MUSICAL ROBOTS FOR CHILDREN WITH ASD USING A CLIENT-SERVER ARCHITECTURE

Ruimin Zhang

Michigan Technological University
Houghton, MI 49931
ruiminz@mtu.edu

Jaclyn Barnes

Michigan Technological University
Houghton, MI 49931
jaclynb@mtu.edu

Joseph Ryan

Michigan Technological University
Houghton, MI 49931
jdryan@mtu.edu

Myounghoon Jeon

Michigan Technological University
Houghton, MI 49931
mjeon@mtu.edu

Chung Hyuk Park

George Washington University
Washington, DC 20052
chpark@gwu.edu

Ayanna M. Howard

Georgia Institute of Technology
Atlanta, GA 30332
ayanna.howard@ece.gatech.edu

ABSTRACT

People with Autistic Spectrum Disorders (ASD) are known to have difficulty recognizing and expressing emotions, which affects their social integration. Leveraging the recent advances in interactive robot and music therapy approaches, and integrating both, we have designed musical robots that can facilitate social and emotional interactions of children with ASD. Robots communicate with children with ASD while detecting their emotional states and physical activities and then, make real-time sonification based on the interaction data. Given that we envision the use of multiple robots with children, we have adopted a client-server architecture. Each robot and sensing device plays a role as a terminal, while the sonification server processes all the data and generates harmonized sonification. After describing our goals for the use of sonification, we detail the system architecture and on-going research scenarios. We believe that the present paper offers a new perspective on the sonification application for assistive technologies.

1. INTRODUCTION

One of the most important characteristics of people with Autism Spectrum Disorders (ASD) comes from difficulties in social communication. ASD may not be diagnosed until adolescence or later, whereas the symptoms appear in early childhood. Children with ASD often have trouble with fundamental social skills, such as turn taking, eye contact, and joint attention. Particularly, their problems in emotional communication are a serious barrier to their social inclusion. Given that the use of interactive robots is known to have high potential for enhancing social skills of children with ASD [1-5] and that music and sound are known to be effective to induce, deliver, and regulate emotions [6], we aim to design robotic sonification for children with ASD. While there has been active research on voice communication between children with ASD and robots [7-8], little empirical research has focused on sonification robots. Sonification is defined as the use of non-speech audio signals to convey information [9]. By definition, sonification can include all types of non-speech sounds, such as musical sounds, natural sounds, sound effects, even noise, etc. With all these sounds in mind, we have investigated the effectiveness of robotic sonification for children with ASD in a more systematic way. We envision not only interaction between a child and a robot, but also interactions between multiple children and multiple robots. To this end, we have developed a client-server architecture. Each robot communicating with a child acts as a client and a sonification server processes all the data and returns the action command to the clients. Our robotic sonification system is expected to facilitate emotional and social communications of children with ASD using music and real-time sonification, generated based on a child's emotional states and physical interaction patterns with robots.

2. APPROACH TO ASD INTERVENTION

While waiting for more effective treatment for ASD, many remedies with diverse disciplines have been applied, including pharmaceutical research, behavioral therapy, and complementary and alternative medicine [10-11]. Here, we focus on an interactive robot approach and music therapy approach.

2.1. Interactive Robot Approach

Some people with autism prefer to communicate with and through computers because they are predictable and place some control on the otherwise chaotic social world [12-13]. While keeping the advantages of using computers, researchers have attempted to use robots for children with ASD to help social interaction skills [1-5]. In addition to their basic computing ability, interactive robots provide a sensing and detecting environment, and save interaction data. Since robots play a role as a toy, we can also expect that psychological "rapport" between children and robots may be well formed.

