Music Education Robot Platform for Children with Autism

*Note: Sub-titles are not captured in Xplore and should not be used

1st Huanghao Feng
Ritchie School of Engineering & Computer Science
University of Denver
Denver, USA
huanghao.feng@du.deu

2nd Mohammad H.Mahoor
Ritchie School of Engineering & Computer Science
University of Denver
Denver, USA
m.mahoor@du.edu

***Abstract*—**

***Index Terms*—Social Robot, Autism, Music Therapy, Turn-Taking, Motor Control, Emotion Classification**

I. Introduction

Recent study indicates that music has played a important role in children's daily life such as waking time, streaming from radios, televisions, cell-phones, computers and toys [23]. Since children with autism spent most of the time with technology product nowadays, music could play an important role in their life as well. The symptoms of autism spectrum disorders, a disorder of neural development, include but not limited impaired social interaction and communication [15]. In order to help this population, different therapy methods have been developed and some are widely use in autism recovery, such as behavior therapy, game therapy, art therapy, music therapy and more [3]. Most of the time, treatment for autistic children, mediators are required because majority of them may not able to play with kids with autism directly, for example, drawing for art therapy, game for game therapy and instrument for music therapy.

Many researches shows that children with autism have less interest in communicating with human due to sensing overwhelming issue. A robot with still face could be a good agent with less intimidating characteristics for helping children with autism. There is also researches show that kids with autism are more attracted to interact with humanoid social robots in daily life [5], [7], [16], [21]. That makes social assistive robot a perfect media for delivering certain therapy method, such as music therapy. Significant amount of reports suggest that using music as a assistive method, also known as music therapy, for helping individuals with autism can be beneficial. Composed songs and improvisational music therapy were used as a music techniques in such activities. However, there was limited evidence to support the use of music interventions under certain conditions to conduct social, communicative and behavioral skills in early age children with autism. Patients can get a feeling for the music by listening, singing, playing instruments, and moving. Music therapy for children is conducted either in a one-on-one session or in a group session, and it can help children with problems in communication, attention, and motivation, as well as with behavioral problems [8]. Motivation and emotion are essential to music education, together they ensure that students acquire new knowledge and skills in a meaningful way. Much has been reported that music has been viewed as a means of engaging the children and therapists as a non-verbal aspect in musical-emotional communication [22].

The rest of this paper is organized as follows: Section II presents some related works concerning human-robot interaction in multiple intervention methods. Section III elaborates the experiment design process including hardware and session details. An intelligent music teaching platform is presented in Section IV and experiment results are given in Section V. Finally, Section VI concludes the paper with some remarks for future research.

II. Related Works

Music is effective method to involve children with autism in rhythmic and non-verbal communication. Besides, music has often been used in therapeutic sessions with children who have suffer from mental and behavioral disabilities [2], [17]. Nowadays, at least 12% of all treatment of individuals with autism consist of music-based therapies [1]. Specifically, teaching and playing music to children with autism spectrum disorders (ASD) in therapy sessions have shown great impact for improving social communication skills [10]. Recorded music or human played back music are used in single and multiple subjects' intervention session from many studies [1], [4]. Different social skills are targeted and reported (i.e. eye-gaze attention, joint attention and turn-taking activities) in using music-based therapy sessions [9], [18]. Noted that improving gross and fine motor skills for ASD through music interventions is a missing part in this field of studies [1].

Identify applicable funding agency here. If none, delete this.

Socially assistive robots are widely used in young age of autism population interventions these years. Some studies are focusing on eye contact and joint attention [7], [11], [12], showing that at some point the pattern of ASD group in perceiving eye gaze are similar to typically developed (TD) kid, and eye contact skills can be significantly improved after intervention sessions. Plus, these findings also provides a strong evidence of ASD kids are easy to attracted to humanoid robots in various type of social activities. Some groups start to use such robots to conduct music-based therapy sessions nowadays. Children with autism are asked to imitate play music based on Wizard of Oz style and Applied Behavior Analysis (ABA) models from humanoid robots in intervention sessions for practicing eye-gaze and joint attention skills [14], [19], [20]. However, some disadvantages of such research due to lack of sample size and no automated system in human-robot interaction. Music can be used as unique window into the world of autism, lots of evidence suggest that many individuals with ASD are able to understand simple and complex emotions in childhood using music-based therapy sessions [13]. Although limited research has found in such area especially using bio-signals for emotion recognition for ASD and TD kids [6] in understanding the relationship between activities and emotion changes.

To this end, in current research a automated music-based social robot platform with activity-based emotion recognition system is presented in the following sections. The purpose of this platform is to provide a possible ultimate solution for assisting children with autism to improve motor skills, turn-taking skills and activity engagement initiation. Further more, by using bio-signals with Complex-Morlet (C-Morlet) wavelet feature extraction [6], emotion classification and emotion fluctuation are analyzed based on different activities. TD kids are participated as control group in order to see the difference from ASD group.

III. Experiment Design

Nine ASD kids (average age: 11.73, std: 3.11) and 7 TD kids (average age: 10.22, std: 2.06) were recruited in this study. For each participant in ASD group, 6 sessions were be delivered including baseline session, intervention sessions and exit session. As for TD control group, only baseline and exit sessions were required for each participant. Each session lasts for 30-60 min total depends on the difficult of each session and performance of individuals.

