Boston University Computer Science Technical Report No. 2005-032

MusicCamera – A Camera-based Music Making Tool for Physical Rehabilitation

Mikhail Gorman[1], Margrit Betke[1], Elliot Saltzman[2,3], and Amir Lahav[2,4]

Boston University**

Abstract. The therapeutic effects of playing music are being recognized increasingly in the field of rehabilitation medicine. People with physical disabilities, however, often do not have the motor dexterity needed to play an instrument. We developed a camera-based human-computer interface called "MusicCamera" to provide such people with a means to make music by performing therapeutic exercises. MusicCamera uses computer vision techniques to convert the movements of a patient's body part, for example, a finger, hand, or foot, into musical and visual feedback using the open software platform EyesWeb. It can be adjusted to a patient's particular therapeutic needs and provides quantitative tools for monitoring the recovery process and assessing therapeutic outcomes. We tested the potential of MusicCamera as a rehabilitation tool with six subjects who responded to or created music in various movement exercises. In these proof-of-concept experiments, MusicCamera has performed reliably and shown its promise as a therapeutic device.

1 Introduction

Music is universal among human cultures [1]. People enjoy both listening to music and making music, and they naturally respond to music with motion. Recent brain imaging studies have shown that, given appropriate auditory inputs, humans seem to be "tuned" to produce corresponding motor outputs [2, 3]. 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 accompanying music [4–6]. Playing an instrument, such as piano, guitar or drums, may be very

** 1 Image and Video Computing Group, Department of Computer Science, Boston University, 2 The Music Mind & Motion Lab, Sargent College for Health and Rehabilitation Science, Boston University, 3 Haskins Laboratories, New Haven, CT, 4 The Music & Neuroimaging Lab, Beth Israel Deaconess Medical Center, Harvard Medical School

difficult or even infeasible for patients with motor dysfunctions. As an alternative, easy-to-use tool for active music therapy, we have designed "MusicCamera," a human-computer interface that converts body movements into musical and visual feedback in real time using the open software platform EyesWeb [7]. MusicCamera 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 MusicCamera detects and interprets these movements. Detection is accomplished without any sensors or markers that must be attached to the patients' bodies. MusicCamera only relies on the video input from a camera to observe patient motion and on computer-vision techniques to analyze the motion.

MusicCamera is a flexible, adaptive interface that supports two kinds of therapeutic exercises. The patient either listens to a piece of music and reacts with movement of a body part, for example, a finger, hand, foot, or limb, or, more importantly, the patient creates a piece of music by such movements. MusicCamera provides auditory and visual feedback which can be adjusted to the patient's needs. Musical feedback could range from a single piano note to a recording of the patient's favorite song. Visual feedback is provided by a graphical display on a computer monitor or wall-mounted screen. MusicCamera uses fun displays, for example, cartoon drawings or pictures of music instruments. Its hardware setup can be adjusted according to the patient's level of impairment, his or her particular therapeutic goals, and the equipment available in a hospital or patient's home (see Figure 1 for sample setups).

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 MusicCamera, we were particularly influenced by the EyesWeb work in the Laboratorio di Informatica Musicale in Italy (e.g., [7–9]) and our own experiences in developing a number of human-computer interfaces for people with severe disabilities who use cameras to access the computer [10–13]. 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 [14], create graphical displays, and provide visual and audio output.

EyesWeb has mostly been used to create tools for human-music interactions in large performance spaces (e.g., to facilitate interactive dance performances [8]), but its potential as a design tool for therapeutic exercises has also been explored previously [9]. Camurri et al. [9] employed camera-based full-body tracking to create pilot exercises such as "Stand and Sit" and "Painting by Aerial Gestures." Our focus has instead been to develop tools for detection and tracking of smaller body parts, such as hands or feet (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. MusicCamera provides quantitative tools for analyzing and monitoring these movement measures, for example, the range of motion of hands or feet, the speed and frequency of

Fig. 1. MusicCamera in use. It observes the movements of the users' hands (top row and bottom left) or 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) and a screen display (top right).

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.

The EyesWeb platform is one of several camera-based human-computer interfaces that produce music from body movements, which include DanceSpace [15, 16], BigEye [17], Music-via-Motion [18] and the Very Nervous System [19]. These systems detect movements within pre-defined performance areas and trigger sounds in real time. DanceSpace uses a computer-vision gesture-recognition system [20] to track a performer's body parts and can map different instruments to these parts, while a melodic base tune is playing in the background. The performer mimes, for example, playing a virtual cello with her hands and a drum with her foot. The spatial extend of movements are matched to pitches of notes. With the BigEye system [17], a user can define objects and spatial zones that are of interest to him or her. Configured in this manner, the BigEye 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.

