UNIVERSITY OF DENVER

XYLO-BOT: A MUSIC PLATFORM FOR CHILDREN WITH AUTISM IN SOCIAL INTERACTIONS AND AFFECTIVE COMPUTING

By

Huanghao Feng

A DISSERTATION

Submitted to the Faculty Of the University of Denver

in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

Denver, Colorado

August 2020

UNIVERSITY OF DENVER

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

XYLO-BOT: A MUSIC PLATFORM FOR CHILDREN WITH AUTISM IN SOCIAL INTERACTIONS AND AFFECTIVE COMPUTING

Huanghao Feng

FENG, HUANGHAO (M.S., Electrical and Computer Engineering)
(August 2020)

Abstract of a comprehensive exam at the University of Denver.

Dissertation supervised by Dr. Mohammad Mahoor.
No. of pages in text 116.

Children with Autism Spectrum Disorders (ASDs) experience deficits in verbal and nonverbal communication skills including motor control, emotional facial expressions, and eye gaze/joint attention. In this manuscript, we focus on studying the feasibility and effectiveness ofusing a social robot, called NAO, and a toy instrument, xylophone at modeling and improving the social responses and behaviors of children with autism. In our investigation, we designed a autonomous social interactive music teaching system to fulfill this mission.

A novel module-based robot-music teaching system will be presented. Module 1 provides an autonomous self awareness positioning system for the robot to localize the instrument and make micro adjustment for arm joints in order to play the note bar properly. Module 2 allows the robot to be able to play any customized song of the user's request. This means that any songs which can be translated to either C-Major or a-minor key can have a well-trained person type in the hex-decimal playable score and allow the robot to be able to play it in seconds. Module 3 is designed for providing real life music teaching experience for system users. Two key features of this module are designed: music detection and smart scoring feedback. Short time Fourier transform and Levenshtein distance are adopted to fulfill the requirement which allows the robot to understand music and provide proper dosage of practice and oral

feedback to users.

A new instrument has designed in order to present better emotions from music due to the limitation of the original xylophone. This programmable new design of xylophone can provide a wider frequency range of play notes, easily switch between Major and minor keys, super easy to control and have fun with. An automated method for emotion classification in children using electrodermal activity (EDA) signals. The time-frequency analysis of the acquired raw EDAs provides a feature space based on which different emotions can be recognized. To this end, the complex Morlet (C-Morlet) wavelet function is applied on the recorded EDA signals. The dataset used in this manuscript includes a set of multimodal recordings of social and communicative behavior as well as EDA recordings of 100 children younger than 30 months old. The dataset is annotated by two experts to extract the time sequence corresponding to three main emotions including "Joy", "Boredom", and "Acceptance". Various experiments are conducted on the annotated EDA signals to classify emotions using a support vector machine (SVM) classifier. The quantitative results show that the emotion classification performance remarkably improves compared to other methods when the proposed wavelet-based features are used. By using this emotion classification, emotion engagement during sessions and feelings between different music can be detected after data analysis.

To my beloved mother,

and

to my friends.

ACKNOWLEDGMENTS

I still remember the exciting moment that I saw NAO for the first time. I still remember how I got excited when Dr. Mahoor assigned me this project. In the past two years I have been working with these special children and I cannot stop thinking of their future life after ten years or even more. Without my late friend, Mr. Masoud Bahramisharif, and his help, I was not able to start this project. I could not accomplish the goals of this project without my dear mother's support. She means everything to me. Without Mr. Hosein M Golshan's help I might not be able to finish this amazing work. I also want to thank my thesis committee members, Dr. Tim Sweeney and Dr. Matthew Rutherfor. I would thank them all in this acknowledgement.

I again would like to thank my advisor Dr. Mahoor for giving me this unique opportunity to work with NAO and supporting me through this research. Meanwhile, I want to thank the supports and help I received from Dr. Anibal Gutierrez from Florida International University and Ms. Francesca Dino from DU psychology department. I also want to appreciate all my lab mates who gave me lots of courage and motivation and to complete this research. Also many thanks to the University of Denver administrators and professors and staffs to have me here and let me study at DU. I learned a lot here. I feel lucky to be a student at DU.

Contents

List of Figures

List of Tables

List of Acronyms

2-D: Two Dimensional

3-D: Three Dimensional

ARG: Attributed Relational Graph

ASM: Active Shape Model

CMC: Cumulative Match Characteristic

EBGM: Elastic Bunch Graph Matching

FDA: Fisher Discriminant Analysis

FM: False Match

FMR: False Match Rate

FNM: False Non Match

FNMR: False Non Match Rate

FRGC: Face Recognition Grand Challenge

FRVT: Face Recognition Vendor Tests

HD: Hausdorff distance

ICP: Iterative Closest Points

LTS-HD: Least Trimmed Square-HD

MSE: Mean Square Error

PCA: Principal Component Analysis

ROC: Receiver Operating Characteristic

SFFS: Sequential Floating Forward Selection

Chapter 1

Introduction

1.1 Autism Spectrum Disorders (ASD)

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 [27]. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, and eye gaze attention [30]. Currently, clinical work such as Applied Behavior Analysis (ABA) [49, 50] has focused on teaching individuals with ASD appropriate social skills in an effort to make them more successful in social situations [86]. 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 [77]. 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 thesis presents the methodology and results of a study that aimed to design a humanoid-robot therapy sessions for capturing, modeling and enhancing the social skills of children with Autism. In particular we mainly focus on gaze direction and joint attention modeling and analysis and investigate how the ASD and Typically Developing (TD) children employ their gaze for interacting with the robot. In the following section, we have a brief introduction of the existing assistive robots in the following section and how they have been used in autism applications.

1.2 Socially Assistive Robotics

Socially Assistive Robotics (SAR) can be considered as the intersection of Assistive Robotics (AR) and Socially Interactive Robotics (SIR), which has referred to robots that assist human with physical deficits and also can provide certain terms of social interaction abilities [17]. SAR contains all properties of SIR described in [77], and also a few additional attributes such as: 1) user populations (different groups of users, i.e. elders; individuals with physical impairments; kids diagnosed with ASD; students); 2) social skills (i.e. speech ability; gestures movement); 3) objective tasks (i.e. tutoring; physical therapy; daily life assistance); 4) role of the robot (depends on the task the robot has been assigned for) [17]. Companion robots [31] is one type of SAR that are widely used for elderly people for health care supports. Research shows that this type of social robots can reduce stress and depression of individuals in elderly stage [16]. Service social robots are able to accomplish a variety of tasks for individuals with physical impairments [27]. Studies have shown that SAR can be used in therapy sessions for those individuals who suffer from cognitive and behavioral disorders (e.g. Autism). SAR provides an efficient helpful medium to teach certain types of skills to these groups of individuals [30, 49, 50]. Nowadays, there are very few companies that have been designing and producing socially assistive robots. The majority of existing SARs are not commercialized yet and because of being expensive and not well-designed user interfaces, they are mostly used forthe research purposes. Honda, Aldebaran Robotics and Hanson Robokind are the top leading companies that are currently producing humanoid robots.

Ideally socially assistive robots can have fully automated systems to detect and express social behaviors while interacting with humans. Some of the existing robot-human interfaces are semi-autonomous and they can recognize some basic biometrics (e.g. visual and audio commands of the user) and behavioral response. Besides, the majority of existing robots are very complicated to work with. Therefore in the last couple of years several companies have started to make these robots more user-friendly and responsive to both the user need and the potential caregiver commands [17].

Intelligent SARs aim to have the capability to recognize visual or audio commands, objects, and specific human gestures. Some of these robots have the ability of detect human face or basic facial expressions. For instance, ASIMO, a robot developed by Honda, it has a sensor for detecting the movements of multiple objects by using visual information captured from two cameras on its head. Plus its “eyes” can measure the distance of the objects from the robot [56]. Another example is from Aldebaran Robotics which designs small size humanoid robots, called NAO. NAO robot has two cameras attached that are used to capture single images and video sequences. This capturing module enables NAO to see the different sides of an object and recognize it for future use. Furthermore, NAO has a remarkable capability of recognizing faces and detecting moving objects.

Both of the aforementioned robots have speech recognition system. They can interpret voice commands to accomplish a certain set of tasks which have been pre-programmed in the system. NAO is able to identify words for running specific commands. However ASIMO is able to distinguish between voices and other sounds. This feature empowers ASIMO to perceive the direction of human's speaker or recognize other companion robots by tracking their voice [3]. These robots can also speak in many different languages. For example, NAO can speak in English, French, Chinese, Japanese and other languages up to more than ten languages. This feature gives the robot a great social communication functionality to interact with humans from all over the world.

1.2.1 Socially Assistive Robots for Autism Therapy

Socially assistive robots are emerging technologies in the field of robotics that aim to utilize social robots to increase engagement of users as communicating with robots, and elicit novel social behaviors through their interaction. One of the goal in SAR is to use social robots either individually or in conjunction with caregivers to improve social skills of individuals who have social behavioral deficits. One of the early applications of SAR is autism rehabilitation. As mentioned before, autism is a spectrum of complex developmental brain disorders causing qualitative impairments in social interaction. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, and gaze regulation. These skill deficits often pose problems in the individual's ability to establish and maintain social relationships and may lead to anxiety surrounding social contexts and behaviors [86]. Unfortunately there is no single accepted intervention, treatment, or known cure for individuals with ASD.

Recent research suggests that children with autism exhibit certain positive social behaviors when interacting with robots compared to their peers that do not interact with robots [59, 80, 82, 17, 77]. These positive behaviors include showing emotional facial expressions (e.g., smiling), gesture imitation, and eye gaze attention. Studies show that these behaviors are rare in children with autism but evidence suggests that robots trigger children to demonstrate such behaviors. These investigations propose that interaction with robots may be a promising approach for rehabilitation of children with ASD.

There are several research groups that investigated the response of children with autism to both humanoid robots and non-humanoid toy-like robots in the hope that these systems will be useful for understanding affective, communicative, and social differences seen in individuals with ASD (see Diehl et al., [77]), and to utilize robotic systems to develop novel interventions and enhance existing treatments for children with ASD [56, 3, 4]. Mazzei et al. [15], for example, designed the robot "FACE" to realistically show the details of human facial expressions. A combination of hardware, wearable devices, and software algorithms measured subject's affective states (e.g., eye gaze attention, facial expressions,vital signals, skin temperature and EDA signals), were used for controlling the robot reactions and responses.

Reviewing the literature in SAR [17, 77] shows that there are surprisingly very few studies that used an autonomous robot to model, teach or practice the social skills of individuals with autism. Amongst, teaching how to regulate eye-gaze attention, perceiving and expressing emotional facial expressions are the most important ones. Designing robust interactive games and employing a reliable social robot that can sense users' socioemotional behaviors and can respond to emotions through facial expressions or speech is an interesting area of research. In addition, the therapeutic applications of social robots impose conditions on the robot's requirements, feedback model and user interface. In other words, the robot that aims for autism therapy may not be directly used for depression treatment and hence every SAR application requires its own attention, research, and development Only a few adaptive robot-based interaction settings have been designed and employed for communication with children with ASD. Proximity-based closed-loop robotic interaction [11], haptic interaction [26], and adaptive game interactions based on affective cues inferred from physiological signals [72] are some of these studies. Although all of these studies were conducted to analyze the functionality of robots for socially interacting with individuals with ASD, these paradigms were limitedly explored and focused on their core deficits (i.e., Facial expression, eye gaze and joint attention skills). Bekele and colleagues [32] studied the development and application of a humanoid robotic system capable of intelligently administering joint attention prompts and adaptively responding based on within system measurements of gaze and attention. They found out that preschool children with ASD have more frequent eye contact toward the humanoid robot agent, and also more accurate respond in joint attention stimulations. This suggests that robotic systems have the enhancements for successfully improve the coordinated attention in kids with ASD.

Considering the existing SAR system and the major social deficits that individuals with autism may have, we have designed and conducted robot-based therapeutic sessions that are focused on different aspects of social skills of children with autism. In this thesis we employed NAO which can be autonomously communicate with the children. We conducted two different designs to examine the music social skills of children with autism and provide feedbacks to improve their behavioral responses.