A. Experiment Room

All the sessions were held in a 11ft x 9.5ft x 10ft room with six HD surveillance cameras installed at corners, side wall and ceiling of the experimental room see Figure 1. One mini hidden microphone attached at the ceiling camera for sending real time audio to the observation room in order to let the care giver to listen to. An external hand-held audio recorder were set in front of the participant during sessions to be able to collect high quality audio for future process.

As shown in Figure 1, the observation room is located at the back of the one-way mirror facing at the back of participants in order to avoid distraction while sessions on going. Real-time video and audio were broadcasting to the observation room during each session, which allowed researchers observe and record in the meantime. Parents behind the mirror may also call off the session in case of emergency.

B. NAO: A Humanoid Robot

All communication content were delivered by a humanoid robot agent called NAO developed by Aldebaran Robotics in France. NAO is 58 cm (23 inches) tall, with 25 degrees of freedom. This robot can conduct most human behaviors. It also features an onboard multimedia system including four microphones for voice recognition and sound localization, two speakers for text-to-speech synthesis, and two HD cameras with maximum image resolution 1280 x 960 for online observation. As shown in Figure 2, these utilities are located in the middle of the forehead and the mouth area. NAO's computer vision module includes facial and shape recognition units. By using the vision feature of the robot, the robot can see the instrument from its lower camera and be able to implement color detection module for self-calibration system which allows the robot to have real-time micro-adjustment for its arm-joints in case of off positioning during xylophone playing.

The robot arms have a length of approximately 31 cm. Each arm has five degrees of freedom and is equipped with sensors to measure the position of each joint. To determine the pose of the instrument and the mallets' heads, the robot analyzes images from the lower monocular camera located in its head, which has a diagonal field of view of 73 degrees. These dimensions allow us to choose a proper instrument.

The four microphone locations embedded on the NAO's head can be seen in Figure 2. According to the official Aldebaran documentation, these microphones have sensitivity of 20mV/Pa +/-3dB at 1kHz, and an input frequency range of 150Hz - 12kHz. Data will be recorded as a 16 bit, 48000Hz, 4 channels wav file which meets the requirements for designing the online feedback audio score system described below.

C. Hardware Accessories

Due to the purpose of this study, some necessary accessories needed to be purchased and build before the robot was able to play music. All accessories will be discussed in the following.

1) Xylophone: A Toy for Music Beginner: In this system, due to NAO's open arms' length, we choose a Sonor Toy Sound SM soprano-xylophone with 11 sound bars of 2 cm in width. The instrument has a size of 31 cm x 9.5 cm x 4

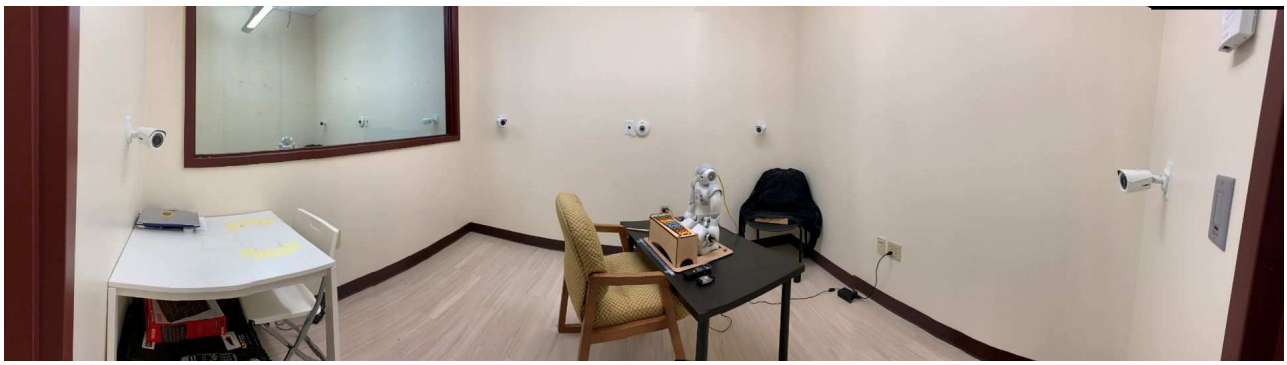

Fig. 1. Experiment Room

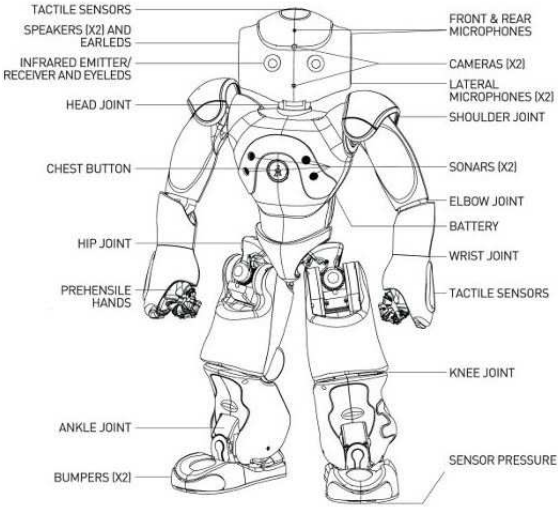

Fig. 2. A Humanoid Robot NAO: 25 Degrees of Freedom, 2 HD Cameras and 4 Microphones

cm, including the resonating body. The smallest sound bar is playable in an area of 2.8 cm x 2 cm, the largest in an area of 4.8 cm x 2 cm. The instrument is diatonically tuned in C-Major/a-minor. For the beaters/mallets, we used the pair that came with the xylophone with a modified 3D printed grip (details in next subsection) to allow the robot's hands to hold them properly. The mallets are approximately 21 cm in length and include a head of 0.8 cm radius. The 11 bars of the xylophone represent 11 different notes (11 frequencies) which covers approximately a one and half octave scale starting from C6 to F7.