Systems without cameras that produce music from body movements have been surveyed by Winkler [21] and Morales-Manzarnares et al. [22]. These include interfaces with touch sensors placed on the floor (e.g., [23, 24]) or attached to the body of the user or his clothes. Paradiso, Hsiao, and Hu [25] created sensors that were placed onto the shoes of dancers and used for interactive dance. The Kontrolldress [26] is a body suit equipped with sensors that a dancer can tap with his or her fingertips to produce sounds. In other systems, such as BodySynth by Van Raalte [27], electromyographical 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 wide variety of movements [28]. The MidiDancer [29] system and the system Siegel and Jacobsen [30] use flex sensors that measure how much a dancer's joints, e.g., elbows, wrists, hips, and knees, are bent and 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 [31], sliders and table rings [32], or light-sensitive drum sticks [33]. Morales-Manzarnares et al. [22] used the Flock-of-Birds sensor system [34] to develop a system for music composition, improvisation, and performance based on body movements. The Flock-of-Birds sensor system is a popular body tracker that has often been used for virtual reality applications. It measures the motions of a person who wears electromagnetic sensors that sample the magnetic fields emitted by an external transmitter. Such 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, since sensors do not need to be attached to the user's body, and are therefore particularly appropriate for therapy purposes.

2 Method

MusicCamera 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, i.e., 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 MusicCamera is shown in Figure 2.

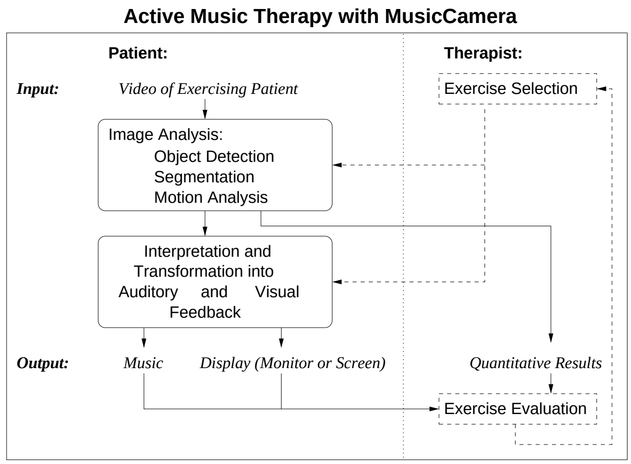

Fig. 2. Conceptual overview of the use of MusicCamera (solid lines) in active music therapy. The therapist can select and monitor exercises and use system outputs to adjust exercises (dashed lines).

2.1 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., [35–37]). 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 (e.g., the red and green values of skin may range between 40–255, and blue values between 0–40). If the object of interest is the patient's foot, the color of the patient's sock or shoe may 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 manual initialization is convenient because it makes the system flexible and allows different kinds of body parts, such as hands, arms, or feet, to be detected. It also makes the system more reliable, since skin tones vary widely over the population and pixel colors depend on the lighting conditions and camera selection. To simplify the detection of the body part even further, a black background, for example, black cloth, may 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 color values and applies a one-pixel erosion operation [38] in order 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 [39] of the foreground pixels. The orientation of the object is computed by determining the orientation of the axis of least inertia [39]. 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 [39] 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 $x_{\max} = fX/Z$ and $y_{\max} = fY/Z$. 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, while 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 $X \times Y$ 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 distanceb $Z = fL/l$ to the camera.

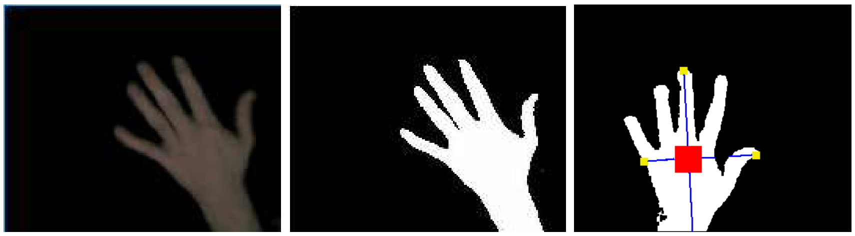

Fig. 3. Left: Camera view of a subject's hand. Middle: Filtered binary image of detected hand. Right: Binary hand image with centroid in red, axes of least and most inertia in blue, intersection points in yellow.

