An Auditory Feedback based system for treating Autism Spectrum Disorder

Massimo Magrini
ISTI-CNR
Via Moruzzi, 1
56124 Pisa - ITALY
+39 050 315 3144
massimo.magrini@isti.cnr.it

Andrea Carboni
ISTI-CNR
Via Moruzzi, 1
56124 Pisa - ITALY
+39 050 315 3144
andrea.carboni@isti.cnr.it

Ovidio Salvetti
ISTI-CNR
Via Moruzzi, 1
56124 Pisa - ITALY
+39 050 315 3124
ovidio.salvetti@isti.cnr.it

Olivia Curzio
IFC-CNR
Via Moruzzi, 1
56124 Pisa - ITALY
+39 050 315 3144
oliviac@ifc.cnr.it

ABSTRACT

A system for real-time gesture tracking is presented, used in active well-being self-assessment activities and in particular applied to medical coaching and music-therapy. The system is composed of a video camera, a FireWire digitalization board, and a computer running own (custom) developed software. During thetest sessions, a person freely moves his body inside a specifically designed room. The algorithms detect and extrapolate features from the human figure, such us spatial position, arms and legs angles, etc. An operator can link these features to sounds synthesized in real time, following a predefined schema. The augmented interaction with the environment helps to improve the contact with reality in subjects having some disability. The systemhas been tested on a set of young subjects affected by autism spectrum disorders (ASD) and a team of psychologists has analyzed the results of this experimentation.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications

J.3 [**Life and Medical Science**]: Health

General Terms

Measurement, Experimentation.

Keywords

Human-computer interaction, Biofeedback, Autism.

1. INTRODUCTION

In the last years specific activity has been carried out for developing sensor-based interactive systems capable to help the treatment of learning difficulties and disabilities in children [1][2]. These systems generally consist of sensors connected to a computer, programmed with special software that reacts to the sensor data with multimedia stimuli.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

REHAB '15, October 01-02, 2015, Lisbon, Portugal

© 2015 ACM. ISBN 978-1-4503-3898-1/15/10…$15.00

DOI: 10.1145/2838944.2838952

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 [3]. The sense of control, which these systems provide, can be a powerful motivator for subjects with limited interaction with reality.

While a great part of systems, like SoundBeam (www.soundbeam.co.uk), totally rely on ultrasonic sensors, our approach is mostly based on real-time video processing techniques; moreover, our solution makes also it possible to easily use additional sets of sensors (e.g., infrared or ultrasonic) in the same scene. The use of video-processing techniques adds more parameters suitable to localize exactly and detail all the human gestures we want detect and recognize. By using a custom software interface, the operator can link the extracted video features to sounds synthesized in real time, following a predefinedschema.

The developed system has been experimented in a test campaign on a set of young patients affected by Low-Functioning Autism (autism spectrum disorder, ASD), in order to provide a personal increased interaction over the operational environment and to reduce pathological isolation [4]. Results were very positive and encouraging, as confirmed by both clinical psychologist and parents of the kids. In particular, the therapists reported a positive outcome from the assisted coaching therapies. Indeed, this positive evolution was crucial to improve the motivation and curiosity for a full communication interaction in the external environment, thus affecting subjects' well-being.

2. AUTISM

ASD is a neurodevelopmental disorder characterized by impaired social interaction and communication. It is a pervasive developmental disorder, characterized by a triad of impairments: social communication problems, difficulties with reciprocal social interactions, and unusual patterns of repetitive behavior [5]. Leo Kanner, a child psychiatrist [6], described it for the first time in 1943. An exhaustive description of this disorder in medical terms is beyond the scope of this paper.

Unfortunately, no medications can cure autism or treat its core symptoms, but rather can help some people affected feel better. A large part of the interventions focuses on behavioral approaches, of which the best known is the ABA (Applied Behavior Analysis) method [7], based on repetitive patterns and reinforcements. Other approaches follow instead the

Developmental Individual Difference Relationship (DIR) model [8]. DIR acts at various levels of involvement, attempting a containing action against the central symptoms of autism according to the following guidelines: 1) involvement against isolation 2) communication and flexibility versus rigidity and persistence 3) gestures against stereotypies and aggressive behaviors. In the design of our system we were inspired, even not strictly, by the DIR model.

