overview of Music Maker is shown in Figure 2.

Figure 3. Left: Camera view of a subject's hand. Middle: Filtered binary image of detected hand. Right: Binary hand image with centroid shown as a gray square, axes of least and most inertia as dark lines, and intersection points as small, light-gray rectangles.



The Image Analysis Module

The image-analysis module locates the object of interest in the image by color analysis, a technique that is often used in computer vision systems to detect faces or hands (e.g., Wu, Chen, and Yachida 1999; Schwerdt and Crowley 2000; Hsu, Abdel-Mottaleb, and Jain 2002). If the object of interest is the patient's hand, pixels of skin color are found by analyzing the 8-bit red, green, and blue color components of each pixel and looking for pixels with relatively large values of red and green but small values of blue. (The stored default values range between 40–255 for the red and green values and 0–40 for the blue values.) If the object of interest is the patient's foot, the color of the patient's sock or shoe is used for localizing the foot in the image.

At the beginning of a therapy session, the therapist uses an initial camera view of the selected body part to determine the range of colors that the image-analysis module must detect in this session. This optional manual initialization is convenient because it makes the system flexible in the sense that it allows different kinds of body parts with and without clothes, such as hands, arms, or feet, to be detected. It also makes the skin-detection method more reliable, because skin tones vary widely over the population, and pixel colors depend on the lighting conditions and camera selection. The manual initialization step may be omitted if a subject's bare foot, for example, is detected reliably with the default values for skin color. To simplify the detection of the body part even further, a single-color background (e.g., black cloth) can be used.

Once the color range of the imaged body part is determined, the image-analysis module creates a binary image of pixels with desired values and applies a one-pixel erosion operation (Jain, Kasturi, and Schunk 1995) to filter the object of interest in the foreground and remove small collections of pixels in the background that also happen to have the desired colors. A camera view of a subject's hand and the filtered binary image of the detected hand are shown in Figure 3, left and middle, respectively.

The image-analysis module computes various properties of the segmented object of interest, such as size, location, orientation, and length of perimeter (Figure 3, right). The object location in the image is represented by the centroid (Horn 1986) of the foreground pixels. The orientation of the object is computed by determining the orientation of the axis of least inertia (Horn 1986). The intersection of this axis with the object perimeter is determined to compute the length of the object. Similarly, the intersection of the axis of most inertia with the object perimeter is determined to compute the width of the object in the image. A comparison of the location of the centroid in consecutive image frames provides information about the direction and speed of the motion of the object in the video. This approximation of velocity is considered to be the first derivative of the location parameter. Similarly, the first and second derivatives of the other parameters are computed to provide information about the type of motion.

For quantitative analysis of the patient's performance, the properties of the imaged body part are converted from the two-dimensional image-coordinate system to the three-dimensional world-coordinate system. The perspective projection equations (Horn 1986) are used for this conversion. In particular, given focal length f of the camera and the dimensions of the exercise space in which the subject moves (i.e., length X, width Y, and height Z), the field of view in the image can be expressed by the image dimensions $x_{\max} = f\,X/Z$ and $y_{\max} = f\,Y/Z$. If the distance Z_o of the patient's body part to the camera remains constant during the exercise, the location of the body part in world-coordinates is then $(X_o, Y_o, Z_o) = (x_o\,Z_o/f, y_o\,Z_o/f, Z_o)$, where (x_o, y_o) is the location of the centroid of the corresponding object in the image.

Using the conversion factor Z_o/f, other parameters of the object in the image in pixel units, such as length and width, can be converted into length units in the world coordinate system. If the object of interest is the patient's hand, the computed length of the object in the image typically corresponds to the distance between the tip of the middle finger and the wrist, and the object width indicates the width of the palm. Changes of the object in the image space can be related to the actual motion of the body part in the exercise space, for example, a side-to-side motion of a hand on a table that is parallel to the image plane or a fist opening and closing. The conversions from image to world coordinates, provided by the image-analysis module, allows the therapist to monitor patient performance by evaluating spatio-temporal information about the location, orientation (rotation angle in the x-y plane), speed, and direction of motion of the body part in the exercise space.

Some exercises may require that the patient moves the body part in all three dimensions, which means that the distance to the camera is no longer constant. In this case, the length L of the object of interest is measured in advance of the therapy session, and its apparent length l in the image is then used to infer its distance $Z = f\,L/l$ to the camera.

The Interpretation Module

The interpretation module analyzes the spatio-temporal information computed by the image-analysis module and provides appropriate visual and auditory feedback for the patient. It maps the processed camera's view of the exercise space to the chosen display, the computer monitor or projection screen.

Both the analysis and feedback provided by interpretation module depend on the kind of exercise selected. In this section, we describe a set of rehabilitation exercises that patients with physical disabilities might perform during music therapy to improve bi-manual and hand–eye coordination, stretch and strengthen upper limb muscles, or practice manual grasping tasks. The patients' movement either modulates the ongoing soundtrack of a prerecorded piece of music (e.g., their favorite song) or creates a short musical sequence composed of a few notes (melody or rhythm). The recorded pieces of music are stored as Microsoft Wave (WAV) files, and the notes played to create a musical piece use Microsoft's software-based GS Wavetable SW MIDI Synthesizer, which can play, for example, piano-like and drum-like sounds using Roland Sound Canvas patches.