2.2 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 the following section, we describe exercises in which the patients must move their hands or feet in order to play the soundtrack of a prerecorded piece of music (e.g. their favorite song) or to create a short musical sequence composed of a few notes (melody or rhythm).

Exercise 1: "Keep the Music Playing." By moving a body part in a certain predefined manner in the exercise space, the patient activates the play back 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, for example, 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 parameters 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 color of the object centroid. A stationary hand is visualized by a red centroid. The faster the hand moves the greener the centroid becomes.

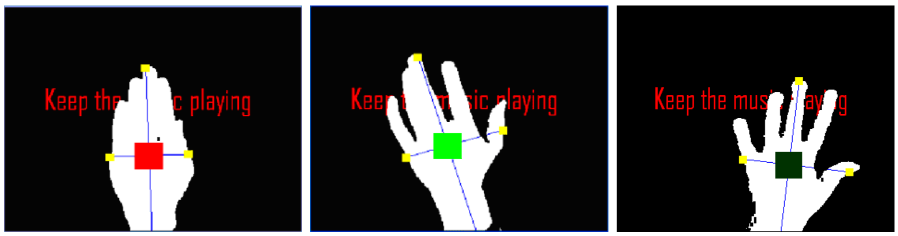

Fig. 4. Left: Stationary hand shown in a red centroid. Middle: Fast moving hand. The speed is indicated by green centroid. Right: Slowly moving hand. The speed is indicated by a dark green centroid.

Exercise 2: "Change the Volume of the Music." 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 which produce loud music. Visual feedback may be selected as in Exercise 1.

The therapist determines 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, since human perception of change in sound intensity is apparently logarithmic [40]. 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." The patient creates rhythmic sounds by moving in the exercise space and watching the 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 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, he or she can select two instruments at the same time. The exercise can also be performed with music playing in the background. In this case, the patient's role may be to accompany the music with a rhythm instrument. An example display that allows the selection of four rhythm instruments is shown in Figure 5.

Exercise 4: "Play a Melody." The patient creates a melody by selecting a sequence of tones while watching the processed image of his or her body part

Fig. 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.

(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 blended together. The virtual instrument used to synthesize the sound and the number of notes and their pitch relations can be selected in advance. These choices allow the therapist to consider the patients auditory capabilities to differentiate between notes 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 a fun graphical display is shown in Figure 7, in which seven frogs correspond to seven different pitches.

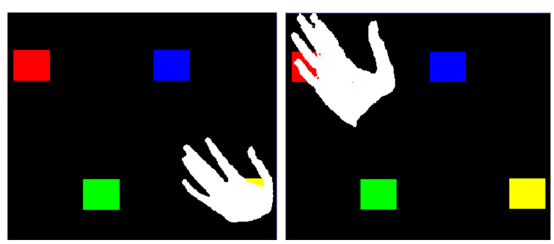

Fig. 6. A display used for the "play a melody" exercise. In this example, a subject was asked to first select the note represented by the yellow block with his left hand and then move it to play the note represented by the red block.

Exercise 5: "Play a Rhythm or Melody with Volume Changes." As in the "play a rhythm" and "play a melody" exercises, the patient selects notes by reaching into certain regions of the exercise space. Here, the instantaneous velocity of the body part, measured as the patient reaches into a region, is mapped to the volume of the note corresponding to that region. The patient thus receives auditory feedback according to how rapidly he or she selected a note. This feedback simulates, as much as possible, the dynamics of playing a

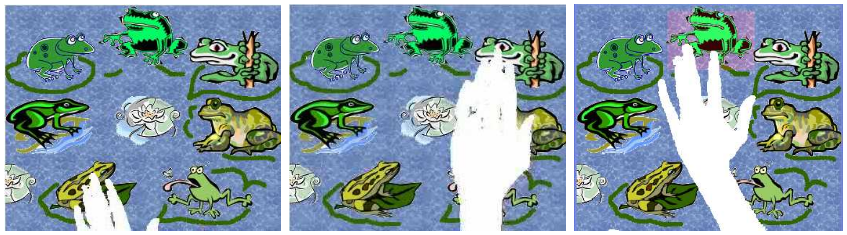

Fig. 7. The seven-frog display used in the "play a melody" and "follow a melody" exercises. Left: A note is unambiguously selected by the patient's hand. Middle: The processed hand image overlays three of objects of the display, requiring ambiguity resolution by the interpretation module. Right: The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

real musical instrument. As in the "change the volume of the music" exercise, a logarithmic transformation from speed to volume is used.