1.3 Music Therapy for ASD

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 [89]. 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 [63]. 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 [8]. 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 [84, 66, 14, 19]. 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 [20]. 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 [85].

1.4 Contributions

The major contributions of this manuscript are as follows:

- Developing a wavelet-based approach to event based emotion classification using Electrodermal activity signal from early age children. In our work, the dataset is first annotated to label perceived emotions (e.g., Acceptance, Joy, Boredom) expressed by each subject. Afterwards, we utilize the continuous wavelet transform to develop a new feature space for classification purposes. Using the complex Morlet function, the wavelet coefficients of the EDA signal at different scales are calculated, providing a more detailed representation of the input signal. The performance of the proposed feature space on emotion classification task is evaluated using the canonical support vector machine (SVM) classifier with different types of kernel functions as well as the K-nearest neighborhood (KNN) classifier. And this method will be implemented to music teaching/playing therapy intervention for better understanding emotion engagement.

- Developing an autonomous social interactive robot music teaching system for children with autism. a novel module-based robot-music teaching system will be presented. Three modules have been built in this intelligent system including module 1: eye-hand self-calibration micro-adjustment in order to prevent minor change of relative position between musical instrument and robot; module 2: joint trajectory generator in order to play any meaningful customized melody; and module 3: real time performance scoring feedback using short time Fourier transform and Levenshtein distance in order to provide a autonomous real-time music learning experience.

- Designing a new instrument call X-Elophone, which allows user to create more types of melody. This new design brings more possibilities for young children who is willing to learn music and music emotion understanding.

- Designing a set of music teaching session using a humanoid social robot NAO to deliver a unique music teaching experience to kids with autism. After intervention sessions, participants will be able to have better eye-gaze/joint attention performance, better motor control skills and better music understanding ability. By using newly designed X-Elophone, participants would learn music emotions.

1.5 Orgizaition

This manuscript is organized as follows: In chapter 2, we present related work related to autism spectrum disorders, emotions classification, music therapy in autism treatment, social robots in autism therapy. Chapter 3 explains our approach of designing the autonomous social interactive robot music teaching system. Chapter 4 proposed a wavelet-based feature extraction approach for emotion classification as a pre-study for music interaction emotion recognition. Chapter 5 illustrates the experimental session design and partial experimental results. Discussion and conclusion will be presented in Chapter 6 and Chapter 7 demonstrates the new design of xylophone.

Chapter 2

Related Works

2.1 Autism

Individuals with autism spectrum disorder experience verbal and nonverbal communication impairments, including motor control, emotional facial expressions, and eye gaze attention. Oftentimes, individuals with high-functioning autism have deficits in different areas, such as (1) language delay, (2) difficulty in having empathy with their peer and understanding others emotions (i.e. facial expressions recognition.), and more remarkably (3) joint attention (i.e. eye contact and eye gaze attention). Autism is a disorder that appears in infancy [45]. Although there is no single accepted intervention, treatment, or known cure for ASDs, these individual will have more successful treatment if ASD is diagnosed in early stages. At the first glance at the individual with autism, you may not notice anything odd, however after trying to talk to her/him, you will understand something is definitely not right [62]. S/He may not make eye contact with you and avoid your gaze and fidget, rock her/his body and bang her/his head against the wall [62]. In early 1990s, researchers in the University of California at San Diego aimed to find out the connections between autism and nervous system (i.e. mirror neurons). Mirror neuron [62] is a neuron that is activated either when a human acts an action or observes the same action performed by others. As these neurons are involved with the abilities such as empathy and perception of other individual's intentions or emotions, they came up with malfunctioning of mirror neuron in individuals with ASD [62]. There are several studies that focus on the neurological deficits of individuals with autism and studying on their brain activities. Figure 2-1 demonstrates the areas in the brain that causes the reduce mirror neuron activities in individuals with autism.

Individuals with autism might also have several other unusual social developmental behaviors that may appear in infancy or childhood. For instance children with autism show less attention to social stimuli (e.g. facial expressions, joint attention), and respond less when calling their names. Compared with typically developing children, older children or adults with autism can read facial expressions less effectively and recognize emotions behind specific facial expressions or the tone of voice with difficulties [61]. In contrast to TD individuals, children with autism (i.e. high-functioning, Asperger syndrome) may be overwhelmed with social signals such as facial behaviors and expression and complexity of them and they suffer from interacting with other individuals, therefore they would prefer to be alone. That is why it would be difficult for individuals with autism to maintain social interaction with others [5].

In order to diagnose and asses the aspects and score the social skill level of an individual with autism, several protocols are available. One of the commercially available protocols is called Autism Diagnostic Observation Schedule (ADOS) [9] that consists of four modules and several structured tasks that are used to measure the social interaction levels of the subject and examiner. We are inspired by ADOS in designing our intervention protocols later described in Chapter 4. Hence, we briefly review ADOS in the next section.

2.1.1 Turn-Taking

Turn-taking is a type of organization in conversation and discourse where participants speak one at a time in alternating turns. In practice, it involves processes for constructing contributions, responding to previous comments, and transitioning to a different speaker, using a variety of linguistic and non-linguistic cues.[87]

While the structure is generally universal,[60] that is, overlapping talk is generally avoided and silence between turns is minimized, turn-taking conventions vary by culture and community.[80] Conventions vary in many ways, such as how turns are distributed, how transitions are signaled, or how long is the average gap between turns.

In many contexts, conversation turns are a valuable means to participate in social life and have been subject to competition.[82] It is often thought that turn-taking strategies differ by gender; consequently, turn-taking has been a topic of intense examination in gender studies. While early studies supported gendered stereotypes, such as men interrupting more than women and women talking more than men,[17] recent research has found mixed evidence of gender-specific conversational strategies, and few overarching patterns have emerged.[77]

2.1.2 Motor control

Motor control is the systematic regulation of movement in organisms that possess a nervous system. Motor control includes movement functions which can be attributed to reflex,[87]. Motor control as a field of study is primarily a sub-discipline of psychology or neurology.

Recent psychological theories of motor control present it as a process by which humans and animals use their brain/cognition to activate and coordinate the muscles and limbs involved in the performance of a motor skill. From this mixed psychological perspective, motor control is fundamentally the integration of sensory information, both about the world and the current state of the body, to determine the appropriate set of muscle forces and joint activations to generate some desired movement or action. This process requires cooperative interaction between the central nervous system and the musculoskeletal system, and is thus a problem of information processing, coordination, mechanics, physics, and cognition.[60, 80] Successful motor control is crucial to interacting with the world, not only determining action capabilities, but regulating balance and stability as well.

The organization and production of movement is a complex problem, so the study of motor control has been approached from a wide range of disciplines, including psychology, cognitive science, biomechanics and neuroscience. While the modern study of motor control is an increasingly interdisciplinary field, research questions have historically been defined as either physiological or psychological, depending on whether the focus is on physical and biological properties, or organizational and structural rules.[82] Areas of study related to motor control are motor coordination, motor learning, signal processing, and perceptual control theory.

2.2 Human Robot Interaction in Autism

Children with ASD experience deficits in appropriate verbal and non-verbal communication skills including motor control, emotional facial expressions, eye-gaze attention, and joint attention. Many studies have been conducted to identify therapeutic methods that can benefit children with ASD [52]. However, only a few groups used humanoid robots for teaching or practicing social communication skills [53, 54, 55, 56, 57, 58, and 59]. For some of the social behaviors, such as eye contact, joint attention, facial expressions recognition, that are rarely seen in interactions of children ASD, several evidence suggest that robots can trigger them more effectively than human [78]. Researchers observed that individuals with ASD have more interest toward a robot therapeutic partner than a human. In most cases participants showed better speech and movement imitation compared with response to a human partner [79]. Although a recent case study [52] which was done by Ricks (2010) suggests that this approach might have clinical utility, still this area is obviously in its infancy. Studies have shown that positive feedback from the robot on the participants' performance is an effective way to encourage children with ASD to communicate more [52]. Other studies have also examined the use of affect recognition (e.g. emotional state, arousal level) based on psychophysiological responses to modify the behaviors during a robotic game. However, there is limited information on the utility of humanoid robots' positive feedback in interventions for individuals with ASD.

2.2.1 Interactive and Therapeutic Robots Designs for Autism

Different types of robots have been used in autism research for various purposes. Some researchers have been attempting to utilize a realistic human appearance [56], while others have created robots with very mechanical forms [54], and others have developed robots with a cartoonish or animal form [58]. Generally speaking different categories of robot that have been used for autism research can be placed either into Non-Humanoid and Humanoid robots group [52], which will be explained in the following sections.

Non-Humanoid Robots

Non-humanoid robots are those robots which do not have the same body joint and facial appearance as human does. It contains those animal like, cartoonish, or non-human like appearances. These robots have been used by several researchers in the last two decades. This category of robots is generally easier to design and develop and less expensive, therefore, several of initial robot-human interaction for individuals with ASD was conducted by non-humanoid robots. The bubble-blowing robot at USC (while children approached it, the robot will node head make voice or blow bubble from lower part of robot body), for instance, was not a human form robot and can be built simply [53]. Another non-humanoid robot used by researchers from University of Hertfordshire called Labo-1 [54], which can play tag games (tip you're it or tig), with children. (In the game, several children play with the robot together, the robot uses its heat sensor to approach kids as a type of interaction.) In Yale University, researchers were using a mobile robotic dinosaur named Pleo who can show emotions and desires by using its sounds and body movements [55]. Children in the clinic have been helped by Pleo's pet-like appearance, expressiveness, and versatility. The reason why researchers using non-humanoid robots is that they found out that when children with ASD see humans, they usually will choose to avoid and not to interact with them. On the contrary, an animal shape or toy shape robot would be easier for kids to engage with and have a better interaction.

Humanoid Robots

Humanoid robots generally provide the human-like appearance and consist of body parts such as humanoid head, body and arms. Advanced humanoid robot would be able to move different parts of it body to walk or dance (NAO), some of the humanoid robot also has the capability to show facial expressions (e.g. ZENO). This type of robot unlike non-humanoid robot, they have the ability to accomplish more complicated social communication tasks than non-humanoid robot, but those tasks will be less complicated than human-human interaction. This capability can help us to design interaction sessions and therapeutic sessions for children with autism and assist them with improving their social behaviors. Robins from University of Hertfordshire, who is one of the pioneers which employed a study to evaluate the importance of robot's appearance for autism research. A doll-like robot called Robota were asked to interact with children with autism [56]. This example shows that children appeared to be more interested in interaction with less-human like robots. Researchers conclude that children with ASD would prefer a simple non-complexity and less details of human but still hold the humanoid form. So, a robot called KASPAR has been developed by Robins to fit this design criteria [57]. Similar conclusions have been made by researchers at the National Institute of Information and Communications Technology (NIICT) in Japan. They found out that when kids with ASD have interaction with their designed robot called Infanoid, the children tend to pay more attention on the mechanical parts of the robot's body than communicating with the robot itself [52]. A small soft snowman-shaped robot, called Keepon, was designed to represent as a simple, repeatable, mechanical robot regarding the reason mentioned above [58]. Keepon can express its emotions conveyed by shaking, rocking, and bobbing up and down which can be used as a super fun toy companion for kids with ASD. Another humanoid robot which has been designed by researchers at the University of Pisa, is known as FACE. The purpose of their project is to create a robot as realistic as possible to a human face for evaluating how human react as the FACE displays different expressions [59]. (During the sessions, child (IQ around 85) with autism did not show any interest in FACE at the beginning. However, with verbal suggestion, kid replied the expression by using a word "damsel" which is from a fairy tale, though the FACE showing a sad expression on it.) This study suggested that by using FACE, it is possible to extend emotional recognition skill to children with autism. In the last few years several different types of non-humanoid and humanoid robots have been used for autism therapeutics sessions that we will discuss about them in the next session.

2.2.2 Different Therapeutic approaches for Individuals with ASD

As explained in Chapter 2, different individuals with autism might suffer from various types of social or developmental behavior. Therefore in order to have an effective therapeutic intervention setting we need to focus on various tasks and treatments. Bellow we will provide different intervention aspects that majority of children with ASD may suffer from.