2) Mallet Gripper Design: According to the size of Nao's hands, we designed and 3D printed a pair of gripers to have the robot be able to hold the mallets properly. Shown in Figure 3.

3) Instrument Stand Design: A wooden base was designed and laser cut to hold the instrument in the proper place for the robot to be able to play music. Shown in Figure 4.

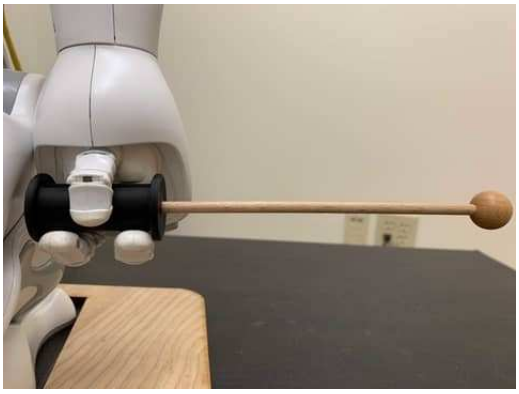

Fig. 3. Mallet Griper

Fig. 4. Instrument Stand Front View.

D. Experiment Sessions

Two parts were included in the baseline and exit session which were 1) music practice and 2) music game play. In the music practice part, participants were asked to complete a full set of music play activities including listen to the music, single note play, multiple notes play, half song play and the whole song play. Three entertaining game modes were designed in the music game play part, participants were

allowed to communicate with robot regarding which mode to play with. Mode 1: robot will randomly play a song from its song bank for kids to listen to; Mode 2: robot randomly generates a sequence of notes with consonance or dissonance style, requests an oral emotion feeling from participants and physical playback afterwards; Mode 3: allows participants to have a 5 seconds of free play and challenge the robot to imitate from the participants what just played. There was no limit for how many times each individual who wants to play each time, but at least play each mode once in single session. The only difference between baseline and exit session was the song which used in them, in baseline session, "Twinkle Twinkle Little Star" was used as a standard entry level song for all participants, and a customized song were chosen by each individual for exit session in order to motivate participants for better learning music, which makes it more difficult from the baseline session.

Each intervention session has divided into three parts: S1) warm up; S2) single activity practice (with color hint); and S3) music game play. Starting from intervention sessions, customized song were used in the following sessions in order to motivate participants and have them more engage to multiple repetitive activities. The purpose of having warm up section is to have the motor control skill been practiced and meanwhile to help participants implement the motor skills in next activities. Single activity was based on music practice from baseline/exit session, other then those sessions, single activity will only have one type of music practice each individual session, for instance, single note play were delivered in the first intervention session, then the next time this practice will become multiple notes play and the level of difficulty for music play were gradually increased session after session. This was in order to make a challenge based engagement activity for ASD group for better motivation and emotion stimuli. As for music game play were remain the same as baseline/exit session.

IV. Xylo-Bot: An Interactive Music Education System

In this section, a novel module-based robot-music education system will be presented. Three modules have been built in this intelligent system including Module 1: color-based self-calibration micro-adjustment system; Module 2: joint trajectory generator; and Module 3: real time performance scoring feedback. See Figure 5.

A. Module 1: Eye-hand Self-Calibration Micro-Adjustment

Knowledge about the parameters of the robot's kinematic model is essential for tasks requiring high precision, such as playing xylophone. While the kinematic structure is known from the construction plan, errors can occur, e.g., due to the imperfect manufacturing. After multiple rounds of testing, it was found the targeted angle chain of arms never actually equals the returned chain. We therefore used a calibration method to accurately eliminate these errors.

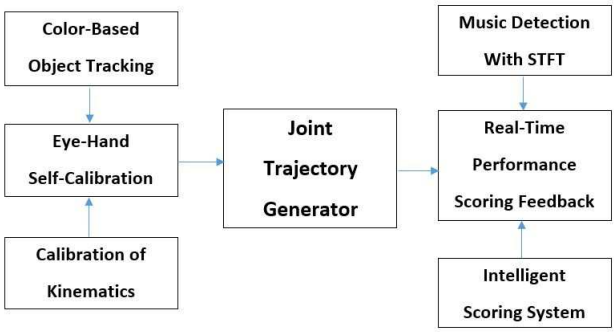

Fig. 5. Block Diagram of Module-Based Acustic Music Interactive System

1) Color-Based Object Tracking: To play the xylophone, the robot has to be able to adjust its motions according to the estimated relative position of the instrument and the heads of the beaters it is holding. To estimate these poses, we used a color-based technique. The main idea is, based on the RGB color of the center blue bar, given a hypothesis about the instrument's pose, one can project the contour of the object's model into the camera image and compare them to actually observed contour. In this way, it is possible to estimate the likelihood of the pose hypothesis. By using this method, it allows the robot to track the instrument with very low cost in real-time. See Figure 6.