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 length 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 regions on the display might be arranged in a way so that the user cannot select a region without traversing regions he or she may not intend to select (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: (1) 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. (2) 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.

3 Experiments and Results

MusicCamera was implemented in the EyesWeb [8] development environment. It runs in real time on a computer with a 3.3 GHz processor and 1GB of RAM. A M-Audio Delta-192 sound card, Microsoft MIDI synthesizer, 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 MusicCamera 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 cm $\times$ 27.9 cm.

Motion Trajectory Analysis for Patient Monitoring. This proof-of-concept experiment illustrates how a therapist may evaluate a patient's performance by analyzing the spatio-temporal information provided by the image analysis module. In this experiment, five healthy subjects 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.

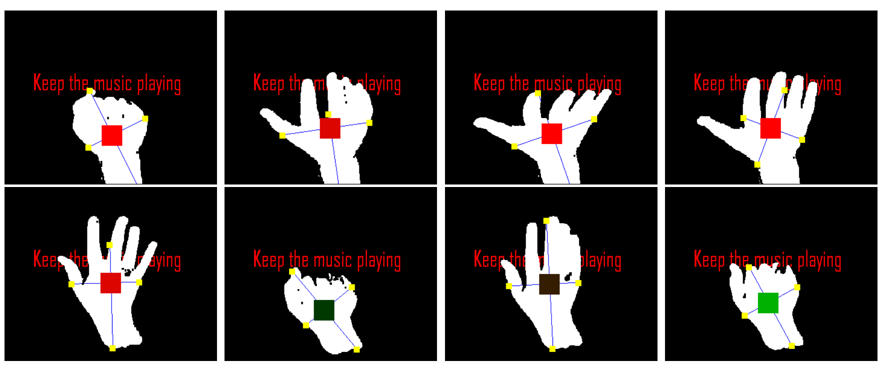

Fig. 8. Opening and closing of the fist of subject 3 (first row) and subject 4 (second row). Subject 4 moves much faster than subject 3, as uindicated by the color of the centroids.

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

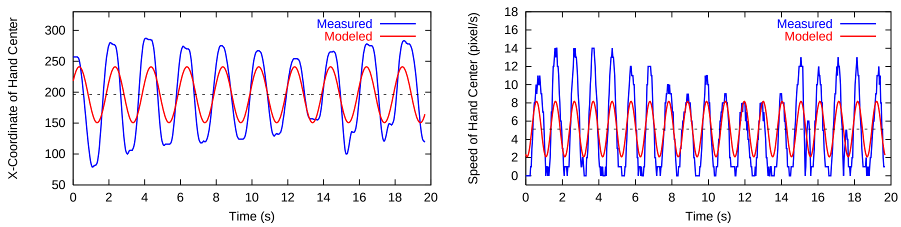

Fig. 9. Subject 5 performed ten side-to-side hand motions in about 20 seconds while listening to a piece of music. Frequency analysis of the position of the centroid (blue graph, left) and its velocity (blue graph, right) indicates that the motion can be modeled by sinusoids with corresponding dominant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

of 33.7 cm/s. The same motion was timed with a stop watch, which measured a speed of 27.5 cm/s, a difference of 18%.

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 degree with the x-axis. The image analysis module determined an orientation of 87 degrees, which is an error of only 3%. 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 is an 11% error.