Self-Initiated Interactions

The difficulty for initiating a social conversation or interaction is one of the impaired social skills of children with ASD. This problem may represent as difficulty for conveying what they want, and why they want it. For example, when a child in early age who wants to urinate, he might have to ask for parent's help, rather than hold it there or let it be. Clinicians try to encourage those kids to ask to play certain toys and a reward will be given after they did it. Instead of human therapists, researcher extended this idea using robots to encourage the children to engage the robot proactively. The robot has built at USC [53] which has a large button on its back, and it was programmed to encourage social interaction with children. For example, the robot will nod its head and make a sound to encourage the kid to approach it; when the kid walk away, it will move its head down and make sad kind of sound to imply the child and ask him/her to come closer to the robot. If the child presses that button on the robot, it will blow bubbles and turn. In this study, one hundred minutes experiments have been recorded, three different conditions have been considered which are the time kids spent near 1) the wall;2) the parent; and 3) behind the robot. Kids have been separated in two groups: 'Group A' (children like the robot) and 'Group B' (children do not like the robot) total number of eight children with ASD. The result shows that the Group A spent more than 60% of the time playing with the robot, and Group B spent more than 50% of the time showing the negative reaction (i.e. go away from robot, play with himself) from avoiding the robot. This study might not be very convincing because it is totally free play with the robot; the experimental settings haven't kept the same, and the limited numbers of participants. Also without control group like typically developing they could not compare the differences of ASD and TD children, within the robot games. However, it shows the capability of encourage children to communication with robot, and lead the conversation [53].

Turn-Taking Activities

At the University of Hertfordshire and the University of California, researchers have built small mobile robots that focused on helping children with ASD in turn-taking behaviors [54 and 53]. It is easily to found out that children with ASD have a hard time to allow their conversation partner to participate. The researchers try to use these robots to help them become accustomed to waiting for responses after they say or do something. Labo-1 built by the University of Hertfordshire, which can play a game called tag with children. This game will forces them to alternate between engaging and avoiding the robot [54]. Labo-1 is a mobile platform that has an AI system resembled in a sturdy flat-topped buggy. Children have been allowed to freely play with Labo-1 as a teacher was deciding about the how to switch between different games/sessions considering children appears (i.e. difference reactions of children like tired or less interested into robot). From their initial trials, children were in overall happy to play with robot. At the beginning of the game, the robot showed several simple behavior patterns, such as going forward and backward. Kids showed positive response to these behaviors and enjoy to keep playing with Labo-1. Children were also enjoyed interacting with the robot while it used a feature called 'heat following behavior', they moved away from the robot and see if the robot can follow or not.

There were five trials in total, three of them lasted around four minutes, and the remaining two had duration of approximately fourteen minutes. Researchers realized that the issues that may cause this difference might be related to the levels of the children's functioning. Since children are not in complete control the robot's actions, and children's response were totally different, some of them either walked or crawled around the room, some of them just simply lay on the floor to interact with robot only use arm movement [54]. During the interactions, it is obvious to notice that robot need more advanced behaviors to be developed and the scenario should have more control for data analysis and get more convincing results. Also the functioning level become another important element that need to be considered.

Expression/Emotion Recognition and Imitation

Another import difficulty of individuals with ASD is to recognize the expressions and emotions, besides appropriately imitating them. Studies show that kids with ASD have a hard time recognizing emotions and facial expressions. It would be difficult for them to deliver their emotions through their faces actions. Researchers pointed out that to kids with ASD, such emotion type information which contained faces or eye contact can result overwhelming or sensory overload. For example, a person could smile twice, and the child with ASD might pick two entire different expressions from those two smiles. Robot can provide more constancy repeatable consistent behaviors than human does, and it would be a better way to teach children expressions and emotions. KASPAR, a child-sized doll like robot which has a silicon-rubber face on it, developed by the University of Hertfordshire has been used to show bodily expressions by move head and arms. KASPAR was operated via wireless remotely, sessions are designed to allow the children to have free play interaction with robot. Some behaviors had been pre-programmed in the robot, those behaviors allows KASPAR to show several facial expressions, hand waving and drumming on the tambourine on its legs to express different emotions. During the interaction, three types of touch using the hands had been identified: grasping (different tension levels), stroking, and poking. The forces of touching can be detected by the tactile sensors equipped different places of KASPAR's arms, hands, face, and shoulders. By detecting different levels of touching, KASPAR would provide different movements or expression to tell the children the emotions or feeling of it. Emotion and facial expressions recognition could be taught via these outputs KASPAR given. The limitation of this study is very few numbers of children (five children in total) had participated in this study. Besides limited facial expressions (happiness, displeasure, surprise etc.) have employed in the robot system, and those expressions are hard to distinguish by the images they provided. There is no verbal communication between kids and robot, which is another weakness of this study [57]. FACE is a robot designed at the University of Pisa point to closely approximate a real human face and show the detail facial expressions. Children would be asked to imitate those expressions to practice their ability in facial expression recognition and imitation. Certain scenarios (i.e. 1) facial expression association: a) facial matching, b) emotion labeling; 2) emotion contextualization) would be given to kids and ask them to pick up an appropriate emotional expression for FACE to make. Several experiments have been implemented to help the children to generalize the information they learn from the therapy sessions. After practicing with FACE, the children were tested using the Childhood Autism Rating Scale and the results showed that while working with FACE, the ability of categories emotions and expressions for all kids (total number of 4 kids) have been improved. Also, researchers found out that those children can imitate facial expressions from FACE better than from humans, and it will be easier for therapist because of the automate repeatable of robots process. However, still very limited number of kids participated in the study that made the results somehow not quite untenable [59].

Joint/ Eye-gaze Attention

One of the major deficiencies of individuals with ASD is the lack of continuous concentration on the same object [33]. Joint attention is a concept of remain focus on specific things. Helping children with ASD on this aspect, would also help them to achieve the success in learning other skills. Keep practicing joint attention would give them more understandings of what others are aware of them, what they are aware of other and they both aware of same object. For this purpose, researchers from National Institute of Information and Communications Technology in Japan developed the Keepon robot. For seeking the possible responses of using interpersonal communication, both ASD and TD kids have been recruited in the study. A silicon-rubber made yellow snowman like body covered above the mechanical parts of Keepon, with two eyes on the upper part of the robot, and a nose (microphone embedded) in between. Lower part which is the belly of Keepon can easily deform whenever it changes posture and when people touch it [58]. With four degrees of freedom ($\pm$40 degrees of nodding, $\pm$180 degrees of shaking, $\pm$25 degrees of rocking, and bobbing with 15 mm stroke), Keepon is able to perform two action mode: Attentive action and Emotive action. In attentive mode, Keepon will orient its face/body to a certain object around it, two CCD cameras in its eyes would be able to making eye contact and joint attention with the target; in emotive mode, Keepon will still its attention in a certain direction, and rocks its body up and down or left and right to express its emotions like pleasure and excitement. In both modes, Keepon will also making little sounds to drag the attention of people around it or give some feedback when people touches it or grabs it [58]. There are two operation mode to control Keepon, either automatic mode or manual mode. In automation mode, locations of a human face, a toy with a predetermined color, and an optically moving region would be detected. An Attention Map are written inside of Keepon, it orients its body (eye gaze) to most salient point on the Attention Map; its emotional expression is determined by the type (face/toy/motion) and the saliency value of the point of interest. In manual mode, based on the onboard cameras and listens to the sounds captured by the onboard microphone, a person can easily control Keepon via a remote computer. The operator only need to click the interest on the panoramic map to displays emotional expression on Keepon [58]. After more than a year and half (over 500 child-sessions), this research provides some interesting results. Children who have autism and PDD, they usually have difficulty with communicating with others, which however were able to approach Keepon with security and curiosity, and had a good time with it. Some of the kids even learned how to share their pleasure with other people which extended the dyadic interaction to triadic interaction. Different children have different style with communicating with Keepon, based on those different reactions researchers might predict different personality of those children [58]. This study shows some promising conclusions, but still cannot provide statistic results to readers. Overall, the experiment settings have been considerate be thoughtful, though the sessions kept in free play mode. Good amount of partisans enrolled in the study, which makes the conclusion more convincible. The aim of the study has been fully illustrated during sessions, joint attention has ran through both action modes and provided a good feedback from the partisans. Keepon's voice needs to be improved, not just making simple noise, but also have a complete conversation would be better. More statistic results needs to be analyzed in the future to compare both children with ASD and TD kids.

2.2.3 Using NAO in Autism

NAO is a multifunctional humanoid robot that was developed by Aldebaran Robotics and as it has capabilities such as making different gesture, moving different arm and leg movement and hear orientations, It has been used for different human-robot interaction sessions. In this section we will talk about the existing interactions sessions that were conducted by NAO and later in the next chapter we will explain about our therapy sessions and designed game based on NAO for children with ASD. In University of Teknologi MARA, NAO was used to conduct seven interactions modules for interacting kids with autism. Each module lasts four minutes, and one minute break was provided between two sessions. Different interaction tasks have been contained in those modules (i.e. static interaction, joint attention, basic language skills).

Frequency of child looking at robot and duration of each occurrence of interaction has been reported. After all, they concluded that those 7 modules can be applied to develop human-robot integration therapy sessions for children with autism [80]. Same year, these researchers use 5 of those 7 modules did a case study, with the same setting, they recruited one high-functioning (with IQ 107) to complete those 5 tasks. They aimed to discover whether that child can provide a better exposure behavior with robot compared with the activity in the class. After running the five tasks for only one instance, they concluded that the child behavior have been improved significantly with robot than in the class, they also suggested that humanoid robot NAO can be used as a major platform to support and initiate interaction with children with ASD [81]. After this case study, they recruited other 5 children with ASD (low IQ, average around 50) and did the same experimental interaction sessions with them. For out of five children showed better performance during robot interaction compared with daily in-class performance [82]. Further research have been done by this group, they added emotion recognition module into the interaction sessions. Five body gesture emotions (hungry, happy, mad, scared, and hug/love) have been implemented in the program. Two boys have been enrolled in this study, and after finished the session, researchers pointed out that NAO has the potential capability to teach head and bod posture related to social emotions for children with autism, without provided any statistical analysis only based on observations [83]. This group has been initiated working with NAO for autism therapeutic session and implementing and compared different scenarios based on NAO. Reviewing the existing papers demonstrate that the number of participants and interaction sessions for these studies are very limited.

They have used only one session for each subject. Therefore they could not analyze the social responses of individuals with ASD statistically. In our study we employ NAO since it has several functionalities that are embedded in it (e.g. text-to-speech, tactical sensor, face recognition, voice recognition etc.). This would help us to build a social communicative tasks for human-robot interaction. Based on the size of the robot and the friendly appearance of the robot we design, conduct and analyze the gaze related responses of ASD individual and compare it with TD control group. The details of our experiment and the results will be discussed in Chapter 4.

2.3 Emotion Classification

2.3.1 Electrodermal activity (EDA)

Emotion is an intense mental experience often manifested by rapid heartbeat, breathing, sweating, and facial expressions. Emotion recognition from these physiological signals is a challenging problem with interesting applications such as developing wearable assistive devices and smart human-computer interfaces. This paper presents an automated method for emotion classification in children using electrodermal activity (EDA) signals. The time-frequency analysis of the acquired raw EDAs provides a feature space based on which different emotions can be recognized. To this end, the complex Morlet (C-Morlet) wavelet function is applied on the recorded EDA signals. The database used in this paper includes a set of multimodal recordings of social and communicative behavior as well as EDA recordings of 100 children younger than 30 months old. The dataset is annotated by two experts to extract the time sequence corresponding to three main emotions including “Joy”, “Boredom”, and “Acceptance”. The annotation process is performed considering the synchronicity between the children’s facial expressions and the EDA time sequences. Various experiments are conducted on the annotated EDA signals to classify emotions using a support vector machine (SVM) classifier. The quantitative results show that the emotion classification performance remarkably improves compared to other methods when the proposed wavelet-based features are used.

EDA has been used as an effective and reproducible electrophysiological method for investigating sympathetic nervous system function [35, 73, 70, 81]. Note that the sympathetic nervous burst changes the skin conductance, which can be traced by analyzing the EDA signals[36, 40, 64]. 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 [33]. 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 [64].