Exercise 1: "Keep the Music Playing"

By moving a body part in a certain predefined manner in the exercise space, the patient activates the playback of a recorded piece of music. The exercise can be used to practice, for example, moving a hand side-to-side or opening and closing a fist. The interpretation module interrupts the music whenever it detects that the patient's actual movements differ from the movements to be practiced (e.g., the patient performs a different type of movement or moves too slowly). To keep the music playing, the subject must move at a speed above a certain threshold, which can be set in advance by the therapist. A patient can thus be challenged to move steadily and quickly.

Recognition of the hand opening and closing motion is performed by evaluating the area, length, and perimeter of the detected hand region in the image over time. The interpretation module decides that the desired hand motion is present if the magnitude of the first and second derivatives of these parame-

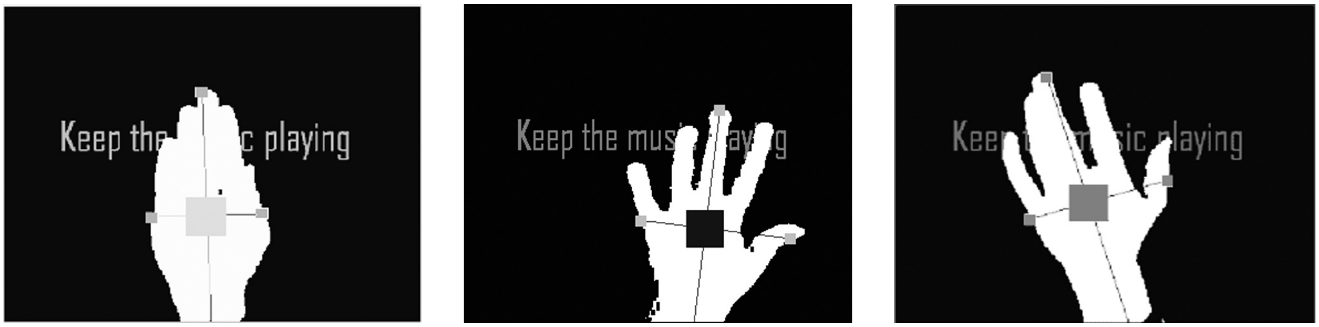


Figure 4. Left: Stationary hand shown with a light-colored centroid. Middle: Fast moving hand, with speed indicated by the dark-gray centroid. Right: Slowly moving hand, with speed indicated by the light-gray centroid.

ters reached certain thresholds. Recognition of side-to-side motion of the hand is performed by evaluating the change of the *x*-coordinate of the hand's centroid over time.

An example display for this exercise is shown in Figure 4. The therapist can select whether or not to provide visual feedback in this exercise, which would be the camera's view of the moving body part projected on the display, as well as a visualization of the velocity by the brightness of the object centroid. A stationary hand is visualized by a light-gray centroid. The faster the hand moves, the darker the centroid becomes.

Exercise 2: "Change the Volume of the Music"

In this exercise, the patient moves a body part in the exercise space while a recorded piece of music is playing. The patient can change the volume of the music by changing the speed of the movement. With this exercise, a patient can be challenged to perform both slow and smooth motions, which produce soft music, and rapid and abrupt motions, which produce loud music. Visual feedback may be selected as in Exercise 1.

The therapist would determine in advance the maximum desired speed, which the system maps to the loudest volume setting. This speed could be based on the distance that a hand can possibly move between two consecutive image captures. We used a transformation from speed to volume that is logarithmic, because human perception of change in sound intensity is apparently logarithmic (Makris 1995). For example, a speed that is one unit slower than the maximum speed is mapped to a sound that is half as loud as the loudest sound.

Exercise 3: "Play a Rhythm"

Here, the patient creates rhythmic sounds by moving in the exercise space and watching visual feedback. Feedback is provided by overlaying the output of the image-analysis module (e.g., the moving hand) onto the display. Different regions of the display correspond to percussion sounds made by different virtual instruments, for example, drums or cymbals. If the patient's body part "touches" a particular region, the corresponding sound is synthesized. If the patient uses both hands, then two instruments can be selected at the same time, allowing the patient to practice bimanual coordination. The exercise can also be performed with music playing in the background; in this case, the patient attempts to accompany the music with a rhythm instrument (e.g., to play as a drummer). An example display that allows the selection of four rhythm instruments is shown in Figure 5.

Exercise 4: "Play a Melody"

In this exercise, patients create a melody by selecting a sequence of tones while watching the processed image of one of their body parts (e.g., hand, finger, or foot) overlaid onto the display. Different regions of the display correspond to notes at different pitches. If the patient uses two hands or fingers to select two regions at the same time, the pitches are played together. The virtual instrument used to synthesize

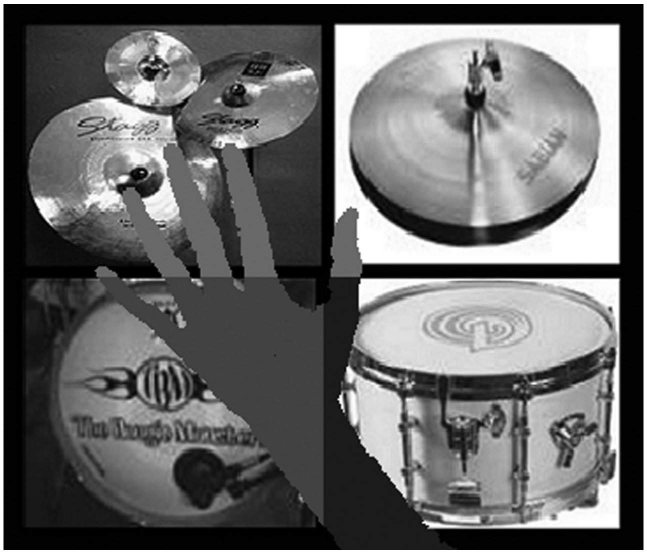


Figure 5. A subject performs the "Play a Rhythm" exercise with a four-region rhythm-instruments display. At the instance shown, most of the hand pixels appear in the lower-left region of the display, covering the drum, which is therefore selected.

