Real Time System for Gesture Tracking in Psycho-motorial Rehabilitation

Massimo Magrini and Gabriele Pieri
Institute of Information Science and Technologies, National Research Council, Via Moruzzi 1, Pisa, Italy

Keywords: Gesture Recognition, Tracking System, Autism Spectrum Disorder, Rehabilitation Systems.

Abstract: In the context of the research activities of the Signal and Images Lab of ISTI-CNR, a system is under development for real-time gesture tracking, to be used in active well-being self-assessment activities and in particular applied to medical coaching and music-therapy. The system uses a video camera, a FireWire digitalization board, and a computer running own developed software. During the test sessions a person freely moves his body inside a specifically designed room. The developed algorithms can extrapolate features from the human figure, such us spatial position, arms and legs angles etc. Through the developed system the operator can link these features to sounds synthesized in real time, following a predefined schema. The system latency is very low thanks to the use of Mac OS X native libraries (CoreImage, CoreAudio). The resulting augmented interaction with the environment could help to improve the contact with reality in young subjects affected by autism spectrum disorders (ASD).

1 INTRODUCTION

In the last years sensor based interactive systems for helping the treatment of learning difficulties and disabilities in children appeared on the specialized literature (Ould Mohamed and Courbulay, 2006) and (Kozima et al. 2005). These systems, like the quite popular SoundBeam (Swingler and Price 2006), generally consist of sensors connected to a computer, programmed with special software which reacts to the sensor's data with multimedia stimuli.

The general philosophy of these systems is based on the idea that even profoundly physically or learning impaired individuals can become expressive and communicative using music and sound (Villafuerte et al., 2012). The sense of control which these systems provide can be a powerful motivator for subjects with limited interaction with reality. Our research department has got a long tradition in developing special gesture interfaces for controlling multimedia generation, even if targeted to new media art.

While systems like SoundBeam totally rely on ultrasonic sensors, our system is based mostly on real-time video processing techniques; moreover it is easily possible to use an additional set of sensors (e.g. infrared or ultrasonic). The use of video-processing techniques adds more parameters which can be used for the exact localization and details about the human gestures to be detected and recognized. By using the implemented software interface, the operator can link these extracted video features to sounds synthesized in real time, following a predefined schema.

The proposed system has been experimented as case study, in a real-patients test campaign over a set of patient affected by autism spectrum disorder (ASD), in order to provide them an increased interaction towards external environment and trying to reduce their pathological isolation (Riva et al., 2013).

Following the case study testing on young subjects, very positive results were obtained, confirmed both by the professional therapists and the parents of the patients. In particular the therapists reported a positive outcome from the assisted coaching therapies. Moreover this positive evolution could bring an improvement in terms of transferring the motivation and curiosity for the full communication interaction in the external environment, thus improving the well-being of the subjects.

Magrini M. and Pieri G..
Real Time System for Gesture Tracking in Psycho-motorial Rehabilitation.
DOI: 10.5220/0004937205630568
In *Proceedings of the International Conference on Health Informatics* (SUPERHEAL-2014), pages 563-568
ISBN: 978-989-758-010-9

2 SYSTEM METHODOLOGY

The system is installed in a special empty room, with most of the surfaces (walls, floor) covered by wood. The goal is to build a warm space which, in some way, recall the prenatal ambient. All system parts such as cables, plugs etc. are carefully hidden, as they are potential elements of distraction for autistic subjects. The ambient light is gentle and indirect, also for avoiding shadows that can affect the motion detection precision.

The whole system is based on an Apple Macintosh computer (Figure 1), running the latest version of Mac OS X. The video camera is connected to the computer thru a firewire digitizer, the Imaging Source DFG1394. This is a very fast digitizer, which allows a latency of only 1 frame in the video processing path. As an output audio card we decided to use the Macintosh internal one, its quality is superior to an average PC, more than sufficient for our purposes. A couple of TASCAM amplified loudspeaker completes the basic system. For using additional sensors (infrared, ultrasonic) we could add a simple USB board which digitize analogic control signals translating them into standard MIDI messages, easy to manage inside the application.

We used the Mac OS platform for its reliability in real time multimedia applications, thanks to its very robust frameworks: Core Audio and Core Image libraries permit very fast elaboration without glitches and underruns.