Lighting conditions. Most of our tests were performed under indoor lighting conditions typical for the intended use of our system, but we also tested MusicCamera under low lighting conditions to evaluate its robustness. MusicCamera works well in a bright laboratory environment, since the imaged body part and the dark background of the exercise space can be set apart reliably by color thresholding. In dark lighting conditions, the patient's body part starts to blend in with the dark background in the video. Low lighting caused dark input images (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 Elderly Subject. We tested three different setups of MusicCamera with an elderly subject. 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 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

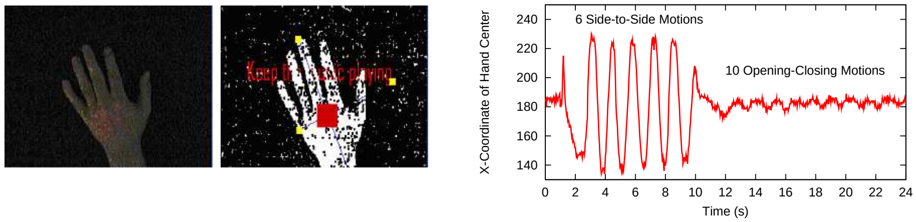

Fig. 10. Impact of Low Illumination. Left: Hand image recorded under low-light conditions. Middle: Resulting noisy hand segmentation results during the "Keep the Music Playing" exercise. Right: Side-to-side movements of the hand were detected based on the pronounced sinusoidal patterns of the x-coordinate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

corresponded to the highest and lowest notes both with his hand and foot. He was also able to play notes by selecting display regions (here, cartoon-pictures of frogs) that were highlighted.

Experiments with "Play a Melody" Exercise. We conducted the test of the "play a melody" exercise with five subjects. The subjects interacted with the four-note display shown in Figure 6. They were asked to play a note by reaching with a hand into the corresponding region of the exercise space. Visual feedback was given by highlighting the appropriate block on the display. Each note was displayed for one second. We recorded the number of correctly and incorrectly matching notes played by five subjects during the exercise. During the 30-second exercise period, the system pseudo-randomly selected and highlighted one of four possible choices for the duration of a second. 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 times. The true positive detection rate, i.e., the number of correctly selected notes divided by the number of notes in the sequence (30), was 93%. Subjects rarely selected notes by mistake. The false positive rate, i.e., the number of falsely selected notes divided by the number of possible notes (120), was only 3%.

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	Av. Rate per 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%)
Av. Rate	90% (5%)	83% (5%)	98% (3%)	100% (2%)	96% (1%)	
Av. Rate of All Correct (Incorrect) Detections:						93% (3%)

Experiments with "Follow a Melody" Exercise. The test of the "Follow a Melody" exercise involved the same five subjects and four-note display as in the previous experiment. A 10-note sequence was displayed visually by highlighting appropriate interface blocks. The subjects were able to move along and thus play the corresponding melody. When asked to repeat the motion and play the melody without visual support of highlighted blocks, they needed 16 trials, on average, until they were able to play it without making any mistakes. This indicates that visual support could be useful in active music therapy to avoid the cognitive load of patients asked to follow a melody.

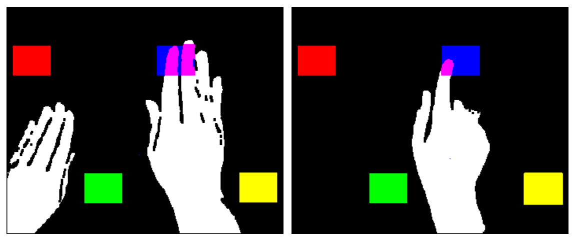

Fig. 11. The hands of subjects 1 and 2 while following a melody.

4 Discussion and Conclusions

Successful physical rehabilitation of patients requires creativity, ingenuity and flexibility. MusicCamera is a human-computer interface that supports these needs. It has the potential to provide a rehabilitation environment that is particularly motivating, effective and safe. The exercises we have described may be used for testing, practicing, and improving a patient's motor functions. It is an important characteristic of MusicCamera that it provides quantitative tools for monitoring the recovery process. It can measure and evaluate the properties of patient movements, such as range, speed, and steadiness. Therapeutic outcomes can thus be described quantitatively.

Our experiments showed that subjects can quickly learn how to use MusicCamera and produce the sought-after sounds. MusicCamera can be used to exercise different body parts involving feet, hands or fingers. MusicCamera is thus a flexible tool that can adjust to the exercise needs of specific patients. MusicCamera can also be used for exercising while in different body positions. The test with the subject lying on a bed was particularly important, because it showed the potential of MusicCamera to help patients who are unable to get out of bed. MusicCamera could provide patients the option to start rehabilitation while still lying in a hospital bed. Patients may also want to use MusicCamera as a rehabilitation tool at home since it uses portable, relatively inexpensive equipment and is easy to set up.