EDA signals are nonstationary and noisy; hence, wavelet-based analysis of EDA signals has been considered in the literature [75, 37] either as a pre-processing step or a feature extraction approach for emotion classification. [75] used a set of wavelet coefficients representing EDA features together with heart rate signal to increase the percentage of correct classifications of emotional states and provide clearer relationships among the physiological response and arousal and valence. [71] used a feature space based on the discrete wavelet transform (DWT) of the EDA signal to distinguish subjects suffering social anxiety disorder (SAD) and a control group. Using MLP and DWT features, they achieved a classification accuracy of 85

2.3.2 Classification Applications

Physiological responses have been identified as reliable indicators of human emotional and cognitive states. This section is dedicated to review some existing methods used for human emotion recognition based on various physiological responses, such as facial expression and other types of bio-signals.

A wearable glass device was designed by [35] to measure both electrodermal activity (EDA) and photoplethysmogram data for emotion recognition purposes. A built-in camera was also used in this device for capturing partial facial expression from the eye and nose area. This approach obtains remarkable performance in facial expression recognition in the subject-dependent cases. However, for subject-independent cases, it results in different accuracies across different types of emotions, which is an undesirable feature.

Several emotion classification methods have been presented in the literature using different bio-signals [43, 38, 69, 88]. Due to the variety of the signals used in these methods, different approaches have been designed to comply with their specific characteristics. Analysis of variance (ANOVA) and linear regression [69] are the commonly used methods to extract features from bio-signals and to recognize different emotional states. These methods are based on the assumption of a linear relationship between the recorded signals and emotional states. A fuzzy-based classification method [43] has been used in to transform EDA and facial electromyography (EMG) to valence and arousal states. These states were then used to classify different emotions.

Artificial neural networks (ANN) have also been applied for emotion classification tasks based on physiological responses. [42] developed a multilayer perceptron network (MLP) architecture capable of recognizing five emotions using various features from Electrocardiography (ECG) and EDA signals, and obtained very accurate classification performance. [54] employed K-nearest neighborhood and discriminant function analysis to perform the emotion classification task using different features extracted from the EDA signals, body temperature and heart rate.

Support Vector Machine (SVM) is a well-known supervised learning algorithm that has extensively been used for pattern classification and regression [13]. The SVM classifier tends to separate dataset by drawing an optimal hyperplane between classes such that the margin between them becomes maximum. The samples of each class that are located within the margin are called support vectors and play the main role in calculating the parameters of the hyperplanes between the corresponding classes. Machine learning algorithms such as SVM, linear discriminant analysis

(LDA), and classification and regression tree (CART) have been employed for emotion classification purposes. For instance, in several works including [68, 29], the authors combined various types of bio-signals such as ECG, skin temperature (SKT), HR, and Photoplethysmogram (PPG) for emotion classification purposes. [2] proposed unsupervised clustering methods for emotion recognition. Their method benefited from several features obtained from different body responses such as SC, HR, and EMG. They showed that only a few statistical features such as the mean and standard deviation of the data can be relevant identifiers for defining different clusters.

To the best of our knowledge there are a few works [38, 55] that have studied and compared different automated classification techniques for emotion recognition of children using EDA signals. This motivated us to conduct this study using an existing dataset, which concentrates on emotion classification of children based on the relationship between their facial expressions and the collected EDA signals.

2.4 Music Therapy

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 [67, 7]. Nowadays, at least 12% of all treatment of individuals with autism consist of music-based therapies [6]. Specifically, teaching and playing music to children with autism spectrum disorders (ASD) in therapy sessions have shown great impact for improving social communication skills [41]. Recorded music or human played back music are used in single and multiple subjects' intervention session from many studies [6, 12]. Different social skills are targeted and reported (i.e. eye-gaze attention, joint attention and turn-taking activities) in using music-based therapy sessions [74, 34]. Noted that improving gross and fine motor skills for ASD through music interventions is a missing part in this field of studies [6].

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 [19, 51, 47], 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 [58, 78, 79]. 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 [52]. Although limited research has found in such area especially using bio-signals for emotion recognition for ASD and TD kids [18] 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 [18], 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.

Chapter 3

Xylo-Bot: An Interactive Music Teaching System

As mentioned above, music can be considered as a effective method for emotion and non-verbal communication. Individuals with ASD are interested in interacting with social robot. Therefore, implement music therapy like intervention sessions using humanoid social robot become possible. A novel interactive human-robot music teaching system design is presented in this chapter. Hardware and software design will be discussed in following sections including experiment room setup, robot selection, instrument accessories and platform design.

In order to make the robot play the xylophone properly and be able to conduct a music-based social interaction scenario, several things needed 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; after that, a set of challenge based experimental sessions require to be constructed including baseline session, intervention sessions and exit session with various level of activities; finally, the module based interactive music teaching system will be designed and programmed which can be implemented into the experimental sessions.

3.1 Hardware Selection and Design

3.1.1 NAO: A Humanoid Robot

NAO, a humanoid robot were selected in the current research which sailed by SoftBank company. NAO is 58 cm (23 inches) tall, with 25 degrees of freedom. Most of the human body movement can be performed by this robot. 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 3.1, these utilities are located in the middle of forehead and mouth area. 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 before music play.

The robot arms have a length of approximately 31 cm. Position feedback sensors

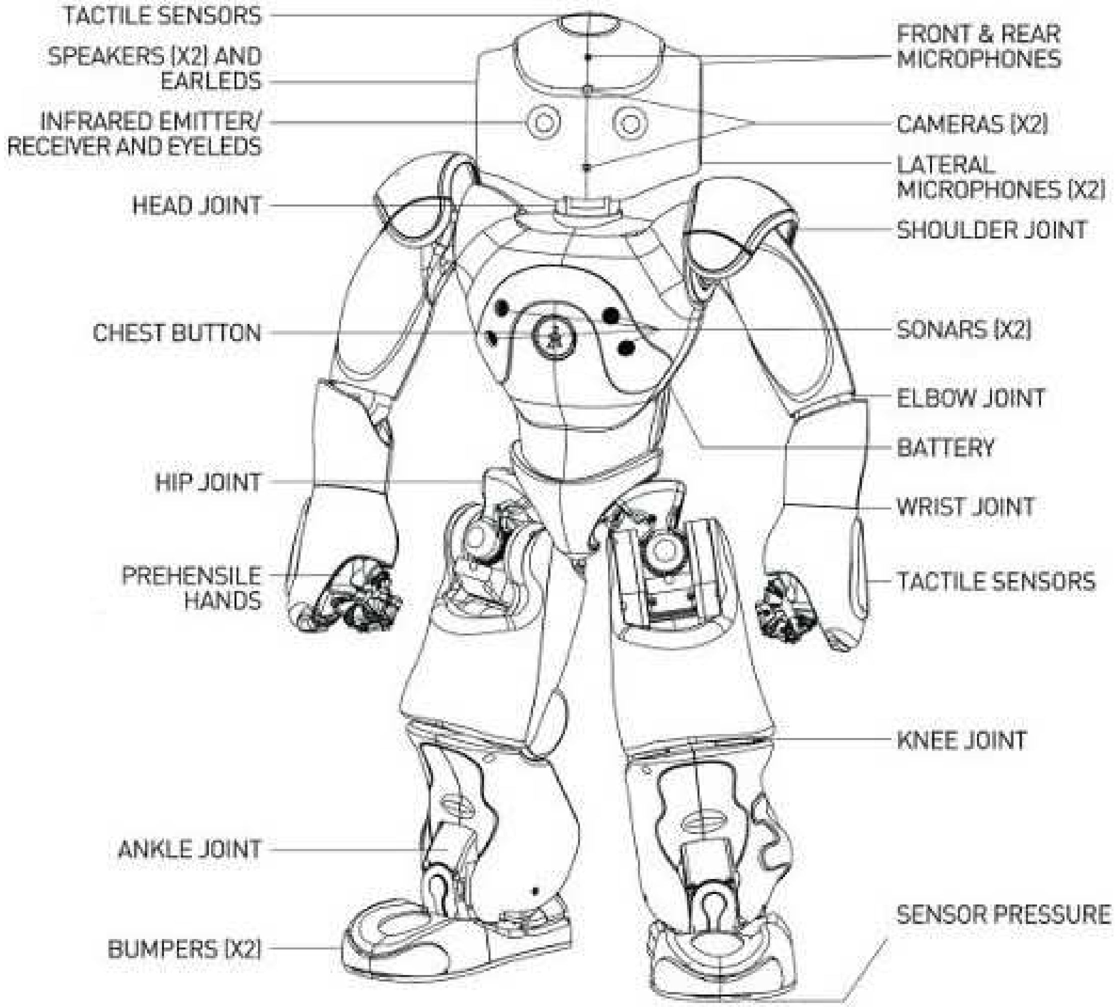


Figure 3.1: *A Humanoid Robot NAO: 25 Degrees of Freedom, 2 HD Cameras and 4 Microphones*

are equipped in each joint of the robot in order to have real time localization information from them which could provide well protection for robot safety. Each arm has five degrees of freedom and is equipped with those sensors to measure the position of joint movement. 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. By using these dimensions, proper size instrument can be selected and more accessories can be built which will be presented in the follow sections.

The four microphone locations embedded on the toy or NAO's head can be seen in figure 3.2. 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 wave file which meets the requirements for designing the online feedback audio score system will be described in later section.

3.1.2 Accessories

The purpose of this study is to have a toy-size humanoid robot to play and teach music for children with autism. Some necessary accessories needed to be purchased and made before the robot able to complete this task. All accessories will be discussed in the following paragraphs.

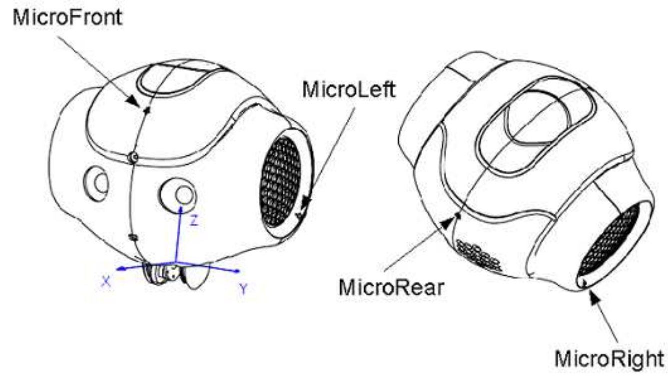


Figure 3.2: *Microphone locations on NAO's head*

Xylophone: A Toy for Music Beginner

In this system, due to NAO's open arms' length, a Sonor Toy Sound SM soprano-xylophone with 11 sound bars of 2 cm in width were selected and purchased. The instrument has a size of 31 cm x 9.5 cm x 4 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. See Figure 3.3. The 11 bars of the xylophone represent 11 different notes (11 frequencies) which covers one and half octave scale starting from C6 to F7.

In order to provide music teaching enviorment system for children with autism, xylophone is one of the best choice for such study. Xylophone, as well known as marimba, has categorized as a percussion instrument which consisting of a set of metal/wooden bars struck with mallets to produce fine musical tones. Other than keyboar or drum, for playing xylophone properly, a unique techique need to be applied. A proper strike movement is required in order to produce a beautiful note coming out of xylophone.

Figure 3.3: *Actual Xylophone and Mallets from NAO's Bottom Camera*

This action is perfect for practice motor control and the melody which played by user could also support the music emotion learning.

Mallet Gripper Design

For the mallets, we used the pair that came with the xylophone with a modified 3D printed griper which allows the robot hands to hold them properly. The mallets are approximately 21 cm in length and include a head of 0.8 cm radius. Comparing to other design, this mallet gripper prides nature hoding position for robot, and set up a proper modal for participants in holding the mallet stick in a good way. See Figure 3.4.