From the previous robot approach, we can learn metrics they successfully used. For example, research has shown that an interactive robotic dog (AIBO) enhances social interaction more than a mechanical toy dog (Kasha), which has no ability to detect or respond to its physical or social environment [4]. Researchers found that children with autism (aged 5-8 years) spoke more words to the AIBO, in comparison to the Kasha. AIBO also elicited more social behaviors such as verbal engagement (e.g., salutations, valedictions, and general conversation), reciprocal interaction (e.g., motioning with arms, hands, or fingers to give direction to the artifact), and authentic interaction (e.g., the speed, tone, and volume of the child's voice is exceptionally well modulated for the circumstances, or the child's body is in a state of repose oriented toward the artifact as a social partner). Moreover, as the children interacted more with AIBO, they engaged in fewer autistic behaviors such as rocking back and forth and flicking fingers or hands. Scassellati and his colleagues [8] have been working on affective prosody production learning of children with autism using Pleo, a dinosaur robot. Their pilot study was conducted with 9 to 13-year-old children with High Functioning Autism. The children had to use an encouraging tone of voice to have the robot move across areas of "water", which they are told the robot fears. Children in these trials have shown increased appropriate prosodic production with the robot trials as opposed to equal sessions with a human instructor.

2.2. Music Therapy Approach

Children with ASD often show enhanced musical ability [14], though we acknowledge that it is not always the case. Music therapy is a broad category that encompasses diverse techniques that are frequently mixed and combined including receptive, recreative, compositional, improvisational, and musical activity therapies [15]. This variety demonstrates that drawing a general conclusion about the effectiveness of music therapy is not simple, but meta-analysis of existing studies has shown positive effects of music therapy [14-15]. We expect the use of various types of sonification (including music) in interactive robotic sessions will result in as effective outcomes as the traditional music therapy sessions or better outcomes than those.

3. GOALS OF ROBOTIC SONIFICATION

A human brain has four different anatomical functions of emotions, which include appraisal, reactivity, emotional understanding, and regulation [16]. We believe our robotic sonification can enhance all of these emotional aspects of children with ASD.

3.1. Enhancing Appraisal

Individuals with ASD typically show limited engagement in expressing their own emotional states. This might be because they have some issues in appraisal of their states and/or they are not good at expressing their states actively. Serving as external stimuli, emotional music and sound will help children with ASD appraise their emotional states. Thus, our robotic sonification system is expected to boost children's ability to appraise their own emotional states.

3.2. Enhancing Reactivity/ Expression

Once they are more capable to assess their emotional states, we can expect they are likely to express their emotions. However, appraisal and reactivity might be an independent process. Once our emotion detection system can estimate the children's emotional states, the sonification system can easily help them express their emotions because music and sound deliver mood and emotions very well. It could be done in a conscious way (e.g., happy music or laughing sounds for the happy state), but it could also be done in an ambient way (e.g., slowly reflecting their mind state using the soundscape).

3.3. Enhancing Emotional Understanding

We anticipate that our robotic sonification system will also serve as a positive reinforcement of the mapping between social situations (e.g., happy or disappointed) and the sounds (e.g., laughter or a sigh) in the session. Therefore, children with ASD can make more strong associations between the situations and appropriate situational sounds. It can further help them learn and estimate the robot's (or other person's) emotional states, which might lead to forming a general capability of empathizing others. To this end, we have designed over 600 non-speech sound cues with about 30 emotional states [17]. Our plan is to construct personalized sound-emotion mapping dimensions based on subsequent mapping experiments.

3.4. Enhancing Emotion Regulation

Children with ASD have a variety of symptoms (e.g., tantrum vs. social isolation). Given that music is frequently used as an intervention for emotion regulation, if appropriately used, sonification might improve some children's emotion regulation. However, for other children with ASD, specific stimulus can make their symptoms more severe. This is why we need to create an individualized sensory stimuli map.

3.5. Enhancing Monitoring of Interaction

From the perspective of a family member, teacher, or clinician, adding sonification can provide an additional way of monitoring of children's interaction in the session. If the sonification is mapped so that the children are uniquely identifiable, they can monitor each child's emotional states. Additionally, the

sonification can provide a representation of the interaction between robots and children overall, which will be helpful assessing the entire class or therapy session.