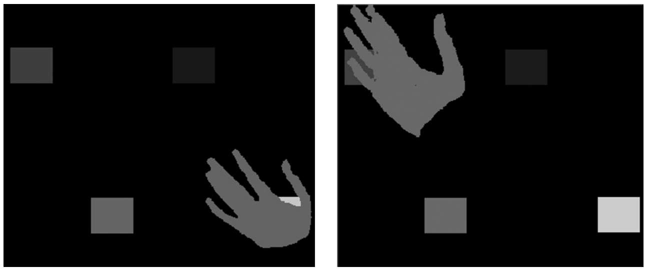


Figure 6. A display used for the "Play a Melody" exercise. In this example, the subject was asked to first select the note represented by the yellow block (bottom right corner of each image) with the left hand and then move it to play the note represented by the red block (top left corner).

the sound and the number of notes and their pitch relations can be selected in advance. These choices allow the therapist to consider the patient's auditory capabilities to differentiate between sounds and to target a particular musical piece. The visual representation of the notes on the display is flexible. One design, shown in Figure 6, uses blocks with different colors to represent the different pitches. An example of an entertaining graphical display is shown in Figure 7, in which seven frogs correspond to seven different pitches.

Exercise 5: "Play a Rhythm or Melody with Volume Changes"

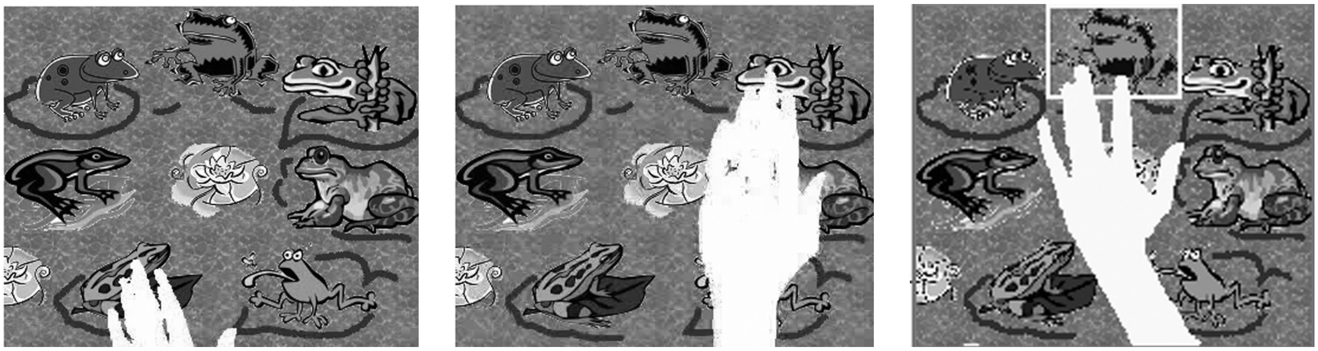
As in the "Play a Rhythm" and "Play a Melody" exercises, the patient here selects notes by reaching into certain regions of the exercise space, and the inter-onset interval from one note to the next depends on the time it takes the patient to move from one region to the next. Additionally, the loudness of auditory feedback depends upon the instantaneous velocity of the body part measured as the patient reaches into a region. A logarithmic transformation is used to map the magnitude of this velocity to the volume of the sound that corresponds to the selected region. Patients thus receive auditory feedback according to how rapidly they selected a note. This feedback simulates, as much as possible, the dynamics of playing a real musical instrument.

Exercise 6: "Follow a Melody"

The "Play a Melody" exercise can be expanded to help a patient learn to play a specific melody. The interpretation module highlights a sequence of display regions—each region for a specific period of time—to teach the patient the pitch and duration of each note of the melody (e.g., the top frog in the display in Figure 7, right). Auditory feedback is given when the patient follows along and reaches to the appropriate regions in the exercise space.

In Exercises 3–6, the display might be arranged such that the user must first traverse regions that were not intended to be selected before selecting the desired area (Figure 7, middle and right). For displays with a small number of easy-to-reach regions, the interpretation module resolves the ambiguity about which region the user intended to select with one of two methods. With the first method, it determines the number of pixels of each display region that is overlaid by the patient's body part. The region most "covered" by the patient's body part is selected. According to the second method, the region closest to the centroid $\{x_o, y_o\}$ of the patient's body part is selected. For the rhythm instrument display in Figure 5, for example, the first method was used. For displays with a larger number of regions, the interpretation module assigns a priority value to each region based on how ambiguous the selection of the region may be. The highest-priority value is assigned to the top-most region of the screen for exercises in which the patient's body part enters the exercise space only from the side that corresponds to the bottom of the display.

Figure 7. The seven-frog display used in the "Play a Melody" and "Follow a Melody" exercises. Left: A note is unambiguously selected by the patient's hand. Middle:The processed hand image overlays three of objects of the display, requiring ambiguity resolution by the interpretation module. Right: The object at the top of the display is highlighted in the "follow a melody" exercise, and the patient's hand is moving appropriately in response.



