# Research Proposal

Huanghao Feng

Jan 2017

**Title:** Xylo-Bot: An Automated Music Teaching Robot Platform for Children with Autism and Beyond

**Author:** Huanghao Feng

**Supervisor:** Professor Mohammad Mahoor

**Degree:** Ph.D.

# Background

Autism is a general term used to describe a spectrum of complex developmental brain disorders causing qualitative impairments in social interaction and results in repetitive and stereotyped behaviors. Currently one in every 88 children in the United States are diagnosed with ASD and government statistics suggest the prevalence rate of ASD is increasing 10-17 percent annually [3]. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, turn-taking, and eye gaze attention [4]. Currently, clinical work such as Applied Behavior Analysis (ABA) [9, 10] has focused on teaching individuals with ASD appropriate social skills in an effort to make them more successful in social situations [16]. With the concern of the growing number of children diagnosed with ASD, there is a high demand for finding alternative solutions such as innovative computer technologies and/or robotics to facilitate autism therapy. Therefore, research into how to design and use modern technology that would result in clinically robust methodologies for autism intervention is vital.

In social human interaction, non-verbal facial behaviors (e.g. facial expressions, gaze direction, and head pose orientation, etc.) convey important information between individuals. For instance, during an interactive conversation, the peer may regulate their facial activities and gaze directions actively to indicate the interests or boredom. However, the majority of individuals with ASD show the lack of exploiting and understanding these cues to communicate with others. These limiting factors have made crucial difficulties for individuals with ASD to illustrate their emotions, feelings and also interact with other human beings. Studies have shown that individuals with autism are much interested to interact with machines (e.g. computers, iPad, robots, etc.) than humans [14]. In this regard, in the last decade several studies have been conducted to employ machines in therapy sessions and examine the behavioral responses of people with autism. These studies have assisted researchers to better understand, model and improve the social skills of individuals on the autism spectrum.

This proposal presents the hypothesis and potential methodology of a study that aimed to design a entertaining humanoid-robot music therapy/teaching-like sessions for capturing, modeling and enhancing the social skills of children with Autism. In particular, we mainly focus on gaze direction; joint attention; turn-taking; motor control; music pitch recognition; and music emotion understanding, investigate how the ASD and Typically Developing (TD) children employ their knowledge for interacting with the robot using music language.

## Research Questions

It is obvious that multiple physical and mental activities are involved during music play and practice. That makes it perfect to have music activity in human-robot interaction for children with autism due to the deficits they are suffering from.

Eye gaze and joint attention are required a significant amount during music instrument practice and play. Participants will be required to not only have eye contact with the music agent but also have joint attention with the music instrument at the same time. This will be one of the main activity among the entire experimental sessions, and will be evaluated at the end. Music teaching scenario creates a meaningful back and forth communication experience. This allows participants to have better understanding of turn-taking skill and be able to adapt this in their daily life.

A good thing about using a percussion instrument is that requires the player to have better motor control in order to deliver a perfect clear note with a proper strike. The strike motion need some practice that would allow the participants be able to adopt motor control technique in other activities of their life.

Listen to someone's favorite music can be cheerful and relax. However, learning instrument play can be stressful, and repetitive practice could be tedious especially in early age. This type of intra-personal feeling changes could be detectable by using affective computing. Emotion recognition and representation can be difficult for children with ASD, that makes it more eager to find out how do they feel inside during music activities. Music is also a great 'language' of representing emotions, it would also be interesting to see if children with ASD would be able to recognize the emotion behind certain piece of melody.

## Experimental Session Design

A novel interactive human-robot music teaching system design will be designed in the next 2 years. In order to make the robot play the xylophone properly, several things need to be done. First is to find a proper xylophone with correct timber; second, we have to arrange the xylophone in the proper position in front of the robot to make it visible and be reachable to play; third, finalize what are the activities should be included in experimental sessions; finally, design the intelligent music system for NAO.

**Session Design Concept:** Intervention sessions will be designed to accomplish the target questions. Due to children concentration time, total activity time in each session should be well controlled between 30 to 40 minutes and the total length of each session will be under one hour including rest time by participants request. In order to make music practice more entertaining which can keep participants attracted and focused, all tasks will be represented as a 'challenge' format. Different levels/difficulties of activities will be designed to fulfill the music teaching concept after all.

**Baseline Session:** Participants will be asked to follow all the instructions contained in all the practices from the following intervention session including single bar strike, multiple bars strike, half song play and the whole song play. We decided to choose a very popular kid's song "Twinkle Twinkle Little Star" for this specific session due to how well-known this song is in almost everyone's childhood.

**Intervention Sessions:** These sessions will be assigned to ASD group particularly and include a single strike with color hints, multiple strikes with colors hint, half-song practice and whole song practice. In this part, a customized song will be used through the rest of the sessions by the request of the individual. In the second half of this intervention session set, single/multiple strikes were also covered before the half/full song practice in order to have participants use the color matching technique during the high level music play due to a lack of professional music background knowledge. In addition, starting from session 2, a single strike warm up practice was added before the formal music practice starts. This particular practice was designed to have better motor control for ASD group so that the robot is able to recognize notes properly and deliver the concept in telling the difference between "making a sound" and "playing a musical note".