4. SYSTEM OVERVIEW

To obtain these therapeutic goals, we have designed robotic sonification platforms. We use a client-server architecture in our system (Figures 2 and 4) that can be connected with any robots and other sensing devices via TCP/IP sockets. The first prototype system can detect children's emotional states during the interaction and play sonification adaptive to their emotions. The second system assumes that a child can interact with the multiple robots at the same time. All the activities and relationships between the child (or children) and multiple robots can be integrated in the sonification server and reflected in the sonification pattern. In this procedure, the sonification server can process different data, and control and harmonize different layers of the sonification in real-time. Depending on the purpose of the sonification, the server can decide which robot will generate the sounds or the server can also generate sonification (e.g., providing feedback to children vs. monitoring the emotional state of the child).

4.1. Prototype System Configuration
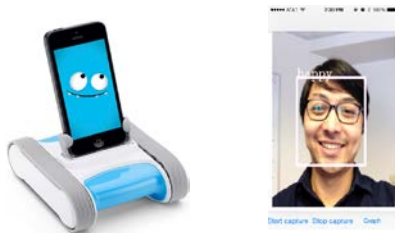
4.1.1. Non-humanoid Robot as a Client

Figure 1. Romo and facial expression detection app.
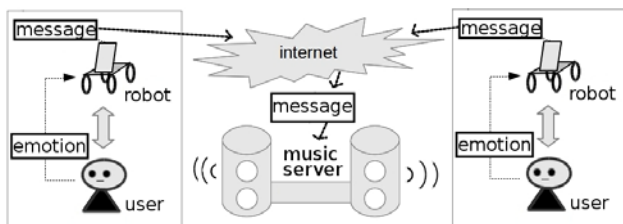

Figure 2. A client-server structure with Romo.

Our first prototype system uses Romo (Figure 1), which is composed of an iPhone and a trackmounted base. Romo can detect the child's emotions at background and generate sonification on the fly in the robot or the data can be sent to the sonification server through the Internet and the server can generate sonification. The emotion detection system was implemented through an iOS application. This application activates the camera included in the iPhone, which captures a video in real-time at a frequency of 15 frames per second, when a child interacts with Romo. When an emotional state is detected, a pink rectangle marks the relevant region in the video and a text label of the emotion is displayed on the left top corner of the rectangle (Figure 1). The detected emotions are recorded with the time they are captured. They can be encoded as messages sent to the server or rendered as a visual graph when the interaction finishes. We used the cascade training method provided by OpenCV [18] (an open-source library that focuses on real-time image processing) to train our customized classifiers for basic emotional states, including happy, surprised, and angry. The training images were from Cohn-Kanade AU-Coded Facial Expression database v2, which includes 592 sequences from 123 posers. Each sequence begins with a neutral expression and proceeds to a peak expression. When the child plays with Romo, our classifiers scan every frame of the video captured by the iPhone camera and send the data to the server.

4.2.2. Sonification Server

The sonification server was implemented with JFugue, which is an open-source Java API for programming music (e.g., MusicString, MIDI, audio file I/O, etc.). It allows us to specify notes, chords, musical instruments, and tracks to play a music piece based on our own algorithms. JFugue provides a class called `Pattern`. We used `Pattern` to create a block of sound/music (e.g., four or eight measures) for each interaction unit (e.g., an emotional state, physical activity, etc.), which we are interested in. The sonification server has an unlimited loop to listen to messages from the clients until it is manually stopped. It processes every message it has received and plays the corresponding sonification piece.