Although our experiments were limited to healthy subjects, they provide a proof of concept that MusicCamera can be used to measure the spatio-temporal properties of movements. The motion trajectories of healthy subjects may be helpful to quantitatively establish the patterns associated with healthy movement and may serve as "base lines" 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, MusicCamera could be used in combination with other video-based interfaces, for example, the Camera Mouse [10], which is a computer access device that has been adopted by numerous users with severe disabilities in the US, Great Britain, and Ireland. Exercises, such as "keep the music playing" or "play a melody" could be performed with the Camera Mouse [10], which tracks body features such as the tip of the user's nose or chin. Another inferface for people with severe disabilities is the BlinkLink [11, 13], 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 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 MusicCamera 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 Moving 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 powerful 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 MusicCamera will continue in the future, and clinical studies will be performed to test its rehabilitative potential.

Acknowledgment

This material is based upon work supported by the National Science Foundation (IIS-0308213, IIS-039009, IIS-0093367, P200A01031, and EIA-0202067).

References

1. M. J. Tramo. Biology and music: Music of the hemispheres. *Science*, 291(5501):54–56, 2001.
2. A. Lahav, E. Saltzman, and G. Schlaug. Auditory mirror neurons: Action representations of sound in the human brain: an fMRI study. Manuscript in preparation.
3. A. Lahav, A. Boulanger, G. Schlaug, and E. Saltzman. The power of listening: auditory-motor interactions in musical training. *Annals of the New York Academy of Sciences*, 1060, 2006. in press.
4. C. Pacchetti, F. Mancini, R. Aglieri, C. Fundaro, E. Martignoni, and G. Nappi. Active music therapy in Parkinson's disease: an integrative method for motor and emotional rehabilitation. *Psychosomatic Medicine*, 62(3):386–393, 2000.
5. S. Paul and D. Ramsey. Music therapy in physical medicine and rehabilitation. *Australian Occupational Therapy Journal*, 47:111–118, 2000.
6. A. Lahav. Musical solution to motor dysfunctions, 2005. Invited Presentation at the Internation Conference on Enabling Environment: Music, Technology, Culture, and Healthcare, United Nations, NY.
7. A. Camurri, P. Coletta, A. Massari, B. Mazzarino, M. Peri, M. Ricchetti, A. Ricci, and G. Volpe. Toward real-time multimodal processing: Eyesweb 4.0. In *Proceedings of the AISB 2004 Convention: Motion, Emotion and Cognition*, Leeds, UK, March 2004.
8. A. Camurri, S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, and G. Volpe. EyesWeb: Toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal*, 24(1):57–69, 2000. Additional information on Eyesweb, developed in the Laboratory of Musical Informatics at the University of Genova, Italy, is available at http://www.infomus.dist.unige.it.
9. A. Camurri, B. Mazzarino, G. Volpe, P. Morasso, F. Priano, and C. Re. Application of multimedia techniques in the physical rehabilitation of Parkinson's patients. *The Journal of Visualization and Computer Animation*, 14:269–278, December 2003.
10. M. Betke, J. Gips, and P. Fleming. 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, March 2002.
11. K. Grauman, M. Betke, J. Lombardi, J. Gips, and G. R. Bradski. Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, 2(4):359–373, November 2003.
12. J. J. Magee, M. R. Scott, B. N. Waber, and M. Betke. EyeKeys: A real-time vision interface based on gaze detection from a low-grade video camera. In *IEEE Workshop on Real-Time Vision for Human-Computer Interaction (RTV4HCI)*, Washington, D.C., July 2004. IEEE Computer Society. 8 pp.
13. M. Chau and M. Betke. Real time eye tracking and blink detection with USB cameras. Technical Report 2005-012, Computer Science Department, Boston University, April 2005. http://www.cs.bu.edu/techreports/pdf/2005-012-blink-detection.pdf.
14. Open Source Computer Vision Library, Intel Corporation (July 2002). http://www.intel.com /research/mrl/research/opencv.