B. Module 2: Joint Trajectory Generator

Our system parses a list of hex-decimal numbers (from 1 to b) to obtain the sequence of notes to play. It converts the notes into a joint trajectory using the beating configurations obtained from inverse kinematics as control points. The timestamps for the control points will be defined by the user in order to meet the experiment requirement. The trajectory is then computed using Bezier interpolation in joint space by the manufacturer-provided API and then sent to the robot controller for execution. In this way, the robot plays in-time with the song.

C. Module 3: Real-Time Performance Scoring Feedback

The purpose of this system is to provide a real-life interaction experience using music therapy to teach kids

Fig. 6. Color Detection From NAO's Bottom Camera Color Based Edge Detection.

social skills and music knowledge. In this scoring system, two core features were designed to complete the task: 1) music detection; 2) intelligent scoring-feedback system. All result were be saved in CSV files for future data analysis.

1) A. Music Detection: Music, in the understanding of science and technology, can be considered as a combination of time and frequency. In order to make the robot detect a sequence of frequencies, we adopted the short-time Fourier transform (STFT) to this audio feedback system. This allows the robot to be able to understand the music played by users and provide the proper feedback as a music teaching instructor.

The short-time Fourier transform (STFT) , is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. In practice, the procedure for computing STFTs is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment. This reveals the Fourier spectrum on each shorter segment. In the discrete time case, the data to be transformed could be broken up into chunks or frames (which usually overlap each other, to reduce artifacts at the boundary). This can be expressed as:

$$\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}$$

likewise, with signal x[n] and window w[n]. In this case, m is discrete and ω is continuous, but in most typical applications, the STFT is performed on a computer using the Fast Fourier Transform, so both variables are discrete and quantized.
The magnitude squared of the STFT yields the spectrogram representation of the Power Spectral Density of the function:

$$\text{spectrogram}\{x(t)\}(\tau,\omega) \equiv |X(\tau,\omega)|^2$$

After the robot detects the notes from user input, a list of hex-decimal number will be returned. This list will be used in two purposes: 1) to compare with the target list for scoring and sending feedback to user; 2) used as a new input to have robot playback in the game session as discussed in the next chapter.

2) B. Intelligent Scoring-Feedback System: In order to compare the detected notes and the target notes, an algorithm which is normally used in information theory linguistics called Levenshtein Distance. This algorithm is a string metric for measuring the difference between two sequences.

In current case, the Levenshtein distance between two string-like hex-decimal numbers a, b (of length $|a|$ and $|b|$ respectively) is given by $\text{lev}_{a,b}(|a|,|b|)$ where

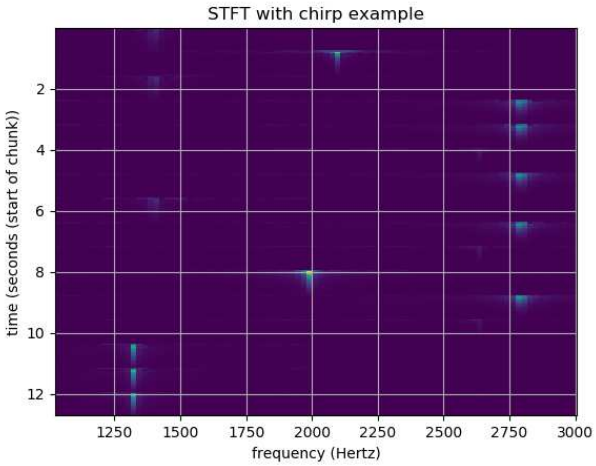

Fig. 7. Melody Detection with Short Time Fourier Transform

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j)+1 \\ \text{lev}_{a,b}(i,j-1)+1 \\ \text{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} \end{cases}$$

where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$ and equal to 1 otherwise, and $\text{lev}_{a,b}(i,j)$ is the distance between the first i characters of a and the first j characters of b.

Note that the first element in the minimum corresponds to deletion (from a to b), the second to insertion and the third to match or mismatch, depending on whether the respective symbols are the same.

Based on the real life situation, we defined a likelihood margin for determining whether the result is good or bad:

$$likelihood = \frac{len(target) - lev_{target,source}}{len(target)}$$

where if the likelihood is greater than 66% - 72%, the system will consider it as a good result. This result will be passed to the accuracy calculation system to have the robot decide whether it needs to add more dosage to the practice. More details will be discussed in the next chapters as it relates to the experiment design.

V. Experimental Results

9 ASD and 7 TD participants finished this study in 8 months, all ASD subjects completed 6 sessions including intervention sessions and TDs for 2 sessions with only baseline/exit session. By using Wizard of Oz control style, a well trained researcher were conducting the baseline and exit sessions for better observation and evaluation quality of performance. With well designed fully automated intervention sessions, NAO were able to initiate music teaching activities with participants.

Since the music detection method was sensitive to the audio input, that requires clear and long lasting sound from xylophone. From Figure 8, it is obvious that majority of subjects were able to strike or play xylophone in proper way after one or two sessions. Notice that subject 101 and 102 had significant improvement curve during intervention sessions. Some of the subjects started at a higher accuracy rate, and kept this rate above 80%, which can be considered as consistent motor control performance even with up and downs. Two subjects (103 & 107) were having difficult time with playing xylophone and following turn-taking cues with agent robot. This fact affected the performance in following activities for both subjects.