Experiments and Results

Music Maker was implemented in the EyesWeb 3.0.0 (Camurri et al. 2000) development environment. It runs in real time on a computer with a 3.3 GHz processor and 1 GB of RAM. An M-Audio Delta-192 sound card and a Logitech Quickcam Pro 4000 USB 2.0 camera were used. The camera collects 352 × 288-pixel frames at a rate up to 30 Hz. The color input image contains three 8-bit values for each pixel in the red/green/blue (RGB) color space. The 3.6-mm lens of the camera has a field-of-view of 54 degrees.

We designed a number of experiments to test the accuracy of the image-analysis algorithms, the reliability of the interpretation module, and the potential of Music Maker as a rehabilitation tool. In most of the experiments, the camera was mounted on a tripod that was placed on a table. It was located at a distance of 42.8 cm to the table top, facing downwards. Subjects performed hand exercises by moving their hand slightly above the table while they were resting their elbow on the table (Figure 1, top). The dimensions of the exercise region were 33.7 × 27.9 cm. In experiments involving movements of bare hands, the default values for skin-color detection were used without the need for manual initialization. Experiments involving movements of socked feet required manual initialization.

Motion Trajectory Analysis for Patient Monitoring

This experiment illustrates how a therapist can evaluate a patient's performance by analyzing the spatio-temporal information provided by the image-analysis module. In this experiment, five healthy subjects (mean age: 21 yr) were asked to perform various exercises while the properties of the motion of the body part—location, speed, orientation, and their first and second derivatives—were computed and analyzed. Examples of processed images and trajectories computed during the "Keep the Music Playing" exercise are provided in Figures 8 and 9.

Measuring Speed of Motion

The accuracy of velocity measurements of the system was tested by analyzing the movement of a hand from one side of the exercise space to the other, which lasted about one second. The image-analysis module computed a speed of 11.5 pixels per second. This corresponded to a speed of 33.7 cm/sec. The same motion was timed with a stopwatch, which measured a speed of 27.5 cm/sec—a difference of 18 percent.

Measuring Hand Orientation

We tested the accuracy of the system in determining hand orientation by placing a hand straight on the table at an angle of 90 degrees with the *x*-axis. The image-analysis module determined an orientation of 87 degrees, which is an error of only 3 percent. The system was able to detect the orientation of a hand tilted sideways by 45 and 135 degrees, as measured with a ruler, but with less accuracy. It computed respective orientations of 50 and 130 degrees, which represents an 11 percent error.

Figure 8. Opening and closing of the fist of subject 3 (first row) and subject 4 (second row). Subject 4 moves much faster than subject 3, as indicated by the colors of the centroids.

Figure 9. Subject 5 performed ten side-to-side hand motions in about 20 sec while listening to a piece of music. Frequency analysis of the position of the centroid (light graph, left) and its speed (light graph, right) indicates that the motion can be modeled by sinusoids with corresponding dominant frequencies of about 0.5 Hz (dark graph, left) and 1 Hz (dark graph, right).