15. F. Sparacino, G. Davenport, and A. Pentland. Media in performance: Interactive spaces for dance, theater, circus, and museum exhibits. *IBM Systems Journal*, 39(3 & 4):479–510, 2000.
16. C. R. Wren, F. Sparacino, A. J. Azarbayejani, T. J. Darrell, T. E. Starner, A. Kotani, C. M. Chao, M. Hlavac, K. B. Russell, and A. P. Pentland. Perceptive spaces for performance and entertainment untethered interaction using computer vision and audition. *Applied Artificial Intelligence*, 11(4):267–284, June 1997.
17. BigEye, a system developed by the Studio for Electro-instrumental Music at the Center for Research & Development of Instruments & Tools, Amsterdam. http://www.steim.org/steim/bigeye.html. Accessed August 2005.
18. K. C. Ng. Music via motion: Transdomain mapping of motion and sound for interactive performances. *Proceedings of the IEEE*, 92(4):645–655, April 2004.
19. Very Nervous System, developed by Media Installation Artist David Rokeby, Toronto, Canada. http://homepage.mac.com/davidrokeby/vns.html. Accessed August 2005.
20. C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. PFINDER: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.
21. T. Winkler. Making motion musical: Gesture mapping strategies for interactive computer music. In *Proceedings of the 1995 International Computer Music Conference*, pages 261–264, San Francisco, CA, 1995. International Computer Music Association.
22. R. Morales-Manzanares, E. F. Morales, R. Dannenberg, and J. Berger. SICIB: An interactive music composition system using body movements. *Computer Music Journal*, 25(2):25–36, 2001.
23. E. Johnstone. A MIDI Foot Controller – The PodoBoard. In *Proceedings of the 1991 International Computer Music Conference*, pages 123–126, San Francisco, CA, 1991. International Computer Music Association.
24. N. Griffith and M. Fernstrom. LiteFoot – A Floor Space for Recording Dance and Controlling Media. In *Proceedings of the 1998 International Computer Music Conference*, pages 475–481, San Francisco, CA, 1998. International Computer Music Association.
25. J. A. Paradiso, K. Hsiao, A. Y. Benbasat, and Z. Teegarden. Design and implementation of expressive footwear. *IBM Systems Journal*, 39(3 & 4):511–529, 2000.
26. Kontrolldress, a body suit developed by Norsk nettverk for Teknologi, Akustikk og Musikk (NOTAM), Norway. http://www.notam02.no/notam02/prod-maskin-kontrolldress.html. Accessed August 2005.
27. C. Van Raalte. Old dog, new tricks. Biofeedback as assistive technology. In *The California State University Northridge Center On Disabilities 1998 Conference*, Los Angeles, CA, 1999. Also available online at http://www.dinf.org/csu_98/csun98_098.htm.
28. T. Marrin Nakra. Searching for meaning in gestural data: interpretive feature extraction and signal processing for affective and expressive content. In M. Wanderley and B. B. Rovan, editors, *Trends in Gestural Control of Music*. Institut de recherche et de coordination acoustique/musique (IRCAM), Paris, France, 1999.
29. MidiDancer, a wireless movement sensing system by T. Ranch, M. Coniglio, and D. Stoppielo, Brooklyn, NY. http://www.troikaranch.org/mididancer.html. Accessed August 2005.

30. W. Siegel and J. Jacobsen. The challenges of interactive dance: An overview and case study. *Computer Music Journal*, 22(4):29–43, 1998. More information is available online at http://hjem.get2net.dk/diem/gen-info.html. Accessed August 2005.
31. J.O. Borchers and M. Muhlheuser. Design patterns for interactive musical systems. *IEEE Multimedia*, 5(3):36–46, 1998.
32. B. Bongers and Y. Harris. A structured instrument design approach: The video-organ. In *Proceedings of the 2002 International Conference on New Interfaces for Musical Expression - NIME02*, pages 86–91, 2002.
33. C. Leider. A realtime MIDI library for an inexpensive microcontroller-based interactive performance system. In *Proceedings of the 1997 International Computer Music Conference*, Thessaloniki, Greece, 1997. International Computer Music Association, San Francisco.
34. Flock of Birds, a magnetic body tracker by Ascension Technology Corporation. http://www.ascension-tech.com/products/flockofbirds.php. Accessed August 2005.
35. R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):696–706, May 2002.
36. K. Schwerdt and J. L. Crowley. Robust face tracking using color. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000.
37. H. Wu, Q. Chen, and M. Yachida. Face detection from color images using a fuzzy pattern matching method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(6):557–563, June 1999.
38. R. Jain, R. Kasturi, and B. Schunk. *Machine Vision*. McGraw Hill, 1995.
39. B. K. P. Horn. *Robot Vision*. The MIT Press, Cambridge, MA, 1986.
40. N. C. Makris. A foundation for logarithmic measures of fluctuating intensity in pattern recognition. *Optics Letters*, 20:2012–2014, 1995.