Figure 9 shows the accuracy result of main music teaching activity for intervention sessions across all participants. Learning how to play one's favorite song can be considered as a motivation for ASD kids understanding and learning turn-taking skill. As described in previous section, the difficulty level of this activity were designed uprising. By this fact, accuracy of the performance from participants were expected to decrease. This activity requires participants able to concentrate and using joint attention skills in robot teaching stage and also respond properly afterwards. Enough waiting time were given after robot says: 'Now, you shall play right after my eye flashes', participants were also received an eye color change cue from the robot in order to complete a desired music-based social interaction. Different from warm up section, notes played in correct sequence of order can be considered as a good-count strike. From Figure 9, most of the participants were able to complete single/multiple notes practice with an average 77.36%/69.38% accuracy rate, although even with color hints, notes' pitch difference still can be a core challenge for them. Due to the difficulty of session 4 and 5, worse performance comparing to previous two sessions were accepted. However, more than half of the participants showed a consistent high performance accuracy or even better result than previous sessions. Combining the report from video annotators, 6 out of 9 subjects showed strong engaging behavior in playing music, especially after first few sessions. Better learn-play turn-taking rotation were performed over time, and significant increase of performance by 3 subjects, reveal turn-taking skills were picked up from this activity.

EDA signal was also collected in this study. By using the annotation and analysis method from previous work [6], a music-event-based emotion classification result will be presented below. In order to find out the emotion secret of ASD group, multiple comparison were made after annotate the videos. Different activities may cause emotion arousal change. As presented above, warm up section and single activity practice section have same activity in different level of intensities, and game play has the lowest difficulty and more relax.

In the first part of analysis, EDA signals were segmented into small event-based pieces according to the number of "conversations" in each section. One "conversation" was defined with 3 movements: a) robot/participant demonstrates the note(s) to play; b) participant/robot repeat the note(s); c) robot/participant presents the result, and each segmentation last about 45 seconds. The continuous wavelet transform (CWT) of the data assuming complex Morlet (C-Morlet) wavelet function was used inside a frequency range of (0.5, 50)Hz, a SVM classifier was then employed to classify "conversation" segmentation among 3 sections using the wavelet-based features. Table I shows the classification accuracy for the SVM classifier with different kernel functions. As can be seen, emotion arousal change between S1 and S2, S2 and S3 can be classified using wavelet- based feature extraction SVM classifier with average accuracy of 76% and 70%. With highest 64% of accuracy for S1 and S3, that may indicates less emotion changes between warm up and game sections.

In order to discover the emotion fluctuation inside of one task, each "conversation" section has been carefully divided into 3 segments as described above. Each segment last about 10 - 20 seconds. Table II shows the full result of emotion fluctuation in warm up (S1) and music practice (S2) sections from intervention session. Notice that all of the segments cannot be classified properly using existing method. Both SVM and KNN show the stable results. This may suggests that ASD group may have less emotion fluctuation or arousal change once task starts even with various activities in it. Stable emotion arousal in single task could also benefit from the proper activity content, including robot agent play music and language usage during conversation. Friendly voice feedback was based on the performance delivered by participants were well written and stored in memory, both positive award while receive correct input and encouragement while play incorrect. Since emotion fluctuation can affect learning progress, less arousal change indicates the design of intervention session were robust.

Cross sections comparison also presented blow. Since each "conversation" contains 3 segments, it is necessary to have specific segments from one task to compare with the other task corresponded to. Table III shows the classification rate in robot demo, kids play and robot feedback across warm up (S1) and music practice (S2) sections. By using RBF kernel, wavelet-based SVM classification rate has 80% of accuracy for all 3 comparisons. This result also matches the result from Table I.

The types of activities and process of the session between baseline session for both group were exactly the same. By using the "conversation" concept above, each of them has been segmented. Comparing with target and control groups using same classifier, 80% of accuracy for detecting different groups. See Table III. Video annotators also reported "unclear" in reading facial expressions from ASD group. These combined messages suggests that, even with same activities different bio-reaction were completely opposite between TD and ASD groups. It has also been reported that, significant improvement of music performance were shown in ASD group, although both groups have similar performance at their baseline sessions. Further more, TD group were shown more willing to try to make their performance as