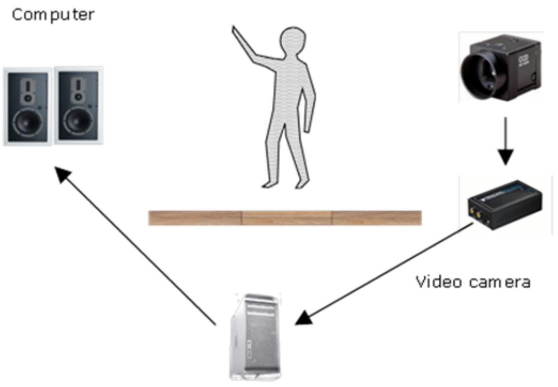


Figure 1: Structure of the System.

3 THE INTERACTIVE SYSTEM

The software is a standalone application, and it is obviously structured in different modules following a strict C++ paradigm (Figure 2). The most important modules are the Sequence grabber, which manages the stream of video frames coming from the video digitize, the Gesture tracking module, which analyses the frames and extrapolate the gesture parameters, and the Mapper, responsible of the mapping between detected gesture parameters and the generated sounds.

The bigger problem regarding the gesture control of sounds is the latency, which is the delay between the gesture and the correspondent effect on the generated sound. Commercial systems like the popular Microsoft Kinect could greatly simplify the system but introduce a latency (around 100 ms) that is not acceptable for our purposes. Our approach guarantees the minimum latency for the adopted frame rate, which is 40 ms at 25 FPS or 33.3 ms at 30 FPS.

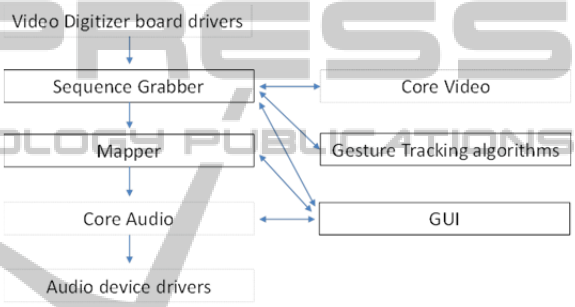
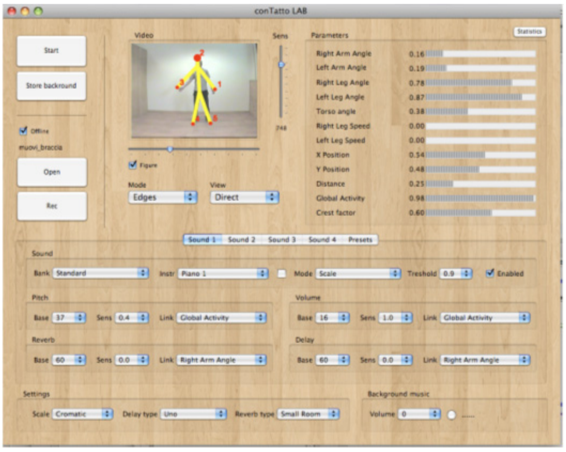


Figure 2: Software Architecture Structure.

3.1 Graphical User Interface

Following the music-therapist specifications we implemented the application graphical user interface as a single window, with subfolders for specific topics (Figure 3). In this way every aspect of the system setup is quickly accessible to the operators during the music-therapy sessions.



☑ Offline

muovi_braccia

☑ Figure

Mode: Edges

View: Direct

Parameters	
Right Arm Angle	0.16
Left Arm Angle	0.19
Right Leg Angle	0.78
Left Leg Angle	0.87
Torso angle	0.38
Right Leg Speed	0.00
Left Leg Speed	0.00
X Position	0.54
Y Position	0.48
Distance	0.25
Global Activity	0.98
Crest factor	0.60

Sound: Bank Standard; Instr Piano 1; □ Mode Scale; Treshold 0.9; ☑ Enabled

Pitch: Base 37; Sens 0.4; Link Global Activity

Volume: Base 16; Sens 1.0; Link Global Activity

Reverb: Base 60; Sens 0.0; Link Right Arm Angle

Delay: Base 60; Sens 0.0; Link Right Arm Angle

Settings: Scale Cromatic; Delay type Uno; Reverb type Small Room

Background music: Volume 0 ○ ......

Figure 3: The appearance of the Graphical Interface.