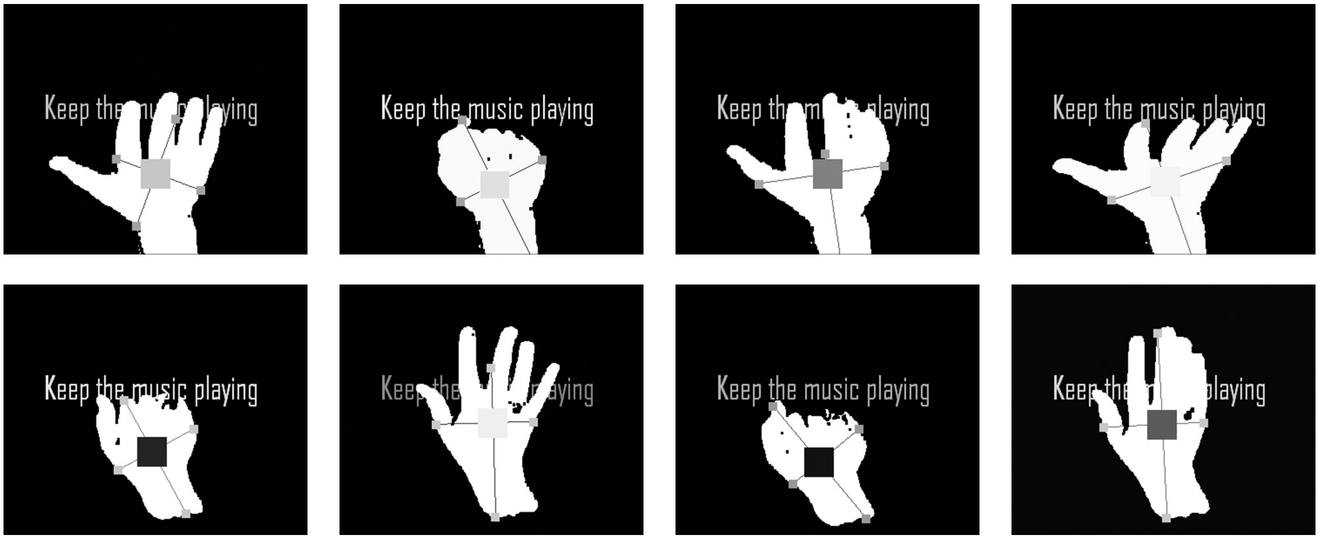


Figure 8

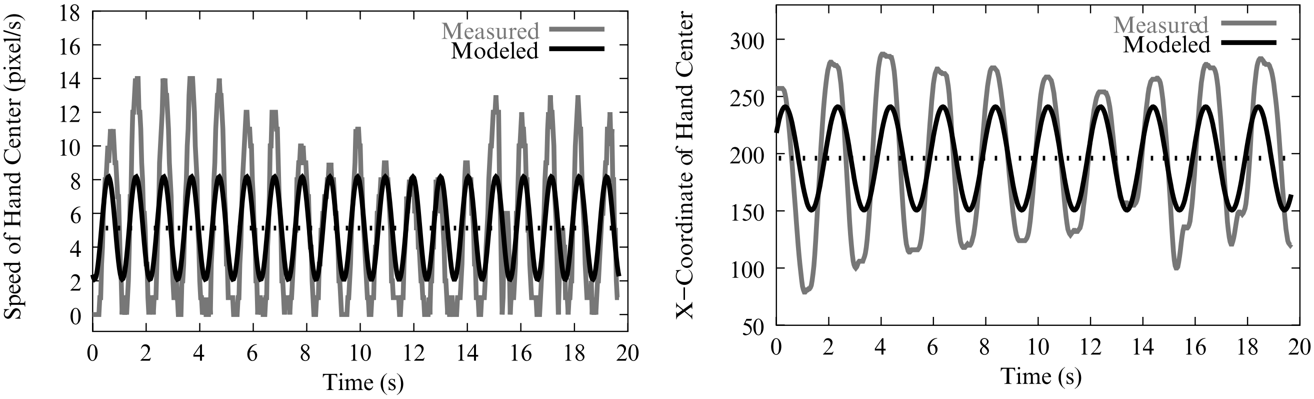


Figure 9

Lighting Conditions

Most of our tests were performed under indoor lighting conditions typical for the intended use of our system, but we also tested Music Maker under low-lighting conditions to evaluate its robustness. Music Maker works well in a bright laboratory environment, because the imaged body part and the dark background of the exercise space can be set apart reliably by color thresholding. In dark lighting conditions, however, the patient's body part starts to blend in with the dark background in the video. Low lighting caused dark input images (see Figure 10, left) and noisy hand segmentation results (Figure 10, middle), but nonetheless allowed motion of hand and fist to be detected reliably (Figure 10, right).

Experiment with an Older Subject

We tested three different setups of Music Maker with an older subject (57 yr) without disabilities

Figure 10. The effect of low illumination. Left: Hand image recorded under low-light conditions. Middle: Corresponding noisy hand segmentation results during the "Keep the Music Playing" exercise. Right: Side-to-side movements of the hand were detected based on the pronounced sinusoidal patterns of the x-coordinate of the hand centroid. The subsequent opening-closing movements of the hand, which produced much less pronounced patterns of the x-coordinate of the centroid, were detected by a combined analysis of the changes in hand area, length, and perimeter over time.

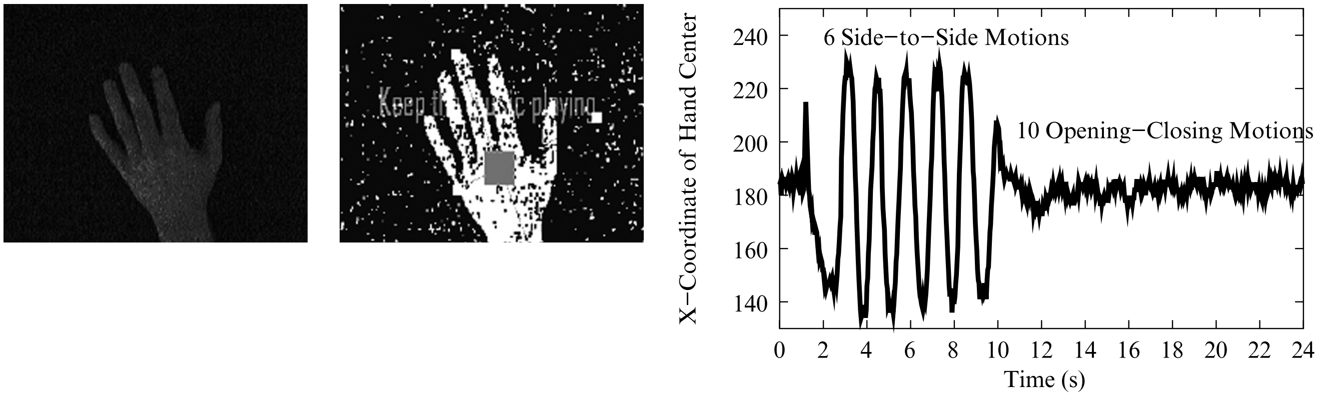


Table 1. Number of Correctly (Incorrectly) Matching Notes among 30 (120) Choices and Average True (False) Positive Detection Rates