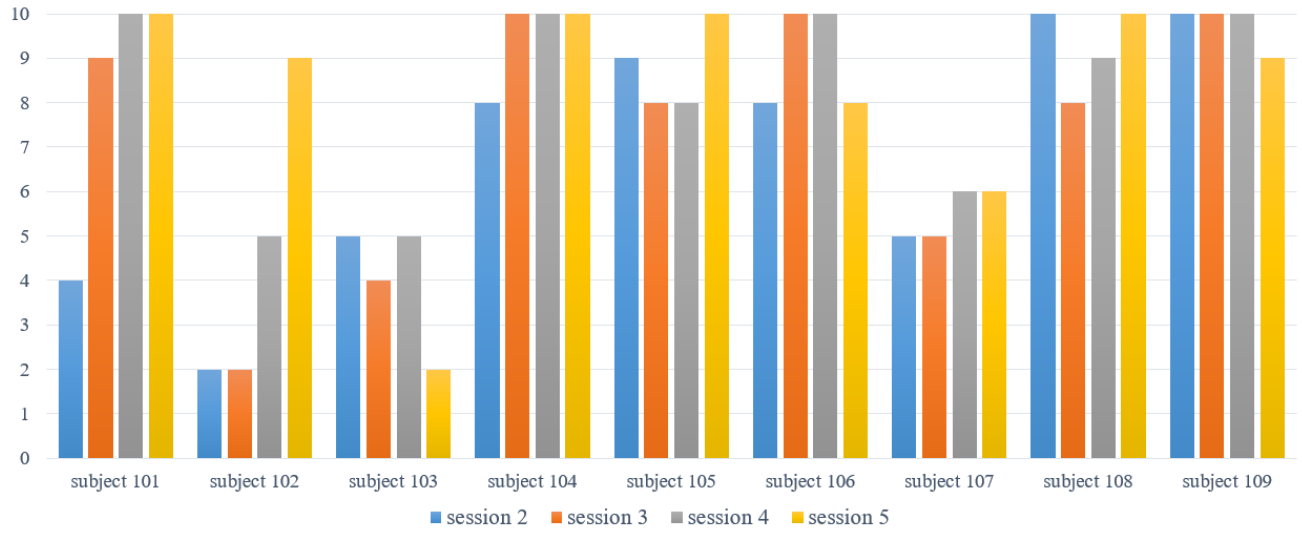

Fig. 8. Motor Control Accuracy Result

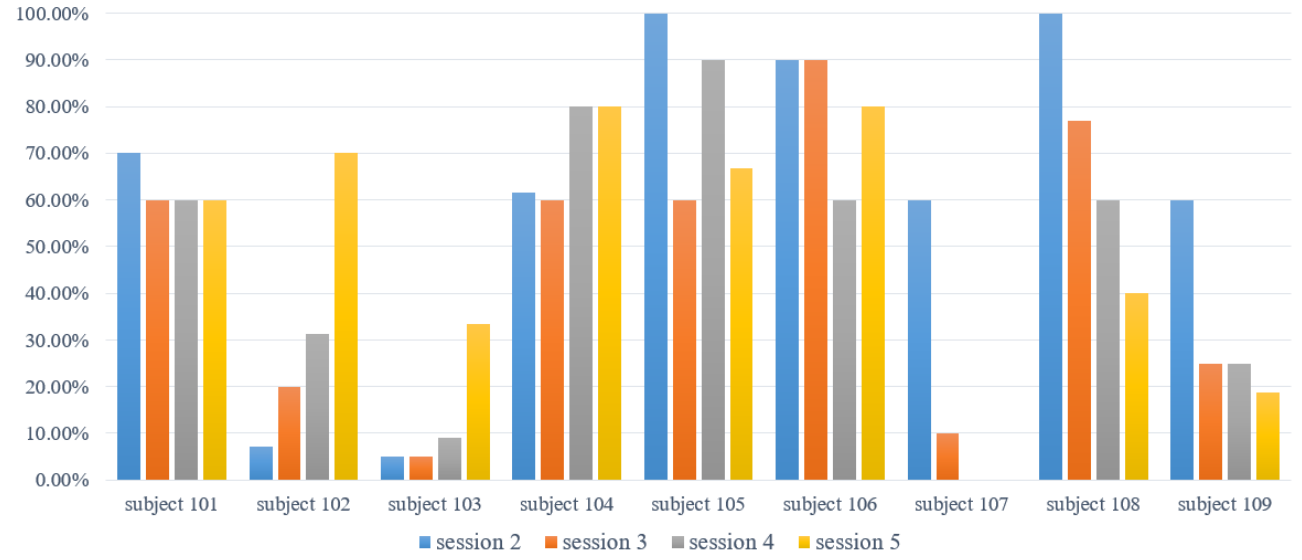

Fig. 9. Main Music Teaching Performance Accuracy

TABLE I
EMOTION CHANGE IN DIFFERENT SECTIONS

	Kernels	Accuracy	AUC	Precision	Recall
S1 vs S2	**Linear**	75	78	76	72
S1 vs S3		57	59	56	69
S2 vs S3		69	72	64	86
S1 vs S2 vs S3		52			
S1 vs S2	**Polynomial**	66	70	70	54
S1 vs S3		64	66	62	68
S2 vs S3		65	68	62	79
S1 vs S2 vs S3		50			
S1 vs S2	**RBF**	76	81	76	75
S1 vs S3		57	62	57	69
S2 vs S3		70	76	66	83
S1 vs S2 vs S3		53			

TABLE II
EMOTION CHANGE IN DIFFERENT SECTIONS

Segmentation Comparison in Single Task								
	Warm up Section				Song Practice Section			
	Kernels	Accuracy	K value	Accuracy	Kernels	Accuracy	K value	Accuracy
learn vs play	Linar	52.62	K = 1	54	Linar	53.79	K = 1	52.41
learn vs feedback		53.38		50.13		53.1		51.72
play vs feedback		47.5		50.38		54.31		50.86
learn vs play vs feedback		35.08		36.25		35.52		36.55
learn vs play	Polynomial	49	K = 3	50.25	Polynomial	53.79	K = 3	50.69
learn vs feedback		50.75		50.13		50.86		50.34
play vs feedback		49.87		49.5		49.14		52.07
learn vs play vs feedback		33.92		35.83		34.71		35.29
learn vs play	RBF	54.38	K = 5	48.37	RBF	50.86	K = 5	50.17
learn vs feedback		55.75		52.75		53.97		50.17
play vs feedback		51.12		50		53.79		52.93
learn vs play vs feedback		36.83		34.17		34.83		33.1