The upper area of the GUI contains the video preview and the detected parameters monitor (see Section 3.3 for details on the parameters), while the lower one permits to setup the mapping between parameters and generated sounds.

3.2 Image Processing

The incoming frame grabbed from the digitizer is processed in several steps, in two alternative modalities: area based or edge based (see Figure 4 top). In the first modality the segmentation process is made on the full area areas, while the edge based one it is based on the edge present in the grabbed frame.

In both modes the image is firstly it is smoothed with a Gaussian filter (fast computed thanks to the *Coreimage* library). In the edge mode the image is processed with an edge detection filter, too.

Then, we use a background subtraction technique for isolate the human figure from the ambient. Pressing the "Store background" button (obviously with no human subjects in front of the camera) we can store the background, area or edge based. When the figure is present in front of the camera the incoming frames are compared with the stored background, using a dynamic threshold, obtaining a binary matrix. The average threshold used in this operation can be tuned by the operator using a simple slider. It is not necessary to set again this sensitivity if the ambient light does not change.

Finally we apply an algorithm for removing unconnected small areas from the matrix, usually generated by image noise. The final binary image is then ready to be processed by the gesture tracking algorithm.

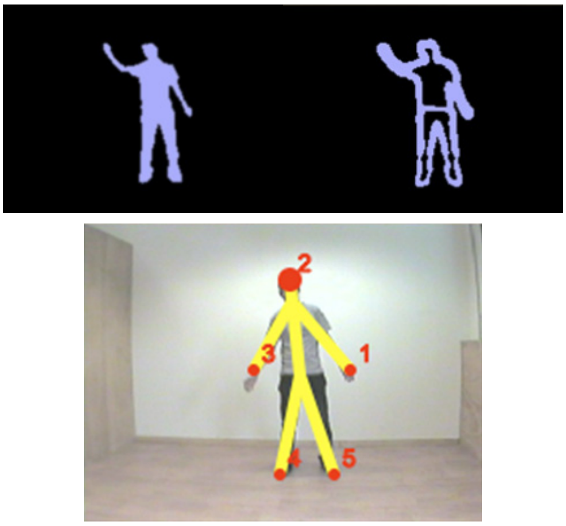


Figure 4: Example of Image elaboration and its rendering.

The whole algorithm, executed for each incoming frame can be described with this pseudo code:

```
START: F0 = <grabbed frame>
F1 = GRAY_IMAGE(F0);
F2 = BLUR_FILTER(F1)
if (AREA_MODE)
     F3 = F2
if (EDGE_MODE)
     F3 = EDGE_FILTER(F2);
if (STORE_AS_BACKGROUND)
     B = F3
F4 = F3-B
F5 = BINARIZE(F4)
F6 = ERODE(F5)

TRACK_MOVEMENT(F6)
goto START;
```

The frame resolution is 320x240, and a full frame rate (e.g. 25 FPS) can be achieved because all the image filters are executed by the GPU.

3.3 Gesture Tracking Algorithm

Starting from the binary raster matrix we apply an algorithm to detect a set of gesture parameters. This heuristic algorithm supposes that the segmented image obtained by the imagine elaboration process is a human figure, and tries to extrapolates some features from it. This process is based on a simplified model of the human figure (see Figure 4 bottom). Additional model for single parts of the body (face, hands) are under development and it can be used for more "zoomed" version of the system. At the moment we can rapidly detect the position of the head, the arms, the legs, and its evolution over time. Starting from these five time dependent positions we decided to compute the following parameters:

- Right Arm angle
- Left Arm angle
- Right Leg angle
- Left Leg angle
- Torso angle
- Right Leg speed
- Left Leg speed
- Barycentre X
- Barycentre Y
- Distance of subject form camera

Their names are self-explaining. The distance from the camera actually is an index related to the real distance: it is computed as the ratio between the frame height and the detected figure maximum height. The leg speed is computed analyzing the last couple of received frames; it is useful for triggering sounds with "kick-like" movements. We also compute these two additional parameters:

- Global activity
- Crest factor

The first is an indicator of overall quantity of movement (0.0 if the subject is standing still with no moments), while the second one is an indication of the concavity of the posture: (0.0 means that the subject is standing with the legs and the arms are united with the body). Some optimizations are performed in order to start the frame analysis from an area centred in the last detected barycentre. Instead of aiming at the design of a very sophisticated detection algorithm, we tried to implement it in a very optimized way for maintaining the target frame rate (25 FPS), in order to avoid latency between gestures and sounds.