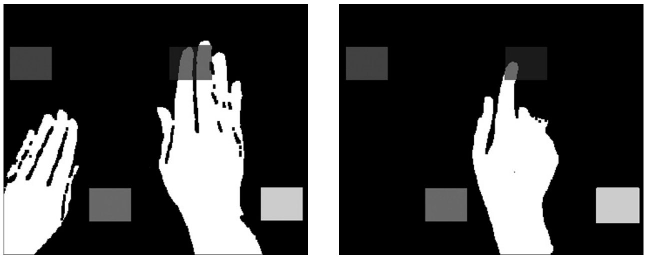
	Subject 1	*Subject 2*	*Subject 3*	*Subject 4*	*Subject 5*	*Avg. Rate/Test, %*
Test 1	30 (2)	22 (6)	28 (6)	30 (3)	27 (3)	91 (4)
Test 2	30 (6)	23 (3)	30 (6)	30 (3)	30 (0)	95 (3)
Test 3	21 (11)	30 (2)	30 (6)	30 (2)	29 (2)	93 (3)
Avg. Rate, %	90 (5)	83 (5)	98 (3)	100 (2)	96 (1)	93 (3)
Average Rate of all Correct (Incorrect) Detections						93 (3)

who did not have a musical background. In the first setup, the subject was moving his hand while sitting on a chair and watching the visual feedback on a computer monitor (Figure 1, top left). In the second setup, the subject was moving his hand while lying supine in a bed (Figure 1, bottom left). In the third setup, he was moving his socked foot while sitting on a chair (Figure 1, bottom middle and right). In all three setups, the subject was asked to play notes using the seven-frog display. The subject was able to reach regions in the exercise space that corresponded to the highest and lowest notes with both his hand and foot.

Experiments to Test Reaching Ability

We conducted an experiment to test the reaching accuracy of five subjects, ages 18–22, who were not musically trained. They were asked to play notes by reaching with a hand into the corresponding regions of the exercise space described in the "Play a Melody" exercise. Music Maker provided visual cues by highlighting the appropriate block on the four-note display shown in Figure 6. If the subject's hand entered the region of the exercise space that corresponded to the cue, it was noted that the subject played a correctly matching note; otherwise, an error was recorded. During the 30-sec exercise period, the system pseudo-randomly selected and highlighted one of four possible choices, each for the duration of 1 sec. The same sequence was used for each subject. The results in Table 1 indicate that the subjects were able to reach the appropriate regions within the given time limit most of the time. The true positive detection rate, that is, the number of correctly selected notes divided by the number of notes in the sequence (30), was 93 percent. Subjects rarely selected notes by mistake: the false-positive rate, that is, the number of falsely selected notes

Figure 11. The hands of subjects 1 and 2 while following a melody.



divided by the number of possible notes (120), was only 3 percent.

Experiments with "Follow a Melody"

The test of the "Follow a Melody" exercise involved the same five subjects and four-note display as in the previous experiment (see Figure 11). During the training phase, subjects were visually cued to play a 10-note melody by moving along a sequence of highlighted interface blocks. In the test phase, they were then asked to repeat the motion sequence and play the melody without visual cues. Subjects needed 16 train-and-test trials, on average, until they were able to play the melody without making any mistakes.

System Latency and Accuracy

We asked five additional subjects, ages 23–30 and also without musical training, to perform the exercises "Play a Melody" and "Play a Melody with Volume Changes" and rank the accuracy of the system on a scale of 0 (poor) to 5 (excellent). For both exercises, four subjects reported that they experienced no discrepancy between the sounds they intended to make and the sounds the system played (rank 5). One subject ranked the accuracy of hand detection in the first exercise to be 3 (average) and the accuracy of the velocity-to-volume mapping in the second exercise to be 4 (very good). All of the subjects reported that they could not perceive any latency between their movements and the corresponding sounds played.

Discussion and Conclusions

Successful physical rehabilitation of patients requires creativity, ingenuity, and flexibility. Music Maker was developed to support these needs. It has the potential to provide a rehabilitation environment that is motivating, effective, and safe. The exercises we have described can be used for testing, practicing, and improving a patient's motor functions. It is an important characteristic of Music Maker that it provides quantitative tools so that the recovery process can be monitored. Music Maker can measure and evaluate the properties of a person's movements, such as range, speed, and steadiness; therapeutic outcomes could thus be described quantitatively.

Our experiments showed that subjects can quickly learn how to use Music Maker and produce desired audible results. Music Maker can be used to exercise different body parts involving feet, hands, or fingers; it is thus a flexible tool that can adjust to the exercise needs of specific patients. Music Maker can also be used for exercising while in different body positions. The test with the subject lying on a bed showed the potential of the system to provide patients the option to start rehabilitation while still lying in a hospital bed. Patients may also want to use Music Maker as a rehabilitation tool at home, because it uses portable, relatively inexpensive equipment and is easy to set up.

Our experiments with healthy subjects showed that Music Maker can be used to measure the spatio-temporal properties of their movements. The motion trajectories of healthy subjects may be helpful to quantitatively establish the patterns associated with healthy movement and may serve as "baselines" against which movement data collected from members of clinical populations could be compared. This may facilitate quantitative assessment of a patient's motor functions and their improvements over time.

To provide music therapy for quadriplegic patients, Music Maker could be used in combination with other video-based interfaces, for example, the Camera Mouse (Betke, Gips, and Fleming 2002), which is a computer access device that has been adopted by many users with severe disabilities in the United States, Great Britain, and Ireland. Exercises, such as "Keep the Music Playing" or "Play a Melody" could be performed with the Camera Mouse, which tracks body features such as the tip of the user's nose or chin. Another interface for people with severe disabilities is the BlinkLink (Grauman et al. 2003; Chau and Betke 2005), which automatically detects a user's eye blinks and accurately measures their durations. Eye blinks could be used to control virtual rhythm instruments, and their durations could be mapped to the length or pitch of a sound.