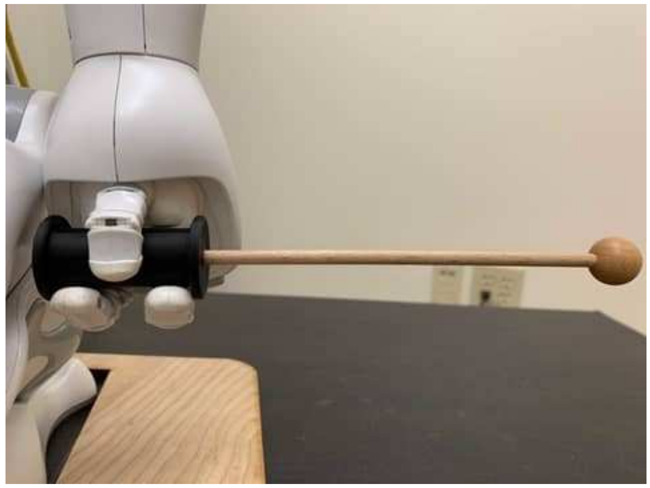


Figure 3.4: *Mallet Griper*

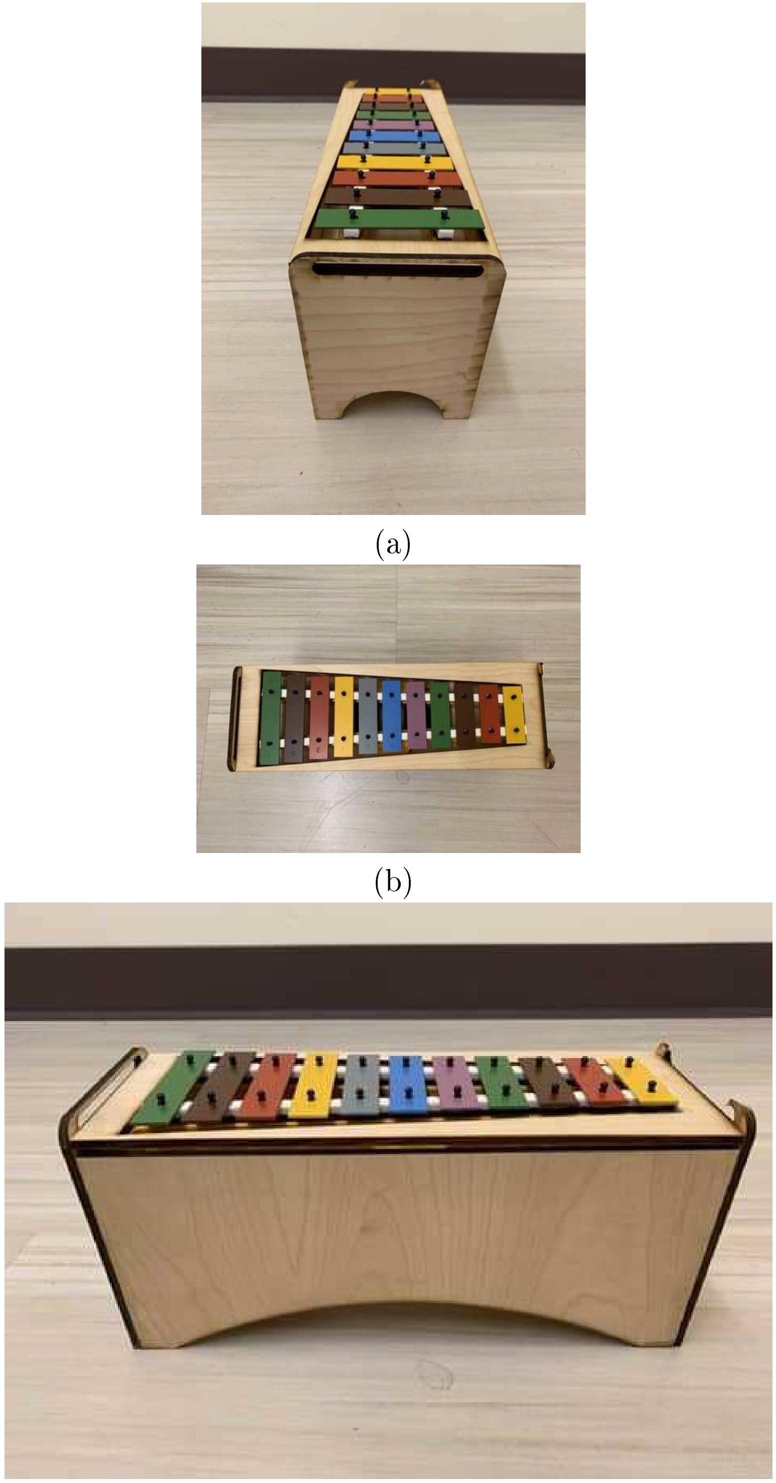
Instrument Stand Design

By carefully measured dimentions, a wooden base was designed and laser cut to hold the xylophone in a proper height for robot crouching position. Using this fine adjusted position, robot can easily fixed in a location and somehow have the same height level with the participants which makes it more nature in teaching activities for the entire time. See Figure 3.5.

3.2 Experimental Sessions Design

3.2.1 Experiment Room

All the sessions were held in a 11ft x 9.5ft x 10ft room located in the Ritchie School of Engineering Room 248, University of Denver. 6 HD surveillance cameras installed at corners, side wall and ceiling of the experimental room see Figure 3.6. One mini



(a)

(b)

(c)

Figure 3.5: *Instrument Stand: (a) Left View (b) Top View (c) Front View.*

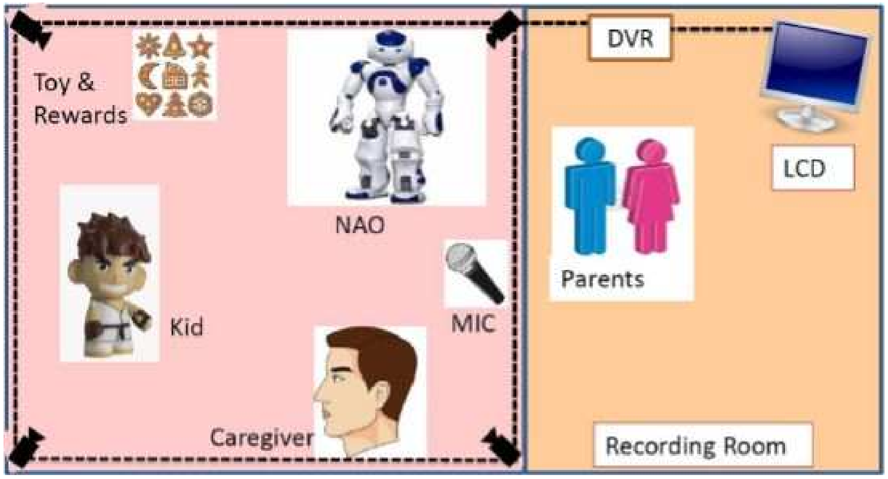


Figure 3.6: *Schematic robot-based therapy session and video capturing setting*

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 for collecting high quality audios for future process.

As shown in Figure 3.7, the observation room is located at the back of the one-way mirror facing at the rear side of participants in order to avoid distraction while sessions on going. Real-time video and audio were broadcasting to the observation room during sessions, which allowed researchers to observe and to record in the meantime. Parents behind the mirror may also call off the session in case of emergency.

3.2.2 Participant Selection

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. All participants were selected from the potential

Figure 3.7: *Experiment Room*

subject pool with help from Psychology department. For each participant in ASD group, 6 sessions were delivered including baseline session, intervention sessions and exit session. As for TD control group, only baseline and exit sessions were required for each participant. Every session lasts for 30-60 minutes total depends on the difficulty of each session and the performance of individuals. Typically, baseline and exit sessions length can be compariable for same subject. As for intervention sessions, duration should gradually increase one after another due to the challenge level uprising.

3.2.3 Session Detail

Baseline and exit session contains 2 activities which are 1) music practice and 2) music game play. Participants were asked to complete a challenge like full song play, starting from single note strike with color hint. Multiple notes, half song play and full song play were coming one by one as long as participants aced the previous task. After practice part done, a freshly designed music game will be presented to participants which contains three novel entertaining game modes in it, participants were allowed to communicate with robot regarding which mode to play with. Mode 1): robot will randomly pick a song from its song bank and play for kids, after each play, a music feeling will be request from participants in order to find out whether music emotion can be recognized from early age of children; Mode 2): a sequence of melody will be randomly generated by robot with consonance (happy or confortable feeling) or dissonance (sad or unfortable feeling) style, an oral emotion feeling from participants will be requested 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 by them, after robot complete playing, performance will be rated by the participants which provide a teaching experience for all subjects. Note that, there was no limitation for how many trails or modes each individual who wants to play for each visit, 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. By using the Module-Based Acoustic Music Interactive System, inputing multiple songs become possible and less time consuming. More than 10 songs are collected in the song bank, such as "Can Can" by Offenbach, "Shake if off" by Talor Swift, "Spongebob Square Pants" from Spongebob cartoon and "You are my Sunshaine" by Johnny Cash etc. Music styles are not only kid's song like "Twinkle Twinkle" but also across clasic, pop, ACG, folk and more can be played by using such platform. Because of the variate music style NAO can play, learning motivation can be successfully delivered among all sessions.

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, a motivated customized song were used in the following sessions 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 following 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. See details in Table 3.2.3

3.3 Module-Based Acoustic Music Interactive System Design

In this section, a novel module-based robot-music teaching system will be presented. Several missions need to be accomplished here: a) make robot play sequence of notes or melody fluently; b) have robot play note accurately; c) be able to adopt multiple songs easily; d) be able to have social communication between robot and participants; e) be able to deliver learning and teaching experience for participants; f) fast response and accurate decision making. In order to fulfill these tasks, a module-based acoustic music interactive system were designed in this wokr. Three modules have

Session #	Session Type	Outline	Session Details	Session Purpose
1	Baseline Session	Full activities with "Twinkle Twinkle Little Star" whole song play, human assessment	1. Computer music listing, the participant will be asked to listen to few music pieces composed by professional; 2. Participant should describe the feeling based on what they just listened, make selection on computer or tell researcher/therapist; 3. Participant free play the music pieces (2 mins), computer/researcher record to see which music which played /liked the most from Step 1.	Have better understanding of how ASD group perceive emotions from music.Compare EDA signals between ASD and TD groups, find out similarity. Compare EDA signals between ASD and TD groups, find out similarity.
		Music Game	1. NAO plays the the music piece select by participant; 2. Teach music piece from Step 1 gradually, form single strike, color strike, rhythm strike, multiple strike the entire music piece. 3. Participant free play xylophone (2 mins), NAO record how kids played for post process.	Getting social baseline from both groups in different tasks for future usage; Compare difference between both groups.
2	Intervention Session	Signal note practice with color hints, robot assessment	1. Participant should hit an arbitrary bar after each robot's struck; 2. Repeat Step 1, while the participant should use both hands in order; 3. NAO strikes single bar associate with a random color, and participant should select the proper color on a computer screen; 4. NAO show Red, Green, Blue, Pink on its eyes, participant strike the matching color on xylophone.	Improve in Joint Attention skills; Colors Recognition.
3		Multiple notes practice with color hints, robot assessment	1. Review Step 4 from Session 2; 2. NAO demonstrates hitting one or two bars in order and verbally repeats "one-two" with the movement; 3. Participant verbally says "one-two" along with robot hitting movement; 4. Repeat Step 2, use "color names" instead of "one-two", kid should imitate the strikes along with oral response.	Improve Verbal short-term Memory; Improve the perception of numbers and counting.
4		First half song practice with color hints	1. Review Step 4 from Session 3; 2. Imitation from the robot while striking 3 or 4 notes in order considering the time interval between each notes with saying "colors"; 3. Repeat Step 2 without oral response. 4. Participant play 3 or 4 notes on a virtual xylophone on tablet/computer, NAO imitate what has been played on real xylophone.	Improve auditory imitation skill; Good cooperation between eye and hands movements; Improve the Visual pursuit; Sharing attention between different tasks; Improve in playing turn-taking and group games in robot-child interactions.
5		Second half song practice with color hints, robot assessment	1. Review Step 3 from Session 4; 2. NAO demonstrates striking 2 or 3 notes with saying "note names from A to G", participant verbally repeats the notes name while robot hitting the bars; 3. Imitation from the robot while striking 3 or 4 notes in order considering the time interval between each notes with saying "note names/colors";	Working memory of children with autism, Dual task performance; Improve in joint attention skill; Improve in shifting attention skill; Improve in auditory memory of children.
6	Exit Session	Full Activities with customized whole song play, human assessment	1. Repeat Pre-Session	Compare results between ASD vs TD, Pre-Session vs Post-Session.