4.2. Current System Configuration

4.2.1. Humanoid Robots as Clients

While Romo was developed as the first prototype system, our long-term research plan involves using various robot platforms, including humanoid robots, non-humanoid robots, and animal robots, each of which requires different hardware and software. Again, this is why we use a client-server structure. We have been working with two robot platforms: ROBOTIS OP2 (formerly known as "DARWIN 2" or "DARWIN-OP2") and Nao (Figure 3), but here we focus on the description of ROBOTIS OP2.
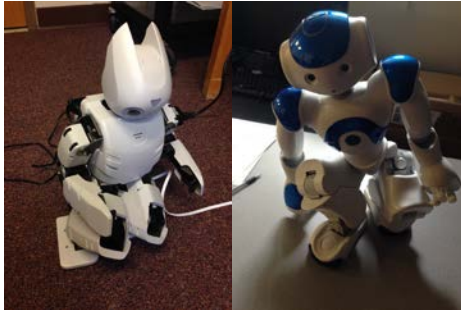
Figure 3. ROBOTIS OP2 and Nao robots for this project.
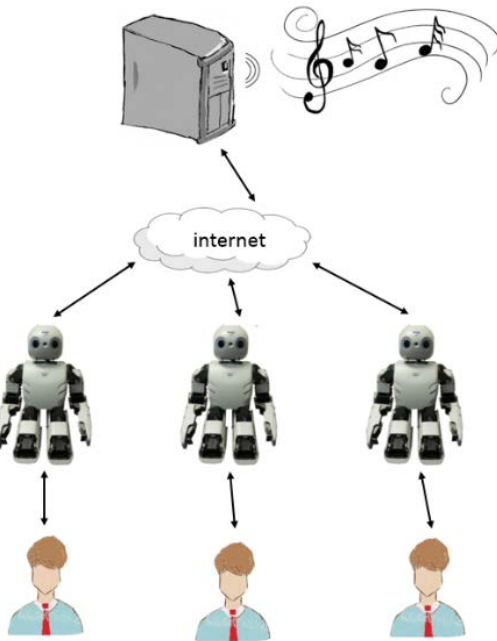

Figure 4. A client-server architecture with ROBOTIS OP2.

Our second version client is ROBOTIS OP2, which is a PC-based humanoid research robot. ROBOTIS OP2 has convenient interfaces, such as a camera and microphone, which enable natural interactions between children and robots. ROBOTIS OP2 has two 6-DOF legs, two 3-DOF arms and one 2-DOF neck. These joints allow ROBOTIS OP2 to simulate a number of human movements, such as walking, falling, standing up, looking around, or waving. ROBOTIS OP2 has a built-in PC with an Intel Atom N2600 @1.6 GHz dual core CPU, 4G DDR3 RAM and 32 GB mSATA, and runs Linux operating system. It provides programmers with a familiar environment and decent computing power to developing applications utilizing ROBOTIS OP2's interfaces. However, compared to other computing devices, such as personal laptops, desktop computers, or commercial servers, ROBOTIS OP2 is low powered and has a particular problem for process that requires intensive computation. Such tasks may involve graphical or audio signal processing, artificial intelligence, and multi-threaded applications.

Therefore, we have designed a human-robot interaction system, using ROBOTIS OP2 as an autonomous terminal to interact with children through natural interfaces including sounds (speech and non-speech) and visuals, and passing computation-intensive tasks to the remote server. The result is returned by the server and used as a reference by ROBOTIS OP2 to decide its next interaction.

4.2.2. Sensing Devices as Clients

Even though each robot has multiple cameras, it is hard to monitor the overall interaction scene between children and robots, using those embedded cameras. Therefore, we have incorporated an RGB-D depth sensor (Microsoft Kinect) to monitor the physical activities of a child and a robot to estimate the social engagement. We also plan to use multiple Kinects and other sensing devices to estimate their interactions more effectively.

4.2.3. Integrated Sonification Server

Our system includes a depth camera, one or more terminal robots (e.g., Romo, ROBOTIS OP2, and the Nao, see Figure 3), and a sonification server. Again, the clients and the server are connected using the reliable TCP/IP protocol. One of the salient differences between the first version prototype (Romo) and the current version system (ROBOTIS OP2) is the enhanced and integrated functions of the server. While the iPhone of Romo processes image data and performs facial expression detection, ROBOTIS OP2 just plays a role as a terminal. The server uses multiple threads to handle connections with each terminal robot and the depth camera. For example, when interacting with children, the terminal robot transfers the data captured by its cameras to the server as a sequence of images. The server performs an analysis of each image to detect a face. A detected face is marked with a rectangle and the position of its center is returned to ROBOTIS OP2. When ROBOTIS OP2 is notified by the server that a face is detected, it rotates its head to the direction where the face is detected and greets the person by a pre-recorded script to introducing itself. If ROBOTIS OP2 does not detect any faces, it will search for a face by rotating its head from side to side.