3.4 Sound Generation

The sound generation is based on the Mac OS CoreAudio library. We used the Audio Unit API for building an Audio graph: 4 instances of DownLoadable Synthesizer (DLS) are mixed together in the final musical signal. These synthesizers produce sounds according to standard MIDI messages received from their virtual input ports. We added two digital effects (echo and Reverberation) to the final mix: for each synthesizer we can control the portion if its signal to be sent to these effects.

Each synthesizer module can load a bank of sounds (in the DLS or SF2 standard format) from the set installed in the system. The user can obviously add his own sound banks, including the sounds he created, to the system. It is also possible to specify a background audio file, to be played together with the controlled sounds.

3.5 Mapper

The mapper module translates the detected features into MIDI commands for the musical synthesizers. Each synthesizer works in independent way, and for each of them it is possible to select the instrument and the instrument banks.

Each parameter of the sounds (pitch, volume, etc.) can be easily linked to the detected gesture parameters using the GUI. For example we can link the Global Activity to the pitch: the faster you move the more high pitched notes you play. The synthesized MIDI notes are chosen from a user selectable scale: there's a large variety of them, ranging from the simplest ones (e.g. major and minor) to the more exotic ones. As an alternative, it is possible to select continuous pitch, instead of discrete notes: in this way the linked detected features controls the pitch in a "glissando" way.

Sound can be triggered in a "Drum mode" way, too: the MIDI note *C* played when the linked parameters reach a selected threshold.

All these links settings can be stored in pre-sets, easily recallable and from the operators.

3.6 Parameters Summary

The detected parameters are shown in real-time with a set of horizontal bars. Their shown value is normalized between 0 and 1; in this way we found that it is easy to understand their role in a link.

At the end of a session it is possible, pressing the Statistics button, to show a simple Statistic of the gesture parameters (currently the average and the variance). These data, together with some other useful information can be saved in a text file for further external analysis.

4 RESULTS AND DEVELOPMENTS

The installed system has been applied for case study tests on several young patients affected by Autism Spectrum Disorder.

Autism is a brain development disorder characterized by impaired social interaction and communication. It appears in the first years of life and arrests the development of affective evolution. It basically compromises social interaction and language expression, and often leads to restricted and repetitive behaviour. Various studies reports about the incidence of autism, they all confirm a large increase in the recent years, rising from 2002 with an increase rate of 78%. In particular on average, but depending on the age of the data retrieved autism affects about 1 children out of 100 (Baio, 2012) in the peak age (8 years old). The following Figure 5 report the rapid increase and trend (Chiarotti and Venerosi Pesciolini, 2012).

Autism has a genetic basis, but a complete explanation of its causes is still unknown. An exhaustive description of this disorder in medical terms is beyond the scope of this document.

Studies have shown that music-therapy has a significant, positive influence when used to treat autistic individuals (Alvin and Warwick, 1992).

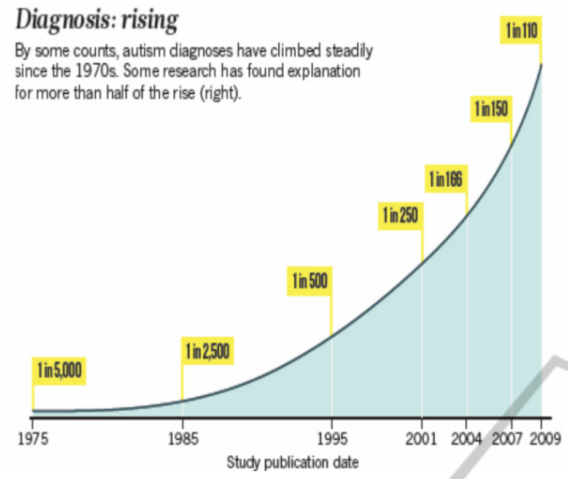


Figure 5: Autism diagnosis rising.

Participating in music therapy allows autistics the opportunity to experience non-threatening outside stimulation, as they do not engage in direct human contact. Music is a more universal language respect to oral language, and it allows a more instinctive form of communication.

The classic music-therapy is mostly based on listening sounds and music, together with a therapist (passive music-therapy). Our system, instead, is active and interactive: providing an augmented interaction with the environment it tries to remove the subject from its pathological isolation.