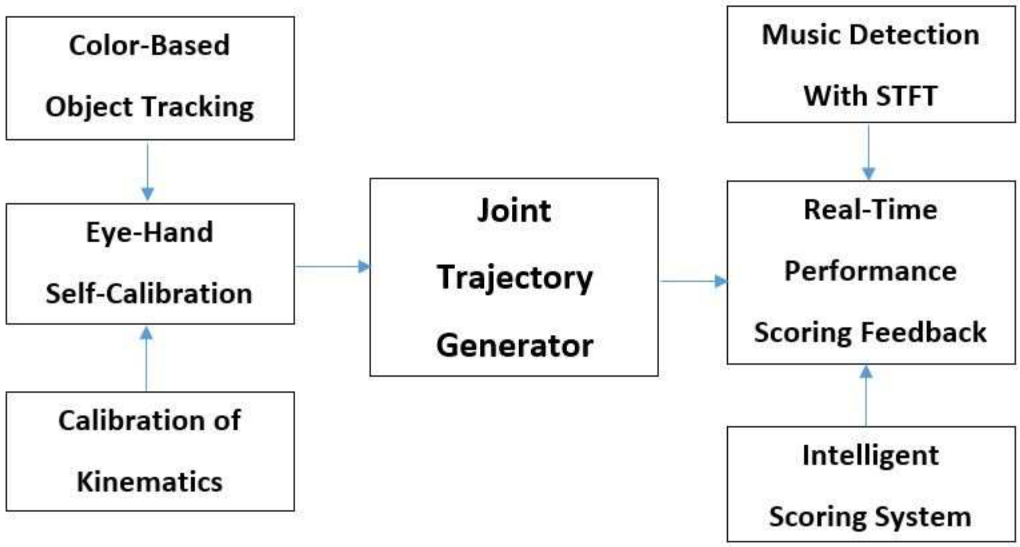


Figure 3.8: *Block Diagram of Module-Based Acustic Music Interactive System*

been 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. See Figure 3.8.

3.3.1 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 used a calibration method to accurately eliminate these errors.

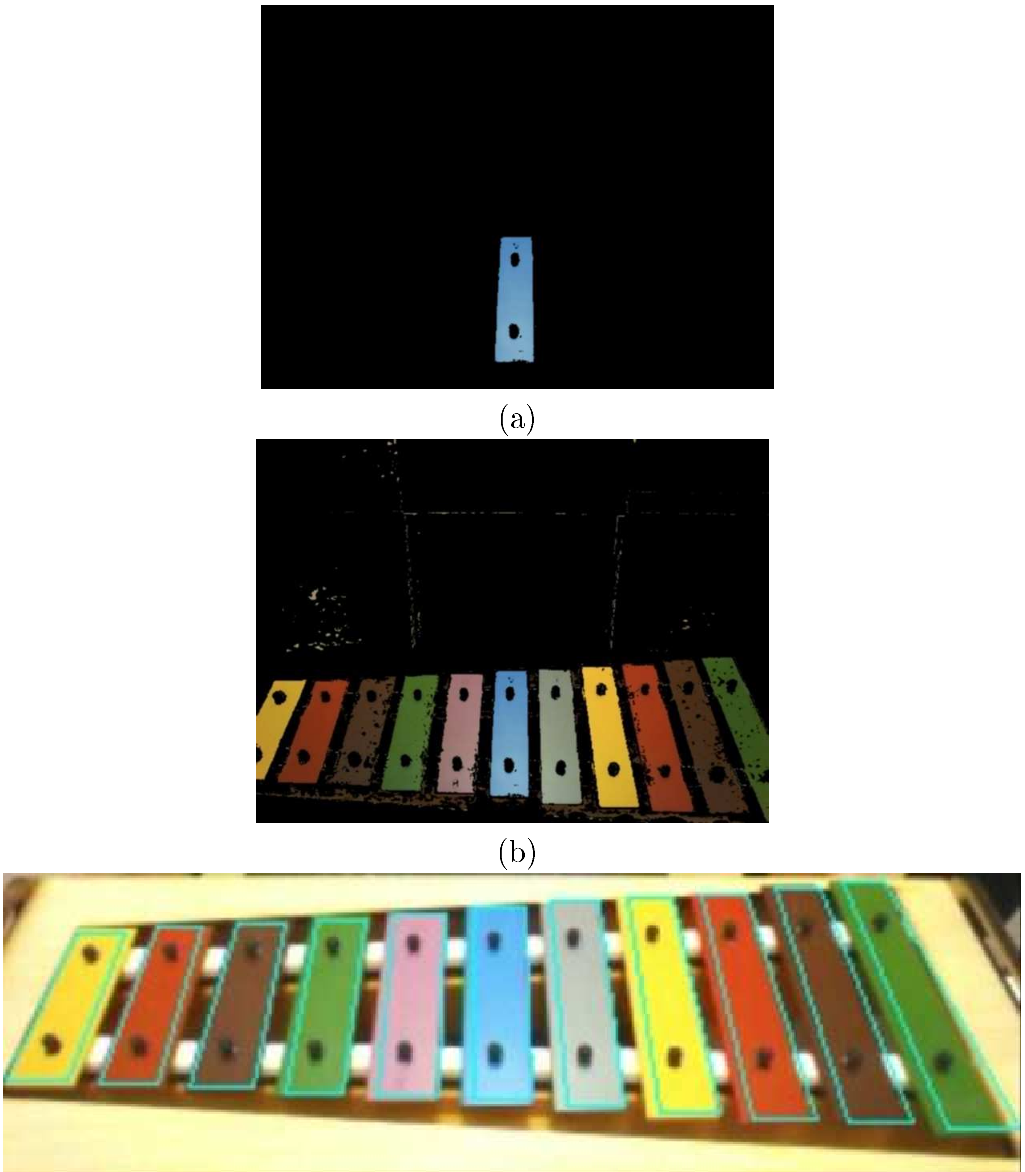
Color-Based Object Tracking

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, 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. See Figure 3.9

3.3.2 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.



(a)

(b)

(c)

Figure 3.9: *Color Detection From NAO's Bottom Camera: (a) Single Blue Color Detection (b) Full Instrument Color Detection (c) Color Based Edge Detection.*

3.3.3 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.

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. One then usually plots the changing spectra as a function of time. 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). Each chunk is Fourier transformed, and the complex result is added to a matrix, which records magnitude and phase for each point in time and frequency. This can be expressed as:

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

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.

B. Intelligent Scoring-Feedback System

In order to compare the detected notes and the target notes, we used an algorithm which is normally used in information theory linguistics called Levenshtein Distance.

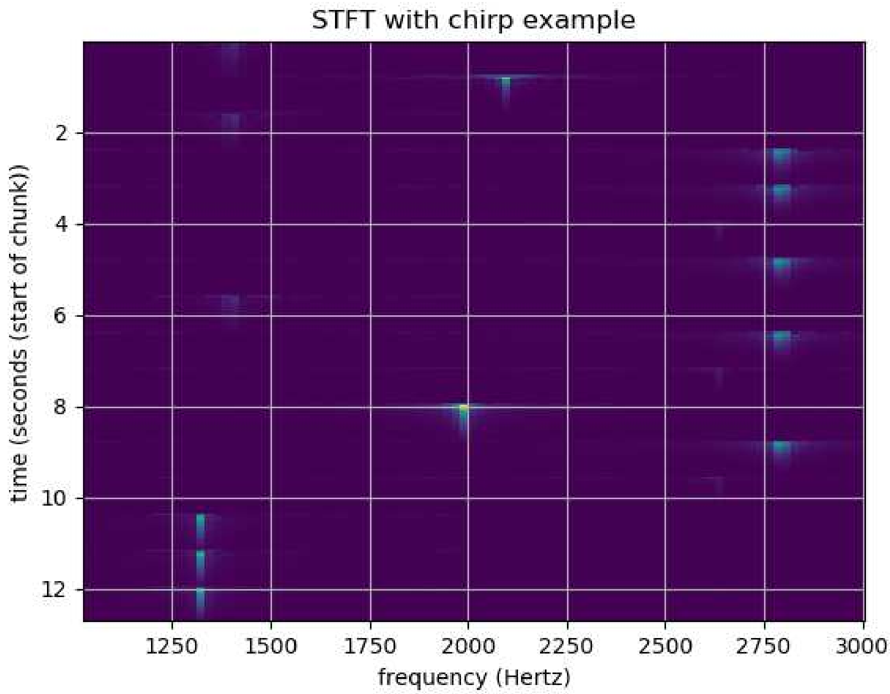


Figure 3.10: *Melody Detection with Short Time Fourier Transform*

This algorithm is a string metric for measuring the difference between two sequences.

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

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j)=0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j)+1 \\ \operatorname{lev}_{a,b}(i,j-1)+1 \\ \operatorname{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

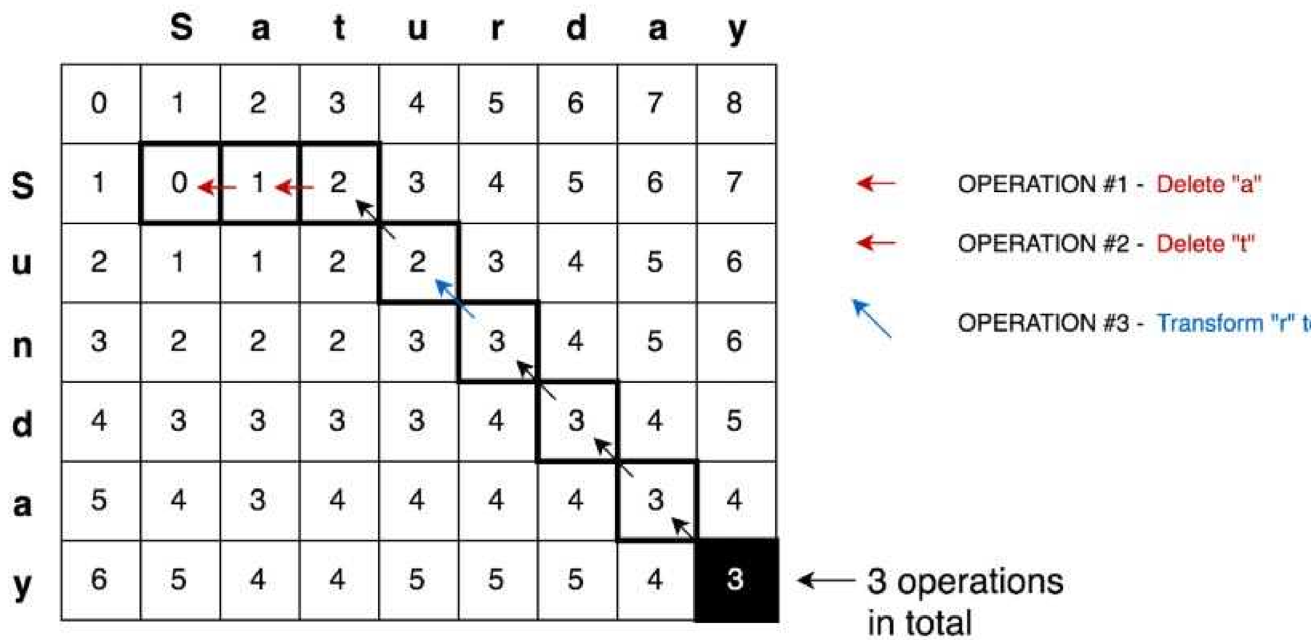


Figure 3.11: *An Example of Compute Levenshtein Distance for "Sunday" and "Saturday"*

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 $\mathrm{lev}_{a,b}(i, j)$ 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. Table 3.11 demonstrates how to apply this principle in finding the Levenshtein distance of two words "Sunday" and "Saturday".

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.

3.4 Summary

In this chapter, we have discussed both hardware and software design for the experiment sessions by using humanoid social assistive robot NAO in music teaching and playing.