On the other hand, to evaluate children's physical activities and social interactions based on the data from the depth camera, we have incorporated metrics from physical therapy and rehabilitation [19]. However, the metrics are usually defined verbally based on physical rehabilitation conditions, and the underlying equations for deriving the metrics are not clearly defined. Thus, for assessing participating children's gestures and small motions, we have determined from the literature that the best approach for our problem is to use the following metrics: range of motion (ROM), path length (PATH), peak velocity (PV), average velocity (AvgV), and movement units (MUs). Besides these parameters, we will also utilize the child's specific motions or gestures to train the robot to attract more attention from the child.

The integrated sonification server was programmed in Java, which enables us to access functionalities of our existing

JFugue platform easily. For example, the server can have a further analysis on the data it receives from the terminal robots and the depth camera, and uses the statistical data to generate an ambient music on-the-fly. For further research using diverse mappings, we specifically developed a sonification mapping platform in the integrated sonification server.
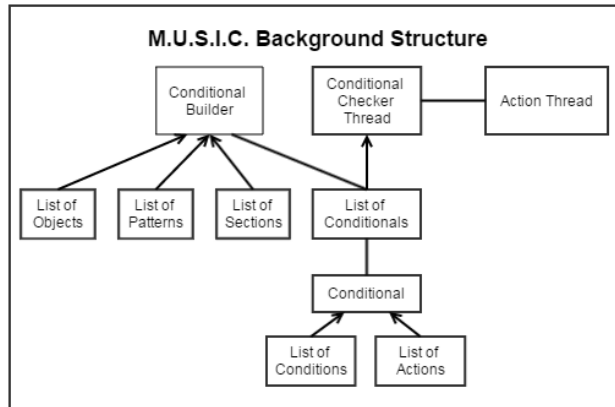
Sonification Mapping Platform

Figure 5. The M.U.S.I.C. platform's background structure.

To make a more flexible, robust, and easy to use interactive environment for sonification mappings with the clients' real-time data (e.g., children's emotion data and physical movement data), we have developed the Musical Utility Software for Interactive Creations (M.U.S.I.C.) platform. The M.U.S.I.C. platform allows researchers to perform actions when the data meet created criteria in the form of conditions. The M.U.S.I.C. allows researchers to create simple one-on-one mappings between tracked emotion and movement data, and pre-defined MIDI notes or melody patterns. In the M.U.S.I.C. structure, researchers can make this mapping both dynamic and static, to load the program from launch or to make changes in real-time. The structure in the diagram (Figure 5) has been simplified to only describe the conditionals rather than the program as a whole. A conditional is a combined set of conditions and actions. When all conditions are met, all actions are performed in a separate thread. This allows for multiple conditionals to be run concurrently as well as maintaining a high speed. The conditionals are built in a conditional builder. The conditional builder has a reference to all of the current objects, patterns, and sections as well as a list of previously made conditionals.

All of these references are acting independently of the conditional builder, and thus, allowing changes to occur at any point in the development process. Conditionals have a flag that determines if the conditional is enabled or disabled. The conditional checker is a separate thread that has a reference to the conditional list and is checking all conditionals that are enabled. Therefore, the process looks as follows:

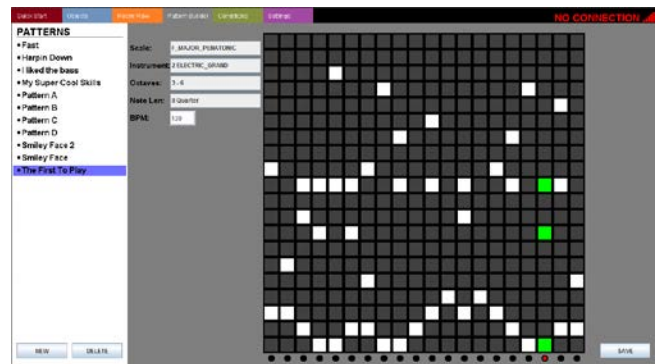**Figure 6. The pseudo code of the conditional checker.**

Figure 7. The pattern view tab.