In future work, we will provide more detailed modeling of the hand and forearm, which would allow monitoring of translational and rotational motions of the palm (ulnar or radial), wrist (flexion or extension), and forearm (in pronation or supination). Once Music Maker can recognize these motions, additional exercises involving these motions will be designed, for example, closing and turning a fist which may help patients practice gripping and turning a door knob. Another example would be the "Reach to a Presented Object" exercise, where objects appear at random locations on the display for a limited time and the patient needs to "hit" them in order to produce a sound. With this exercise, the patient could practice to move quickly towards a target and aim accurately. The goal of the exercise could be formulated as a goal of a game, in which the patient wins if he or she successfully "hits" all objects in a given amount of time. Exercises in form of games may serve as great motivators in the long and tedious process of rehabilitation. By competing to win the game, a patient may reach the exercise goal faster.

In summary, we provided a camera-based interface that may serve as a therapeutic device for physical rehabilitation, helping patients with physical disabilities to learn or regain motor skills that they need for the activities of daily living. The technological development of Music Maker will continue in the future, and clinical studies will be performed to test its rehabilitative potential.

Acknowledgments

Financial support by the NSF (IIS-0093367, IIS-0329009, IIS-0308213, EIA-0202067), NIH (R01-DC-03663), and Dudley Allen Sargent Research Fund is gratefully acknowledged.

References

Akram,W., L. Tiberii, and M. Betke. 2006. "A Customizable Camera-Based Human-Computer Interaction System Allowing People with Disabilities Autonomous Hands-Free Navigation of Multiple Computing Tasks." In *Universal Access in Ambient Intelligence Environments: 9th ERCIM Workshop on User Interfaces for All, Königswinter, Germany, September 27–28, 2006, Revised Papers,* Ed. Constantine Stephanidis and Michael Pieper, 28–43. Berlin: Springer.

Betke, M., J. Gips, and P. Fleming. 2002. "The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People with Severe Disabilities." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10(1):1–10.

Bongers, B., and Y. Harris. 2002. "A Structured Instrument Design Approach: The Video-Organ." *Proceedings of the 2002 International Conference on New Interfaces for Musical Expression.* New York: Association for Computing Machinery, pp. 86–91.

Borchers, J. O., and M. Muhlheuser. 1998. "Design Patterns for Interactive Musical Systems." *IEEE Multimedia* 5(3):36–46.

Brooks, T., et al. 2002. "Interaction with Shapes and Sounds as a Therapy for Special Needs and Rehabilitation." *Proceedings of the 4th International Conference Disability, Virtual Reality, and Associated Technologies.* Veszprem, Hungary, pp. 205–212.

Burns, A.-M., and M. M. Wanderley. 2006. "Visual Methods for the Retrieval of Guitarist Fingering." *Proceedings of the 2006 Conference on New Interfaces for Musical Expression.* New York: Association for Computing Machinery, pp. 196–199.

Camurri, A., et al. 2000. "EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems." *Computer Music Journal* 24(1):57–69. (Additional information on Eyesweb is available on-line at www.infomus.dist.unige.it.)

Camurri, A., et al. 2003. "Application of Multimedia Techniques in the Physical Rehabilitation of Parkinson's Patients." *Journal of Visualization and Computer Animation* 14:269–278.

Camurri, A., et al. 2004. "Toward Real-Time Multimodal Processing: Eyesweb 4.0." Paper presented at the AISB 2004 Convention on Motion, Emotion, and Cognition, March, Leeds, UK.

Chau, M., and M. Betke. 2005. "Real-Time Eye Tracking and Blink Detection with USB Cameras." Technical Report 2005-012, Computer Science Department, Boston University, April. Available on-line at www.cs.bu.edu/techreports/pdf/2005-012-blink-detection.pdf.

Coniglio, M., and D. Stoppielo. 2007. "MidiDancer: A Wireless Movement Sensing System." Available online at www.troikaranch.org/mididancer.html. Accessed January 2007.

Dobrian, C., and F. Bevilacqua. 2003. "Gestural Control of Music: Using the Vicon 8 Motion-Capture System." *Proceedings of the 2003 Conference on New Interfaces for Musical Expression.* New York: Association for Computing Machinery, pp. 161–163.

Grauman, K., et al. 2003. "Communication via Eye Blinks and Eyebrow Raises: Video-Based Human-Computer Interfaces." *Universal Access in the Information Society* 2(4):359–373.

Griffith, N., and M. Fernstrom. 1998. "LiteFoot: A Floor Space for Recording Dance and Controlling Media." *Proceedings of the 1998 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 475–481.

Harris, B., and M. Summa-Chadwick. 2005. "A Computerized System for Neurologic Music Therapy." *Journal of Computing Sciences in Colleges* 21(2):250–257.

Hsu, R.-L., M. Abdel-Mottaleb, and A. K. Jain. 2002. "Face Detection in Color Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5):696–706.

Horn, B. K. P. 1986. *Robot Vision.* Cambridge, Massachusetts: MIT Press.

Hunt, A., R. Kirk, and M. Neighbour. 2004. "Multiple Media Interfaces for Music Therapy." *IEEE Multimedia* 11(3):50–58.

Jain, R., R. Kasturi, and B. Schunk. 1995. *Machine Vision.* New York: McGraw-Hill.