From Chapter One, we determined to have NAO as a music teaching instructor be able to both teach children simple music and deliver social content simultaneously. In order to have the system ready, we first chose the proper agent, a robot named NAO, which is kid-friendly and has complex social abilities. Second, based on the size of the robot, some necessary accessories were purchased and handcrafted. A toy-sized color coded xylophone became the best option and based on the size and position, a wooden based xylophone stand was customized and assembled. Due to the limitation of NAO hands size, a pair of mallet gripers were 3-D printed and customized. At last, an intelligent module-based acoustic music interactive system was designed fully from scratch in order to complete the well designed experiment sessions. With all preparation, three modules were able to have the robot play, listen and teach the music freely. This allows NAO to become a great companion for children in both music learning and social communication.

Experiment session contains 3 types: baseline session, intervention sessions and exit session. A set of difficulty gradually increased music teaching activities were well designed among all sessions in order to deliver better music content for participants. Both ASD and TD group kids were experiencing a game like music challenge from playing single note to the whole song. Music game were designed to keep the session less tedious comparing to the learning part, it also provides a oppertunity for us to learn more about the difference in learning and teaching music from our participants.

Module-based acoustic music interactive system design were challenging. Several problems needed to be solved. In order to imporve the robot strike xylophone accuracy, Module 1 provides an autonomous self awareness positioning system for the robot to localize the instrument and make micro adjustment for arm joints that helps NAO plays the note bar properly. Multiple songs were required be able to played by NAO, program each song with specific arm movements sounds ridiculous. A easy music score inputing method need to be completed before session starts. Module 2 allows the robot to be able to play any customized song of the user's request. This means that any songs which can be translated to either C-Major or a-minor key can have a well-trained person type in the hex-decimal playable score and allow the robot to be able to play it in seconds. Music teaching requests real-time feedback, Module 3 was designed for providing real life music teaching experience for system users. Two key features of this module are designed: music detection and smart scoring feedback. Short time Fourier transform and Levenshtein distance are adopted to fulfill the requirement which allows the robot to understand music and provide proper dosage of practice and oral feedback to users.

Chapter 4

Pre-Study: A Wavelet-based Approach for Emotion Classification

In this chapter we are going to discuss the emotion classification method which will be used in the music teaching platform. The purpose of this pre-study is to have better understanding of emotion changes with young children and discover an automated method for emotion classification in children using Electrodermal activity signals (EDA). The purpose of this pre-study is to find a possible method for comparing the difference between non-autistic and autistic groups regarding the emotion changes in music social stimuli.

4.1 Data Acquisition

The dataset utilized in this pre-study constitutes a collection of multimodal recordings of social and communicative behavior of a hundred kids whose younger than thirty months provided by the Georgia Institute of Technology [65]. All data was collected within the Child Study Lab (CSL) at Georgia school, under a university-approved IRB protocol. The laboratory was 300-square feet area, and also the temperature/humidity of the area for all sessions was kept an equivalent. Based on the dataset description, every session lasted 3–5 minutes and the EDA signals (the frequency rate is 32 Hz) were collected from two Q-sensors attached to left and right wrists, and also the entire experiment was video-recorded. A collection of semi-structured play interactions with adults, called Multi-modal Dyadic Behavior (MMDB), was designed for the experimental sessions to stimulate different emotions: event 1: "greeting", event 2: "playing with a ball", event 3: "looking at a book and turning its pages", event 4: "using the book as a hat", and event 5: "tickling". These experiments are aimed at analyzing and decipherment the children's social communicative behavior at early ages and are in keeping with the Rapid-ABC play protocol [57].

The annotation was administrated supported the temporal relation between the video frames and the recorded EDA sequences of every subject. In different words, the annotators went through the entire video file of every event frame by frame, and designated the frames regarding the initiation finish of an emotion. Meanwhile, the corresponding sequences of the EDA signals were hold on to come up with the dataset for every

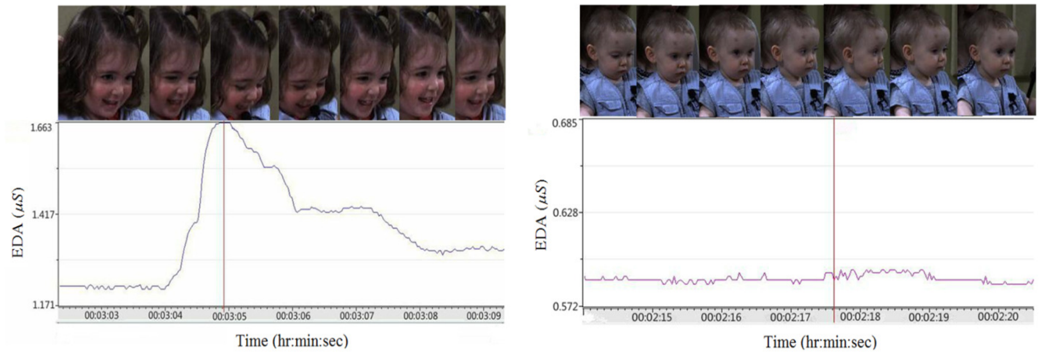


Figure 4.1: *Two samples of the annotation process, left shows some video frames associated with event 5 "tickling" and the right one shows the video frames of event "using the book as a hat". The corresponding EDA signals are shown under each case. While for the event 5 the EDA signal contains meaningful information, the EDA signal of event 4 does not contain useful information, likely due to the disengagement of the subject.*

perceived emotion. During the annotation, 2 dominant emotions were recognizable; events two (with average duration of forty five seconds) and 5 (average period of thirty five seconds) stimulate the “Joy” feeling and event three (with a mean period of sixty seconds) stimulates “Boredom”. With respect to event 1, “greeting”, it had been tough to assign a selected feeling to it; but, the annotators most frequently used “Acceptance” for this event. Additionally, we tend to excluded event 4 from our experiments since the length of this event (on average nine seconds) was terribly short compared with different events (on average fifty seconds for other events), and the annotators weren’t ready to determine any specific feeling triggered by this event. Figure4.1 shows the above-described procedure diagrammatically. Besides, the distribution of different emotions across all subjects and events is given in Figure 4.1.

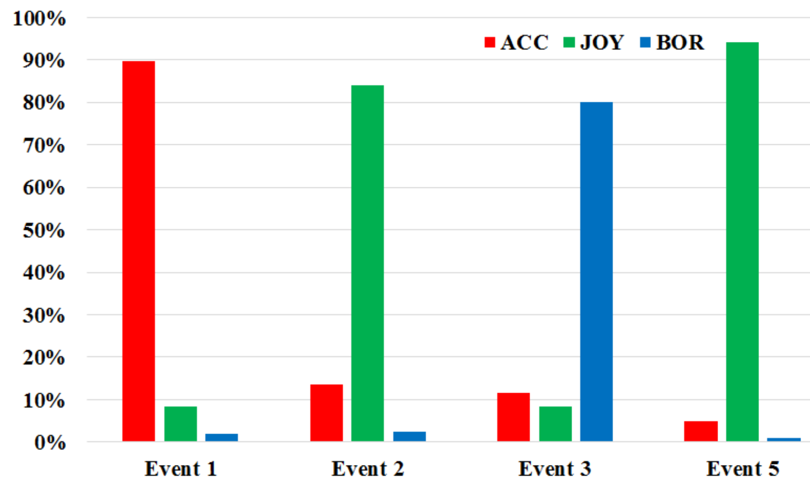


Figure 4.2: *The distribution of the emotions across all subjects and events. The abbreviations "ACC", "BOR", and "JOY" respectively correspond to the emotions "Acceptance", "Boredom", and "Joy".*

4.2 Classification Method

Since we have a tendency to developed our emotion classification methodology supported the time-frequency analysis of the EDA signals, the most properties of the continuous wavelet transform assumptive complex Morlet wavelet is first given here. Then, the pre-processing steps, as well as the wavelet-based feature extraction stage, are mentioned. Finally, we shortly review the characteristics of the support vector machine as the classifier used with our approach.

4.2.1 Continuous Wavelet Transform

The EDA data recorded using the SC sensors are categorized as non-stationery signals[53, 75]. Hence, multiresolution analysis techniques are essentially suitable to study the qualitative components of these kinds of bio-signals [53]. Note that continuous wavelet transform (CWT) is one of the strongest and most widely used analytical tools for multiresolution analysis. CWT has received considerable attention in processing signals with non-stationary spectra [83, 46]; therefore, it is utilized here to perform the time-frequency analysis of the EDA signals. In contrast to many existing methods that utilize the wavelet coefficients of the raw signal to extract features, our proposed method is essentially based on the spectrogram of the original data in a specific range of frequency (0.5, 50)Hz, which provides more information for other post-processing steps (i.e., feature extraction and classification). We apply the wavelet transform at various scales corresponding to the aforementioned frequency range to calculate the spectrogram of the raw signal (i.e., Short Time Fourier Transform (STFT), can also be used to calculate the spectrogram of the raw signal). In addition, as opposed to many related studies that utilize real-valued wavelet functions for feature extraction purposes, we have employed the complex Morlet (C-Morlet) function with the proposed approach, as it takes into account both the real and imaginary components of the raw signal, leading to a more detailed feature extraction.

The wavelet transform of a 1-D signal provides a decomposition of the time-domain sequence at different scales, which are inversely related to their frequency contents[46, 21] . This requires the time-domain signal under investigation to be convolved with a time-domain function known as "mother wavelet". The CWT applies the wavelet function at different scales with continuous time-shift of the mother wavelet over the input signal. As a consequence, it helps represent the EDA signals at different levels of resolution. For instance, it results in large coefficients in the transform domain when the wavelet function matches the input signal, providing a multiscale representation of the EDA signal.

Using a finite energy function Ψ(t) concentrated in the time domain, the CWT of a signal x(t) is given by X(α,b) as follows [83]:

$$X(a, b) = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{a}}\Psi(\frac{t-b}{a})dt$$

where, α, is the scale factor and represents dilation or contraction of the wavelet function and b is the translation parameter that slides this function on the time-domain sequence under analysis. Therefore, Ψ(α,b) is the scaled and translated version of the corresponding mother wavelet. "*" is the conjugation operator.

Note that the wavelet coefficients obtained from Eq. (1) essentially evaluate the correlation between the signal x(t) and the wavelet function used at different translations and scales. This implies that the wavelet coefficients calculated over a range of scales and translations can be combined to reconstruct the original signal as follows:

$$x(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} X(a, b)\Psi(\frac{t-b}{a})dadb$$

4.2.2 Wavelet-Based Feature Extraction

The time-frequency analysis of varied bio-signals has been addressed in many related literature [22, 23, 24, 39]. It has been shown that the wavelet-domain feature area will improve the classification performance of various human activities using the signals emanated from the body responses. Therefore, it primarily enhances the classification performance because of the additional eminence area provided.

In this pre-study, we tend to specialise in the time-frequency analysis of the EDA signal to produce a new feature area supported that emotion classification task is done. As opposed to some connected studies that use the raw time-domain signals for classification purposes [25, 28], we use the amplitude of the CWT of the EDA signals to get the options and drive the classifier. Operating within the wavelet-domain is actually advantageous since the wavelet remodel probes the given signal at completely different scales, extracting a lot of information for alternative post-processing steps. Additionally, the localized support of the wavelet functions permits CWT-based analysis to match to the native variations of the input time sequence [83]. As a result, a lot of elaborate representation of the signal is provided as compared with the raw time-domain signal.

Figure 4.3 shows the amplitude of the CWT of a sample EDA signal at different scales using a complex Morlet (C-Morlet) wavelet function. Different scales of the wavelet functions are convolved with the first EDA signal to spotlight completely different options of the data. As may be seen, thanks to the localization property of the CWT, completely different structures of the signal are extracted at every level of decomposition, providing helpful info for analyzing the recorded EDA signals.

This work has employed 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 [22]. Figure 4.4 shows the wavelet-based feature extraction, 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.