TABLE III
EMOTION CHANGE IN DIFFERENT TASKS

	Accuracy of SVM			Accuracy of KNN		
	Linear	Polynomial	RBF	K = 1	K = 3	K = 5
learn 1 vs learn 2	73.45	69.31	80.86	73.28	71.03	65
play 1 vs play 2	75.34	68.79	80	74.48	69.14	64.31
feedback 1 vs feedback 2	76.38	69.48	80.34	74.14	69.14	66.9

better as possible while they made mistakes.

VI. DISCUSSION, CONCLUSION AND FUTURE WORK

The results indicates that the presented music education platform can be considered as a good tool for help improving fine motor control, turn-taking skills and social activities engagement. The automated music detection system created a self-adjusting environment for participants in early sessions. Most of the ASD kids started to pick up the strike movement after first two intervention sessions, some even can master the motor skill during the very first warm up activity. Although the robot could provide verbal instructions and demonstrations by voice command input from participants whenever they need it. However, majority of the participants did not request such service while playing with the robot. This finding suggests that fine motor control skill can be learned from specific well-designed activities for young ASD population.

The purpose of using music teaching scenario as the main activity in the current research is to create a fine and natural turn-taking behavior chance during social interaction. By observing all experimental sessions, 6 out of 9 subjects could dominate proper turn-taking after one or two sessions. Note that subject 107 had significant improvement in last few sessions comparing to the baseline session. Subject 109 had trouble with focus on listening to the robot for most of the time. However, with researcher interfering, this kid can perform better back and forth music activity for a short time period. For practicing turn-taking skill, a fun motivated activity should be designed for children with autism. Music teaching could be a good example for accomplish this task by taking the advantage of customized songs which selected by individuals.

Starting the later half of the sessions, participants can start to recognize their favorite songs, over half of the participants were getting more into the activities, although the difficulty for playing proper notes were much higher. It is easy to notice that older kids who spent more time engaging with the activities during the song practice session comparing to younger kids, especially in half/whole song play sessions. Several reasons can explain this situation, one is because of the more complex the music, the more challenge and more concentration participants will face. Thus, older individuals may willingly accept the challenge and enjoy the sense of accomplishment afterwards based on their verbal feedback to the research at the end of each session. The music knowledge base could also be one of the reason that conducts this result, since older participants may have more chance to learn music at school. Game section of each session provides the highest engagement level of all time, not only because of this is for relax and fun play, but also offers an opportunity to participants regarding challenge the robot to mirror the free play from them, this interesting phenomenon can be considered as ”revenge”. Especially for subject 106, who spent significant amount of time in free play game mode. According to the session executioner and video annotators, this particular subject shows high level of engagement for all activities, including free play. Based on the conversation and music performance with robot, subject showed strong interest in challenging the robot with a friendly way.

Emotion study for children with autism is difficult. Bio-signal provides a possible way of doing that. Event-based emotion classification method presented in current research suggests that same activity with different intensities can cause emotion change in arousal dimension, although it is difficult

TABLE IV
TD VS ASD EMOTION CHANGES FROM BASELINE AND EXIT SESSIONS

	Linear	Polynomial	RBF
Accuracy	75	62.5	80
Confusion Matrix	63 37 12 88	50 50 25 75	81 19 25 75

to label the emotions based on facial expression change in video annotation phase for ASD group. Less emotion fluctuate in certain activity presented in Table II suggests that a mild friendly game like teaching system may motivate better social content learning for children with autism, even with repetitive movements. These well designed activities could provide a relaxed learning environment which helps participants to focus on learning music content with proper communication behaviors. This may explains the improvement for music play performance in song practice (S2) through intervention sessions in Figure 9. Comparing emotion patterns from baseline and exit sessions between TD and ASD groups in Table III, difference can be found. This may suggests a potential way of assist autism diagnose using bio-signal in early age. According to annotators and observers, TD kids showed strong passion in this research. Excitement, stressful, disappointment were easy to be recognized and labeled from the videos recorded. On the other hand, limited facial expression changes can be detected in ASD group. That makes it difficult to learn whether they have different feelings or they have same feelings but different bio-signal activities comparing to TD group. This could be a interesting research to dig into in the future. Further more, due to the limitation of the sample size, future research can be continued with different classification methods with larger population.

TABLE V
ADD CAPTION

#	Session Type	Outlines
1	Baseline Session	Include all music activities in one session, play song
2		Single music activity practice, all notes coming from
1-13-5		
3		Multiple notes music practice, all notes coming from
	Intervention Session	
1-13-5		
4		First half customized song practice
1-13-5		
5		Second half customized song practic
6	Exit Session	Include all music activities in one session, play cu

ACKNOWLEDGMENT

REFERENCES

[1] Anjana Narayan Bhat and Sudha Srinivasan. A review of "music and movement" therapies for children with autism: embodied interventions for multisystem development. *Frontiers in integrative neuroscience*, 7:22, 2013.