The experimentation of the system with real cases was performed during last year and half over 10 patients.

The therapists reported positive outcome confirmed also by the parents of the young patients involved. In particular there is a continuous and positive trend, evolving from the first meetings till the latest. One of the most evident positive changes is the appearance of sight contact with the therapist coupled with smiling and vocalization to imitate the sounds or the voice of the therapist. In some case will to verbal communication appeared in some subjects that were lacking.

Moreover there has been also tentative of imitative actions and own free initiatives from the patients upon inputs from the therapists.

Particular relevance of the positive outcome of the therapies, among others, can be given by the following evolution signs:

- Some semblance of game with the operator, signs of understanding with his head to invite the operator to repeat certain gestures;
- Imitative action and / or free initiative by the children on operator input;
- Interest to the acoustic stimuli, with attempts to adapt the movements and the gaming action

In general the experimentation is confirmed to be particularly promising for a very important and challenging goal: verify the conservation of the improvements obtained within the case study setting also in the external environment, transferring the motivation and curiosity for the full communication interaction in the real world.

Regarding the future development of the system, there is an active plan to add a database to the control software in order to store the patient's data, including all parameters statistics for the music therapy sessions. In this way the therapists would like to investigate the relationship between the patient's gesture evolution and his autistic disorders. We are developing new models for the gesture recognition module, specialized for single parts of the body. For example, we would like to give the possibility to concentrate the video camera only on the patient's face, detecting movements and positions of eyes and mouth.

A 3D version of the system is also under study, in this way it will be not necessary to stand in front of the camera, but it could be possible to rotate around the body axis, still capturing the correct arms and legs angles.

5 CONCLUSIONS

We described an interactive, computer based system based on real-time image processing, which reacts to movements of a human body playing sounds. The mapping between body motion and produced sounds is easily customizable with a software interface. This system has been used for testing an innovative music-therapy technique for treating autistic children. The experimentation performed of the system with real cases was performed during last period confirmed several benefits from the application of the proposed system. These have been confirmed both by the therapists and the parents of the young patients. The most interesting outcome of the testing was the improvement obtained which could lead also to promising transfer of the attitudes displayed in the test case study to the external environment, in particular referring to the motivation and curiosity for the full communication interaction in the real world.

Moreover, during the case study testing, it was found that along with treating autism spectrum disorder, the system could be successfully used for other diseases, such as Alzheimer and other pathologies typical of older people.

ACKNOWLEDGEMENTS

The activity which brought to the implemented system was partially sponsored by the Foundation Cassa di Risparmio di Lucca. The authors would like to thank Dr. Elisa Rossi for the precious contribution given in the test case, and in the evaluation of results.

REFERENCES

Baio, J., 2012. Prevalence of Autism Spectrum Disorders – Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. In *Morbidity and Mortality Weekly Report (MMWR) Surveillance Summaries*, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, March 30, 2012, pp. 1-19, n. 61 (SS03).

Chiarotti, F., Venerosi Pesciolini, A., 2012. Epidemiologia dell'autismo: un'analisi critica. In *Congresso Nazionale e Workshop formativi – Autismo e percorsi di vita: il ruolo della rete nei servizi*, Azienda ASL di Ravenna, October 4-6, 2012, Ravenna, Italy.

Kozima, H., Nakagawa, C., Yasuda, Y., 2005. Interactive robots for communication-care: a case-study in autism therapy. In *International IEEE Workshop on Robot and Human Interactive Communication*.

Ould Mohamed, A., Courbulay, V., 2006. Attention analysis in interactive software for children with autism. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. Portland, Oregon, USA.

Riva, D., Bulgheroni, S., Zappella, M., 2013. *Neurobiology, Diagnosis & Treatment in Autism: An Update*, John Libbey Eurotext.

Swingler, T., Price, A., 2006. *The Soundbeam Project*.

Alvin, J., Warwick, A., 1992. *Music therapy for the autistic child*, Oxford University Press, USA.

Villafuerte, L., Markova, M., Jorda, S., 2012. Acquisition of social abilities through musical tangible user interface: children with autism spectrum condition and the reactable. In *Proceedings of CHI EA '12-CHI '12 Extended Abstracts on Human Factors in Computing Systems*, pp. 745-760, ACM New York, NY, USA.