This is the generalized pseudo code of the conditional checker (Figure 6). It checks every enabled conditional, and then checks all conditions within that conditional. Notice how the flag doActions is set true. This is so that in situations where there are no conditions, all actions can still be performed. In turn, this allows for chosen actions to always be performed. This code does run at $O(n^2)$ time. In the future, this conditional checker will be run by multiple threads so that runtime can be improved.

This structure is the reason the program works and allows for dynamic adjustments of conditionals. Although this structure may seem complex, the actual GUI is easy to understand (Figure 7). Researchers can make any patterns for each emotion or each movement of each child. Also, they can easily map the entire data stream with a certain sonification pattern. The structure makes it very convenient to add new conditions and actions. For example, when the algorithms for a new movement recognition are created (e.g., drawing a circle in a 3D space), it is just a matter of adding a new condition option for researchers to utilize.

We plan to optimize the code to further improve performance and run usability tests, such as heuristic evaluation or cognitive walkthrough to improve the research environment. Since this application is currently built in Java, we are exploring the option of moving this application to Android as well for better mobility. Being able to use the program on the go, and with extra additions of more robots and sensing devices to connect with the sonification server, could really take this program to the next level of usability and accessibility for the researchers and clinicians without a programming background.

5. RESEARCH SCENARIOS

To conduct therapeutic sessions with children with ASD, we have developed relationships with Autism Awareness Groups, local elementary schools, and the Center for Autism in three different sites (Michigan, Washington DC, and Georgia). As mentioned, we have developed our research scenarios based on the phases in music therapy: receptive, recreative, compositional, improvisational, and musical activity therapies. Some of our initial research scenarios are as follows.

5.1. Receptive Therapy

"Musical walk with a robot" is a type of receptive activity. For example, we selected Antonio Vivaldi's "Four Seasons" as a thematic music, because it is a well-known classical piece with various themes and a sensational storyline, which can be smoothly related to different emotional sensations. Throughout the scenario, the robot will take the child to the four physical stages, each of which consists of different sounds and images evoked during spring, summer, fall, and winter seasons.

5.2. Musical Activity Therapy

As a musical activity scenario, we consider a traditional children's game, "Dance Freeze". With the child and the robot facing each other, music will be played and stopped at random (or pseudorandom) intervals. While the music is going, the players dance or move around the room. When it stops, they must freeze in place. If they move before the music starts again, the robot will play a disappointing sound. If they do not, a happy reward sound will be played, and then the music will start again for more dancing. The music can be varied to regulate the speed of the game and the arousal of the child. Depending on the aptitude of the child, it could be made more complicated by dictating what dance moves should be done and alternating who controls the game.

5.3. Recreative and Compositional Therapy

In addition to receptive or reactive scenarios, we also plan to conduct a recreative or co-creational sonification activity session. To illustrate, the robot will initially play a portion of music and perform actions (e.g., move or dance) correlated to the music. Then, the child can respond to the robot's actions. The task will be interactive, such as kicking a ball toward each other or mimicking behaviors of each other. Each action sequence of the robot and the responses of the child will be translated into harmonized sonification. The music and the robotic actions will be built a priori according to our music-action mapping algorithm. The child's approximate movement and action can generate melodious sounds. If the child does not respond to the stimulus, the robot can continue to act while dynamically changing the music based on the behavior, distance, or interaction patterns of the child. If the child responds to the stimulus, then the robot will take turns and lead the child by suggesting a gesture or task to enhance the sonification outcomes and the social interactions.

6. CONCLULSION AND FUTURE WORKS

To promote social and emotional interactions of children with ASD, we have designed an orchestration robot platform using a client-server architecture. Given that children with ASD literally show a large "spectrum", we need to be cautious in selecting musical variables and making sonification. Even though the participating child prefers auditory stimuli in general, we need to empirically construct an optimized sonification zone to avoid the plausible reverse effect.

Another aspect to consider is the selection of an appropriate robot. Some children might prefer animal robots, whereas others might prefer humanoids. To this end, we are conducting robot acceptance research simultaneously. Our client-server structure will allow us to flexibly adapt to the environments of using different types of robots for each child. We believe this multifaceted approach will ultimately lead to more efficient and effective interventions for children with ASD.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Juslin, P. N., and Vastfjall, D. 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31, 559-621.