3. SYSTEM

The system developed is based on an Apple Macintosh computer running the latest version of Mac OS X. A video camera is connected through a Firewire digitizer, the Imaging Source DFG1394, a very fast digitizer that allows a latency of only one frame in the video processing path. As output audio card we use the Macintosh internal one, sufficient for our purposes. A couple of TASCAM amplified loudspeakers completes the basic system.

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.

Finally, 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, recalls the prenatal ambient. All system parts, such as cables, plugs, and so on are carefully hidden as they could be potential elements of distraction. The ambient light is gentle and indirect, thus avoiding shadows that could also affect the precision of motion detection.

3.1 Software architecture

The implemented software (Figure 1) is a standalone application, organized in several specialized coordinated modules. The most important are (i) the *Sequence grabber*, which manages the stream of video frames coming from the video digitizer, (ii)the *Gesture tracking module*, which analyses the frames and extrapolates gesture parameters, and (iii) the *Mapper*, responsible for transforming the detected gesture parameters into sounds.

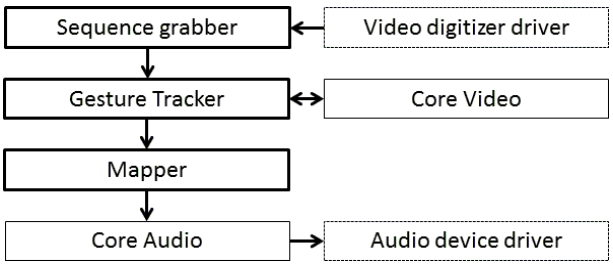

Figure 1. Software architecture.

The biggest problem regarding the gesture control of sounds is latency, which is the delay between the gesture and the correspondent effect on the generated sound. Commercial products, like for instance the popular Microsoft Kinect, could greatly simplify the system but introduce a latency (around 100 ms in ver.1) 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.

3.1.1 Processing algorithm

The whole algorithm, executed for each incoming frame is depicted by the diagram in Fig.3. In the first step of the algorithm each grayscale frame grabbed in real time from the video camera is smoothed with a Gaussian filter (fast computed using the CoreImage library). The output is then processed in one of two alternative operating modes: *area-based* or *edge-based* (Fig.2).

In the area-based modality, the segmentation is performed considering the entire envelope (area) of the figure of the subject examined.

Figure 2. Area (left) and edge (right) operating modes.

In the edge-based modality, instead, an edge detection filter is applied to the image. The next step consists in a background subtraction technique computed to isolate the human figure from the ambient. In order to fulfill this task, each time the background changes, it has to be stored, area or edge based, with no human subject in front of the camera. If this background exists it is used in the following iterations of the algorithm and compared with the incoming frames containing the human figure using a dynamic threshold, obtaining a binary matrix. The average threshold used in this operation can be tuned by the operator in real time. It is not necessary to set again this sensitivity if the ambient light does not change.

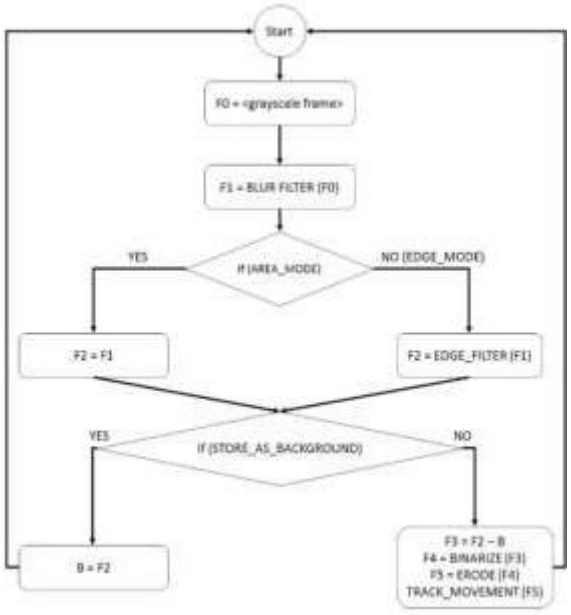

Figure 3. Processing algorithm diagram

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.