[2] Marianna Boso, Enzo Emanuele, Vera Minazzi, Marta Abbamonte, and Pierluigi Politi. Effect of long-term interactive music therapy on behavior profile and musical skills in young adults with severe autism. *The journal of alternative and complementary medicine*, 13(7):709–712, 2007.

[3] Susan E Bryson, Sally J Rogers, and Eric Fombonne. Autism spectrum disorders: early detection, intervention, education, and psychopharmacological management. *The Canadian Journal of Psychiatry*, 48(8):506–516, 2003.

[4] Blythe A Corbett, Kathryn Shickman, and Emilio Ferrer. Brief report: the effects of tomatis sound therapy on language in children with autism. *Journal of autism and developmental disorders*, 38(3):562–566, 2008.

[5] Sandra Costa, Hagen Lehmann, Ben Robins, Kerstin Dautenhahn, and Filomena Soares. " where is your nose?": developing body awareness skills among children with autism using a humanoid robot. 2013.

[6] Huanghao Feng, Hosein M Golshan, and Mohammad H Mahoor. A wavelet-based approach to emotion classification using eda signals. *Expert Systems with Applications*, 112:77–86, 2018.

[7] Huanghao Feng, Anibal Gutierrez, Jun Zhang, and Mohammad H Mahoor. Can nao robot improve eye-gaze attention of children with high functioning autism? In *2013 IEEE International Conference on Healthcare Informatics*, pages 484–484. IEEE, 2013.

[8] Timothy Gifford, Sudha Srinivasan, Maninderjit Kaur, Dobri Dotov, Christian Wanamaker, Gregory Dressler, Kerry MARSH, and Anjana BHAT. Using robots to teach musical rhythms to typically developing children and children with autism. *University of Connecticut*, 2011.

[9] Jinah Kim, Tony Wigram, and Christian Gold. The effects of improvisational music therapy on joint attention behaviors in autistic children: a randomized controlled study. *Journal of autism and developmental disorders*, 38(9):1758, 2008.

[10] Hayoung A Lim and Ellary Draper. The effects of music therapy incorporated with applied behavior analysis verbal behavior approach for children with autism spectrum disorders. *Journal of music therapy*, 48(4):532–550, 2011.

[11] S Mohammad Mavadati, Huanghao Feng, Anibal Gutierrez, and Mohammad H Mahoor. Comparing the gaze responses of children with autism and typically developed individuals in human-robot interaction. In *2014 IEEE-RAS International Conference on Humanoid Robots*, pages 1128–1133. IEEE, 2014.

[12] Diana Mihalache, Huanghao Feng, Farzaneh Askari, Peter Sokol-Hessner, Eric J Moody, Mohammad H Mahoor, and Timothy D Sweeny. Perceiving gaze from head and eye rotations: An integrative challenge for children and adults. *Developmental Science*, 23(2):e12886, 2020.

[13] Istvan Molnar-Szakacs and Pamela Heaton. Music: a unique window into the world of autism. *Annals of the New York Academy of Sciences*, 1252(1):318–324, 2012.

[14] Ying-Hua Peng, Cheng-Wei Lin, N Michael Mayer, and Min-Liang Wang. Using a humanoid robot for music therapy with autistic children. In *2014 CACS International Automatic Control Conference (CACS 2014)*, pages 156–160. IEEE, 2014.

[15] Isabelle Rapin and Roberto F Tuchman. Autism: definition, neurobiology, screening, diagnosis. *Pediatric Clinics of North America*, 55(5):1129–1146, 2008.

[16] Ben Robins, Kerstin Dautenhahn, and Paul Dickerson. Embodiment and cognitive learning–can a humanoid robot help children with autism to learn about tactile social behaviour? In *International Conference on Social Robotics*, pages 66–75. Springer, 2012.

[17] Nicole Roper. Melodic intonation therapy with young children with apraxia. *Bridges*, 1(8):1–7, 2003.

[18] Carolyn E Stephens. Spontaneous imitation by children with autism during a repetitive musical play routine. *Autism*, 12(6):645–671, 2008.

[19] Alireza Taheri, Minoo Alemi, Ali Meghdari, Hamidreza Pouretemad, Nasim Mahboub Basiri, and Pegah Poorgoldooz. Impact of humanoid social robots on treatment of a pair of iranian autistic twins. In *International Conference on Social Robotics*, pages 623–632. Springer, 2015.

[20] Alireza Taheri, Ali Meghdari, Minoo Alemi, Hamidreza Pouretemad, Pegah Poorgoldooz, and Maryam Roohbakhsh. Social robots and teaching music to autistic children: Myth or reality? In *International Conference on Social Robotics*, pages 541–550. Springer, 2016.

[21] Joshua Wainer, Kerstin Dautenhahn, Ben Robins, and Farshid Amirabdollahian. Collaborating with kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. In *2010 10th IEEE-RAS International Conference on Humanoid Robots*, pages 631–638. IEEE, 2010.

[22] Auriel Warwick and Juliette Alvin. *Music therapy for the autistic child*. Oxford University Press, 1991.

[23] Susan Young and Julia Gillen. Toward a revised understanding of young children's musical activities: Reflections from the" day in the life" project. 2007.