[2] Dautenhahn, K. 1999. Robots as social actors: Aurora and the case of autism. In *Proceedings of the Third International Cognitive Technology Conference, August, San Francisco*. Vol. 359. CT'99

[3] Scassellati, B. 2009. Affective prosody recognition for human-robot interaction. *Microsoft Research's External Research Symposium*. Redmond, WA, USA. 2009.

[4] Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J. and Miner, N. e. a. 1999. The sonification report: Status of the field and research agenda. Report prepared for the National Science Foundation by Members of the International Community for Auditory Display (Santa Fe, NM: International Community for Auditory Display (ICAD)).

[5] Prelock, P. A., and McCauley, R. J. 2012. Treatment of Autism Spectrum Disorders. Baltimore: Brooks Publishing.

[6] Hanson, E., Kalish, L. A., Bunce, E., Curtis, C., McDaniel, S., Ware, J., and Petry, J. 2007. Use of complementary and alternative medicine among children diagnosed with autism spectrum disorder. *Journal of autism and developmental disorders*, 37, 4, 628-636.

[7] Moore, D., McGrath, P., and Thorpe, J. 2000. Computer-aided learning for people with autism—a framework for research and development. *Innovations in Education and Training International*, 37, 3, 218--228.

[8] Hailpern, J. 2007. Encouraging speech and vocalization in children with autistic spectrum disorder, *SIGACCESS NEWSLETTER*, 89, (Sept. 2007).

[9] Feil-Seifer, D. and Mataric, M. 2008. Robot-assisted therapy for children with Autism Spectrum Disorders, *IDC Proceedings – Workshop on Special Needs*, June 11-13, Chicago, IL.

[10] Michaud, F. and Theberge-Turmel, C. 2002. Mobile robotic toys and autism. In Dautenhahn, K. (Ed.), *Socially Intelligent Agents – Creating Relationships with Computers and Robots*, Springer, 125-132.

[11] Scassellati, B. 2007. How social robots will help us to diagnose, treat, and understand autism, In S. Thrum, R. Brooics, H. Durrant-Whyte (Eds.), *Robotics Research*, STAR 28, 552-563.

[12] Stanton, C. M., Kahn, Jr., P. H. Severson, R. L., Ruckert, J. H., and Gill, B. T. 2008. Robotic animals might aid in the social development of children with autism, In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (Amsterdam, Netherlands, March, 12-15), HRI '08. ACM.

[13] Werry, I., Dautenhahn, K., and Harwin, W. 2001. Investigating a robot as a therapy partner for children with autism. In *Proceedings of the European Conference for the Advancement of Assistive Technology* (Ljubljana, Slovenia. Sept.) AAATE'01.

[14] Accordino, R., Comer, R., and Heller, W. B. 2007. Searching for music's potential: A critical examination of research on music therapy with individuals with autism. *Research in Autism Spectrum Disorders*, 1, 1, 101-115.

[15] Gold, C., Wigram, T., and Elefant, C. 2006. Music therapy for autistic spectrum disorder. *The Cochrane Library*.

[16] Rosen, H. J., and Levenson, R. W. 2009. The emotional brain: Combining insights from patients and basic science. *Neurocase,* 15, 3, 173-181.

[17] Sterkenburg, J., Jeon, M., and Plummer, C. 2014. Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons. In M. Kurosu (Ed) *Human-Computer Interaction, Part II, HCII 2014, Lecture Notes in Computer Science (LNCS) 8511*, 633-640. Springer International Publishing Switzerland.

[18] Liao S., Zhu, X., Lei, Z., Zhang, L., and Li, S. Z. 2007. Learning multi-scale block local binary patterns for face recognition. In: Lee, S.-W., Li, S. Z. (eds.) *ICB 2007. LNCS.* 4642. 828–837. Springer, Heidelberg.

[19] Brooks, D., and Howard, A. M. 2010. A computational method for physical rehabilitation assessment. In *Proceedings of the 3rd IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics. BioRob'10* (pp. 442-447). IEEE.