The frame resolution is 320x240 pixels, full frame rate (e.g. 25 FPS) can be achieved because all the image filters are executed by the GPU. 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 image elaboration process is a human figure, extracting data from it. This process starts searching a simplified model of the human figure (Fig. 4).

Figure 4. Simplified basic human model

Additional models for detailed parts of the body (face, hands) are currently under development, so that they can be used for more "zoomed" versions of the system. Starting from time dependent position of detected body joints 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 , Barycenter X, Barycenter Y, Distance of subject form camera.

Their names are self-explaining. The distance from the camera actually is just an index related to the real distance: it is simply computed as a 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 in standing position with arms kept along the body). Few optimizations are performed starting the frame analysis from an area centered in the last detected barycenter. Generally speaking we tried to implement the detection algorithms in a very optimized way in order to maintain the target frame rate (25 FPS), minimizing the latency between gestures and sounds.

3.1.2 Sound generation

The sound generation is based on the Mac OS CoreAudio library. We used the Audio Unit API for building a so called *Audio graph*. Four 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 of 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 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.

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 from different 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: thefaster 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 presets, easily selectable from the operators.

4. EXPERIMENTATION

The experimentation [9] was performed in 2011-2012 in the school environment on 4 subjects (5-7 years, all males) diagnosed with Low-Functioning Autism (autism spectrum disorder, ASD). The weekly intervention lasted about 30 minutes. The children involved in the experimentation were evaluated in a cross sectional and follow up pilot study. Clinical features evaluated by mental health centers and information from the "Questionnaire on motor control and sensory elaboration" [10], compiled by parents, and from the "Short Sensory Profile" [11] filled up by the teacherswere analyzed at baseline. Three clinical psychologists, not previously involved in the experimentation, analyzed the first eight videos of the intervention, completing an observation grid for every session. The grid was structured ad hoc by the research team on the basis of the DIR Floortime model and technique in relation to the benchmarks of the main sensory profiles. This grid was partially taken from the questionnaire of Politi and colleagues
[12] aimed at assessing the sensitivity of music in children affected by autism spectrum disorder. The instrument is made up of nineteen items relating to the child's behavior during the sessions and measured the characteristics of each sensory profile in terms of the "four A": *Arousal, Attention, Affection and Action* [13]. In this experimentation to consider the sensory profile of the infant that undergoes the sound stimulation and interaction was crucial. The choice of sound stimuli related to the movement has been made individually for each child during the first sessions through broad-spectrum stimuli. All the interventions were calibrated on the basis of the observations drawn from the videoof the previous meeting, viewed by the reference clinician, a child neuropsychiatrist. It is important to highlight that the clinical reference as well as the operator who conducted the interventions with children had formal training in DIR Floortime method.

4.1 RESULTS

The concordance rate between the three psychologists behavioral observation grid was calculated with the interclass correlation coefficient. Moderate to good inter-rater agreement [intraclass correlation coefficient (ICC) comprised between 0.596 (95%CI 0.41-0.853) and 0.799 (95%CI 0.489-0.933)] were found. A repeated measures design was performed to evaluate change over

time for each child for the first eight sessions (T1-T8). The analysis of variance was performed to assess if there has been an improvement in specific symptomatic areas. The repeated measures analysis of variance indicated an overall increase of the scores drawn up by psychologists (T1-T8; $p<0.05$) (Figure 5). Concerning statistical indexes our study highlights that participants had improved several skills. These variations in behavioral expressions reflect a relational evolution indicating the beginning of an opening attempt to someone no longer perceived as a threat but as someone from which to draw contentment, through playful interaction with the sounds. This pilot study suggests effectiveness of the treatment of autistic children using our approach. The entire project addressed many challenges. Each autistic child is unique in the sense that improvements in abilities are very subjective and some study limitations have to be mentioned: first of all the small size and the non-homogeneity of the sample; this is due to the difficulty of enrolling Low- Functioning Autism children with similar profiles. Participation is self-selected and sample bias cannot be excluded. A more rigorous assessment and selection of participants enrolled by medical staff with more homogeneous competences and profiles and the selection of a matched control group will guarantee results of higher value. Moreover progress trends are very subjective and could also depend upon external factors such as family involvement and health conditions; it would be important also to take into account what kind of treatment the child and/or family are undergoing in health structures. These data would be extremely interesting for creating the bases for using accessible technology-enhanced environments.