**Exit Session:** Both groups were assigned to go through the same steps as the baseline session in choice of their own songs. We would like to see the difference between the two groups in learning a beloved song by their own.

# Hardware and Software Requirements

**NAO: A Humanoid Robot** We are going to use a humanoid robot called NAO which developed by Aldebaran Robotics in France. 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. NAO's computer vision module includes facial and shape recognition units. By using the vision feature of the robot, it can see the instrument from its lower camera and be able to implement an eye-arm self-calibration system which allows the robot to have real-time micro-adjustment of its arm-joints in case of off positioning during music playing.

**Module Based Robot Music Teaching Platform:** A novel module-based robot-music teaching system will be designed in order to accomplish the mission. Three modules will be built in this intelligent system including module 1: eye-hand self-calibration micro-adjustment; module 2: joint trajectory generator; and module 3: real time performance scoring feedback.

**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 the 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 are going to use a calibration method to accurately eliminate these errors.

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, adopted in this thesis, 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.

**Module 2: Joint Trajectory Generator** Our system parses a list of hexdecimal 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.

**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 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.

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.

In order to compare the detected notes and the target notes, we adopted 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.

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

**Electrodermal activity (EDA) Based Emotion Classification Approach**
EDA has been used as an effective and reproducible electrophysiological method for investigating sympathetic nervous system function [6, 13, 12, 15]. Note that the sympathetic nervous burst changes the skin conductance, which can be traced by analyzing the EDA signals[7, 8, 11]. The Q-sensor is a convenient wireless-based EDA device with no need for cables, boxes, or skin preparation. This device can track three types of data including EDA, temperature, and acceleration at the same time [5]. It is worth mentioning that as of today, there has been no published work on emotion classification using the EDA signals collected by this dataset collected at the Georgia Institute of Technology [11].

In this research, we will employ the C-Morlet wavelet function to process the acquired EDA signals, as it has been well used for time-frequency analysis of different bio-signals and classification [2]. Using the C-Morlet mother wavelet, the real and imaginary wavelet coefficients are calculated at different scales. Then, the amplitude of these coefficients is calculated to provide the corresponding spectrogram. This spectrogram is then used as the feature space. More details can be found in [1].

# Bibliography


[1] H.M. Feng, H. Golshan and M.H Mahoor. A wavelet-based approach to emotion classification using eda signals. *Expert Systems with Applications*, 112(1).

[2] Hebb A. O. Hanrahan S. J. Nedrud J. Golshan, H. M. and M. H. Mahoor. A multiple kernel learning approach for human behavioral task classification using stn-lfp signal. In *In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, 2016.

[3] H. Huttenrauch. Fetch-and-carry with cero: observations from a long-term user study with a service robot. September 2002.

[4] Werry J. Rae P. Dickerson P. Stribling K. Dautenhahn, I and B. Ogden. Robotic playmates: Analysing interactive competencies of children with autism playing with a mobile robot. 2002.

[5] Küster D. Basedow C. Kappas, A. and P. Dente. A validation study of the affective q-sensor in different social laboratory situations. In *In 53rd Annual Meeting of the Society for Psychophysiological Research*, 2013.

[6] Kim D. H. Park W. Kwon, J. and L Kim. A wearable device for emotional recognition using facial expression and physiological response. In *In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, August 2016.

[7] A. Kylli:ainen and J. K Hietanen. Skin conductance responses to another person's gaze in children with autism. *Journal of autism and developmental disorders*, 36(4), 2006.

[8] L. Lidberg and B. G. Wallin. Sympathetic skin nerve discharges in relation to amplitude of skin resistance responses. *Psychophysiology*, 18(3), 1981.

[9] F. Michaud and S. Caron. Roball. The rolling robot. *Autonomous Robots*, 12(2), March 2002.

[10] F. Michaud and C. Thberge-Turmel. Mobile robotic toys and autism. 2002.

[11] Abowd G. Rozga A. Romero M. Clements M. Sclaroff S. Rehg, J. and H Rao. Decoding children's social behavior. In *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013.

[12] Halperin J. J. Boulu P. H. Shahani, B. T. and J Cohen. Sympathetic skin response–a method of assessing unmyelinated axon dysfunction in peripheral neuropathies. *Journal of Neurology, Neurosurgery and Psychiatry*, 47(5), 1984.

[13] Davis R. Stagg, S. D. and P Heaton. Associations between language development and skin conductance responses to faces and eye gaze in children with autism spectrum disorder. *Journal of autism and developmental disorders*, 43(10), 2013.

[14] I. Nourbakhsh T. Fong and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4), 2003.

[15] Karjalainen P. A. Koistinen A. S. Tarvainen, M. P. and M. V Valkonen-Korhonen. Principal component analysis of galvanic skin responses. In *In Engineering in Medicine and Biology Society, Proceedings of the 22nd Annual International Conference of the IEEE*, 2000.

[16] S. Wolff and S. Chess. A behavioral study of schizophrenic children. 1964.