Jensenius A. R., R. Godoy, and M. M. Wanderley. 2005. "Developing Tools for Studying Musical Gestures within the Max/MSP/Jitter Environment." *Proceedings of the 2005 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 282–285.

Johnstone, E. 1991. "A MIDI Foot Controller:The PodoBoard." *Proceedings of the 1991 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 123–126.

Kirk, R., et al. 1994. "Computer Music in the Service of Music Therapy: The MIDIGRID and MIDICREATOR Systems." *Medical Engineering and Physics* 16(3):253–258.

Lahav, A. 2005. "Musical Solution to Motor Dysfunctions." Invited presentation given to the International Conference on Enabling Environment: Music, Technology, Culture, and Healthcare. New York, 9 June.

Lahav, A., et al. 2005. "The Power of Listening: Auditory-Motor Interactions in Musical Training." *Annals of the New York Academy of Sciences* 1060:189–194.

Lahav, A., E. Saltzman, and G. Schlaug. (2007). "Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions." *Journal of Neuroscience* 27(2):308–314.

Leider, C. 1997. "A Real-Time MIDI Library for an Inexpensive Microcontroller-Based Interactive Performance system." *Proceedings of the 1997 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 482–484.

Lewis-Brooks, A., and S. Hasselblad. 2004. "Creating Aesthetically Resonant Environments for the Handicapped, Elderly, and Rehabilitation." *Proceedings of the 5th International Conference on Disability, Virtual Reality, and Associated Technologies,* Oxford, UK, pp. 191–198.

Magee, J. J., et al. 2004. "EyeKeys: A Real-Time Vision Interface Based on Gaze Detection from a Low-Grade Video Camera." Paper presented at the 2004 IEEE Workshop on Real-Time Vision for Human-Computer Interaction, 2 July, Washington, D.C.

Makris, N. C. 1995. "A Foundation for Logarithmic Measures of Fluctuating Intensity in Pattern Recognition." *Optics Letters* 20:2012–2014.

Miranda, E. R., and M. M. Wanderley. 2006. *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard.* Middleton, Wisconsin: A-R Editions.

Morales-Manzanares, R., et al. 2001. "SICIB: An Interactive Music Composition System Using Body Movements." *Computer Music Journal* 25(2):25–36.

Nakra, T. M. 1999. "Searching for Meaning in Gestural Data: Interpretive Feature Extraction and Signal Processing for Affective and Expressive Content." In M. Wanderley and B. B. Rovan, eds. *CD: Trends in Gestural Control of Music.* Paris: Institut de Recherche et de Coordination Acoustique/Musique.

Pacchetti, C., et al. 2000. "Active Music Therapy in Parkinson's Disease: An Integrative Method for Motor and Emotional Rehabilitation." *Psychosomatic Medicine* 62(3):386–393.

Paradiso, J. A., et al. 2000. "Design and Implementation of Expressive Footwear." *IBM Systems Journal* 39(3/4):511–529.

Paradiso, J. A., et al. 2004. "Interactive Therapy with Instrumented Footwear." *Proceedings of CHI '04 Extended Abstracts on Human Factors in Computing Systems.* New York: Association for Computing Machinery, pp. 1341–1343.

Paul, S., and D. Ramsey. 2000. "Music Therapy in Physical Medicine and Rehabilitation." *Australian Occupational Therapy Journal* 47:111–118.

Rémus, J. 2006. "Non-Haptic Control of Music by Video Analysis of Hand Movements: 14 Years of Experience with the 'Caméra Musicale.'" *Proceedings of New Interfaces for Musical Expression.* New York: Association for Computing Machinery, pp. 250–253.

Schwerdt, K., and J. L. Crowley. 2000. "Robust Face Tracking Using Color." *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition.* New York: Institute of Electrical and Electronics Engineers, pp. 90–95.

Siegel, W., and J. Jacobsen. 1998. "The Challenges of Interactive Dance: An Overview and Case Study." *Computer Music Journal* 22(4):29–43. (More information is available on-line at hjem.get2net.dk/diem/gen-info.html.)

Sparacino, F., G. Davenport, and A. Pentland. 2000. "Media in Performance: Interactive Spaces for Dance, Theater, Circus, and Museum Exhibits." *IBM Systems Journal* 39(3/4):479–510.

Tarabella, L. 2004. "Improvising Computer Music: An Approach." Paper presented at the International Conference on Sound and Music Computing '04, 20–22 October, Paris.

Tramo, M. J. 2001. "Biology and Music: Music of the Hemispheres." *Science* 291(5501):54–56.

Van Raalte, C. 1999. "Old Dog, New Tricks: Biofeedback as Assistive Technology." Paper presented at the California State University Northridge Center On Disabilities 1998 Conference, Los Angeles, CA. Available on-line at www.csun.edu/cod/conf/1998/proceedings/csun98.htm.

Winkler, T. 1995. "Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music." *Proceedings of the 1995 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 261–264.

Wren, C., et al. 1997a. "PFINDER: Real-Time Tracking of the Human Body." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):780–785.

Wren, C., et al. 1997b. "Perceptive Spaces for Performance and Entertainment Untethered Interaction Using Computer Vision and Audition." *Applied Artificial Intelligence* 11(4):267–284.

Wu, H., Q. Chen, and M. Yachida. 1999. "Face Detection from Color Images Using a Fuzzy Pattern Matching Method." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(6):557–563.