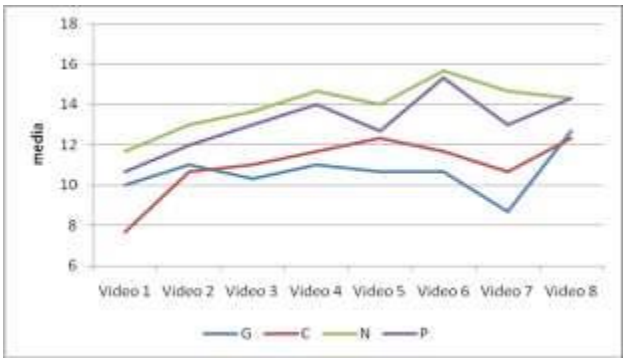

Figure 5. Mean total scores of the behavior observation grid to assess change over time for each child (T1-T8).

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 graphical user interface. This system has been used for testing an innovative music-therapy technique for treating autistic children. The experimentation with real cases demonstrated several benefits from the application of the proposed system. These have been confirmed both by the team of clinical psychologists (using a validation protocol) and by the parents of the young patients. The most interesting outcome of the experimentation was the relational improvement. This promises to transfer the behavior shown in the setting to the external environment, increasing communication and interaction in the real world. We are continuing our experiments combining our technology with Microsoft Kinect v.2. This device is interesting but still not perfect, overall latency is now around 50ms (against 30ms we currently have) at 30fps. Depending on occlusions and other issues skeleton tracking can get unreliable (e.g. when the subjectis very close to a wall). Our future approach will combine the Kinect's proprietary technology with the image processing techniques previously described in this paper.

6. ACKNOWLEDGMENTS

The authors would like to thank Dr. Elisa Rossi for the precious contribution given during the experimentation of system. This work has been partially supported by Project SI DO RE MI, funded by Fondazione Telecom Italia.

7. REFERENCES

[1] 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.

[2] 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.

[3] 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

[4] Riva, D., Bulgheroni, S., Zappella, M., 2013. Neurobiology, Diagnosis & Treatment in Autism: An Update, John Libbey Eurotex

[5] Wing L & Gould J (1979). Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. J Autism Dev Disord 9, 11–29.

[6] A. Vismara, L. J. Rogers, S. 2010. Behavioral Treatments in Autism Spectrum Disorder: What Do We Know? Annual Review of Clinical Psychology, Vol. 6, pp 447-468.

[7] Kanner L (1943). Autistic disturbances of affective contact. Nervous Child 2, 217–250.

[8] Greenspan S., Wieder S., The child with special needs, Perseus Pub. New York 1998.

[9] Massimo Magrini et.al. Progetto "SI RE MI" Sistema di Rieducazione Espressiva del Movimento e dell'Interazione.. Autismo e disturbi dello sviluppo, Erickson (2015) (In press)

[10] De Gangi G. e Berck R. (1983), DeGangi-Berck test of sensory integration. Los Angeles: Western Psychological Services.

[11] Dunn W. (2006), Sensory Profile-School Companion manual. San Antonio, TX: Psychological Corporation.

[12] Politi P., Emanuele E. e Grassi M. (2012), The Invisible Orchestra Project. Development of the "Playing-in-Touch" (PiT) questionnaire, Neuroendocrinology Letter, vol. 33(5): 552-558

[13] Meini C., Guiot G., Maria Teresa Sindelar M.T. (2012) Autismo e musica. Il modello Floortime nei disturbi della comunicazione e della relazione, Erickson, 2012

View publication stats