On the other hand, the detailed structures of the signal are better extracted when the scaling factor decreases. Note that the impact of different families of the wavelet functions (e.g., Symlets, Daubechies, Coiflets) on the emotion classification will be evaluated in the next subsection. The equation of the C-Morlet mother wavelet with fc as its central frequency and fb as the bandwidth parameter is given as follows:

$$\Psi(t) = \frac{\exp(-t^2/f_b)}{\sqrt{(\pi f_b)}} \exp(j2\pi f_c t)$$

4.2.3 Support Vector Machine

The SVM classifier tends to separate data

$$D = \{x_i, y_i\}_{i=1}^{N}, x_i \in^d, y_i \in \{-1, +1\}$$

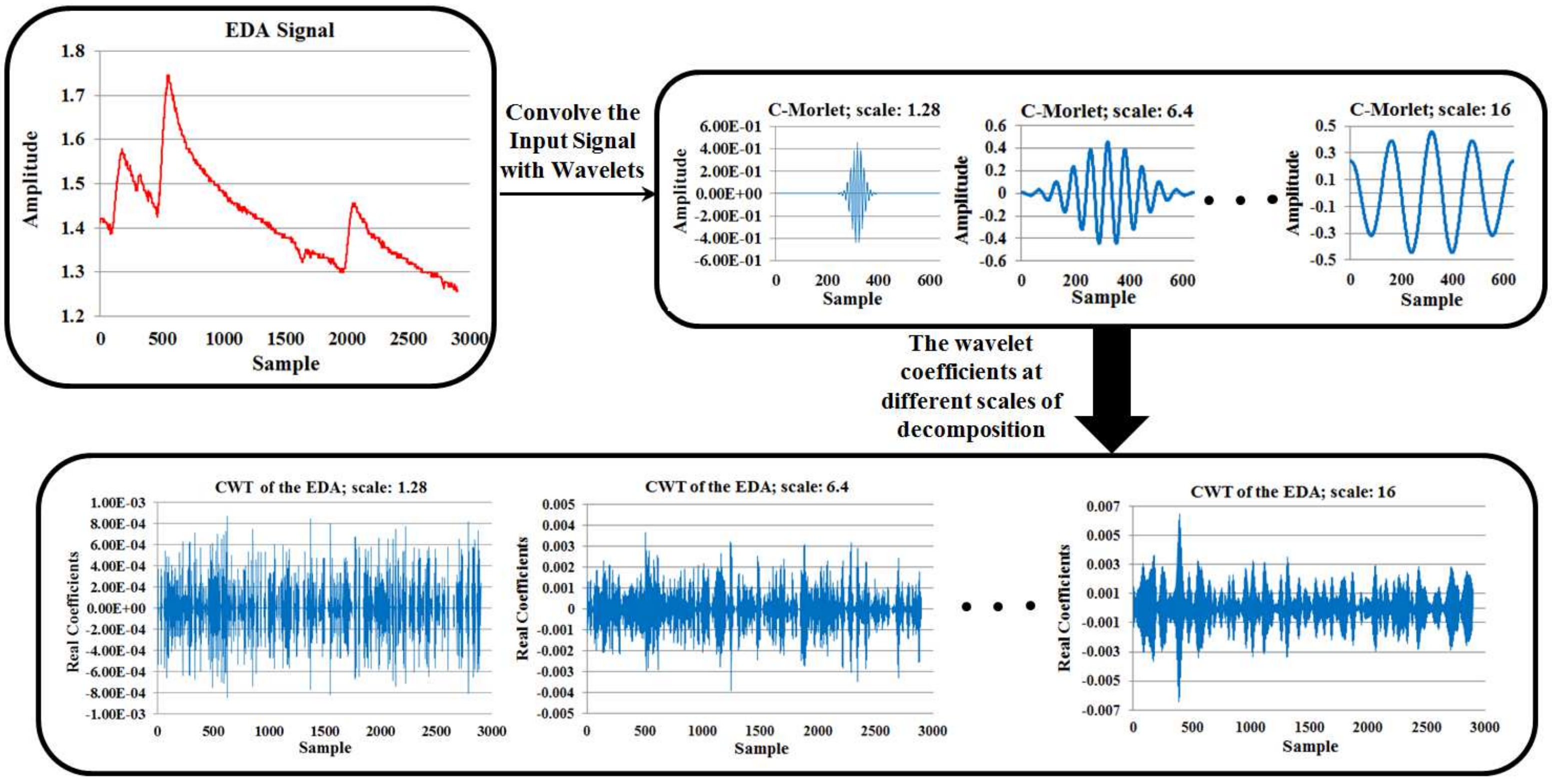


Figure 4.3: *The CWT of a typical EDA signal using the C-Morlet mother wavelet. Different scales of the wavelet functions are convolved with the original EDA signal to highlight different features of the raw data. As can be seen inside the bottom box, when the scaling parameter of the wavelet function increases, the larger features of the input signal are augmented. On the other, the detailed structures of the signal are better extracted when the scaling decreases.*

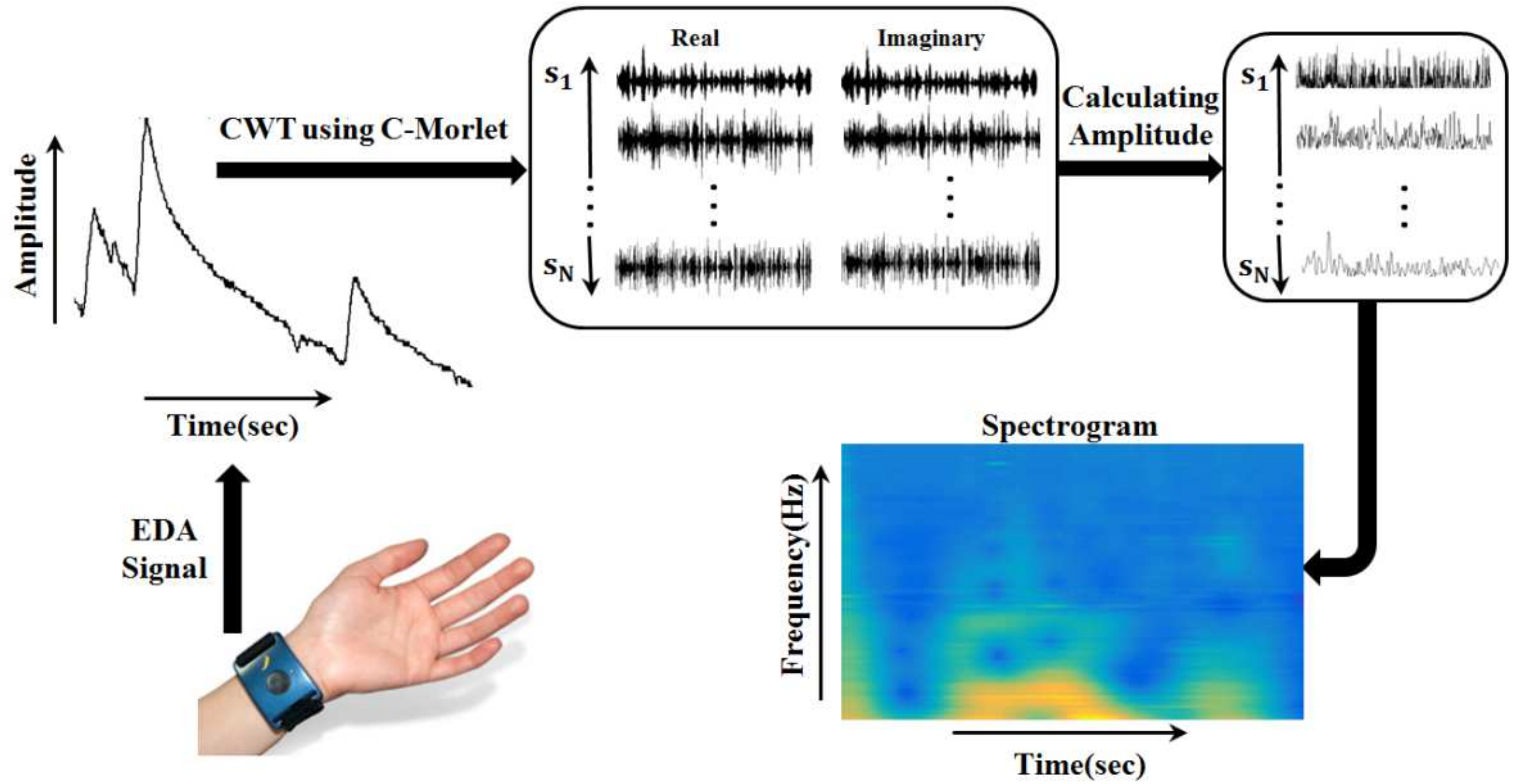


Figure 4.4: *The wavelet-based feature extraction. Uisng the C-Morlet mother wavelet, the real and imaginary wavelet coefficients are calculated at different scales. Then the amplitude of these coefficients is calculated to provide the corresponding spectrogram. This spectrogram is then used as the feature space.*

by drawing an optimal hyperplane $<$w,x$>$+b=0 between classes such that the margin between them becomes maximum [13]. With reference to Figure 4.5, The decision boundary is shown by OH. Two hyperplanes H1 and H2 pass the support vectors that are circled inside the figure. H1 and H2 are the supporting planes and the optimal hyperplane (OH) splits this margin such that it stands at the same distance from each supporting hyperplane. This implies that the margin between H1 and H2 is equal to 2 / ‖w‖. In terms of linearly separable classes, the classifier is obtained by maximizing the margin 2 / ‖w‖, which is equivalent to minimizing ‖w‖ / 2 with a constraint in convex quadratic programming (QP) as follows:

$$\min \frac{1}{2}||w||^2 s.t. y_i(< w, x_i > +b) \geq 1$$

where, w and b are the parameters of the hyperplane and $<.,.>$ is the notation of the inner product.

However, different classes are seldom separable by a hyperplane since their samples are overlapped in the feature space. In such cases, a slack variable $\xi_i \geq 0$ and a penalty parameter C $\geq$ 0 are used with the optimization step to obtain the best feasible decision boundary. It is given as:

$$\min \frac{1}{2}||w||^2 + C(\Sigma_{i=1}^{N} \xi_i) s.t. y_i(< w, x_i > +b) \geq 1 - \xi_i$$

Usually, various kernel functions are used to deal with the nonlinearly separable

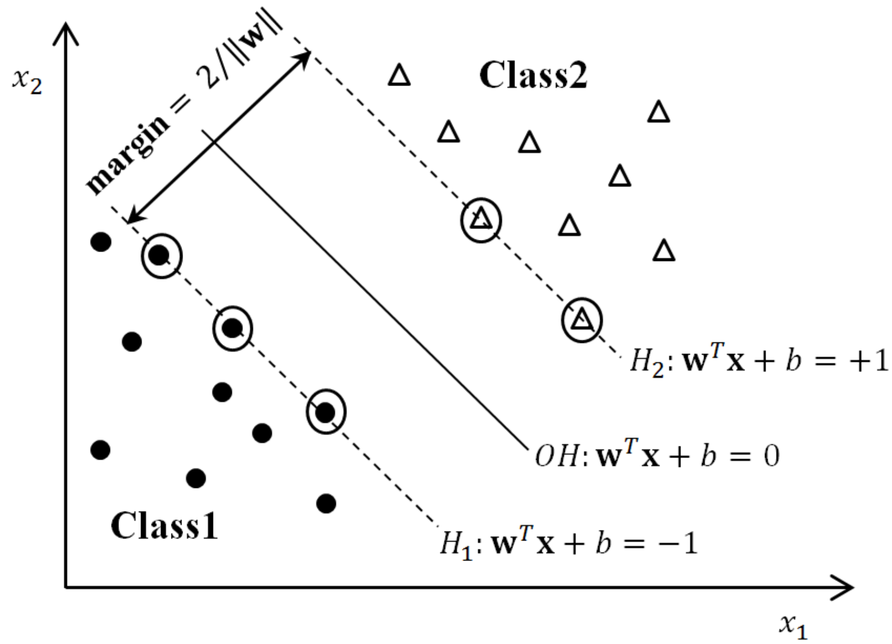


Figure 4.5: *Canonical SVM for classifying two linearly separable classes. The decision boundary is shown by OH. Two hyperplanes H1 and H2 pass the support vectors that are circled inside the figure.*

data. As a result, the original data xi is mapped onto another feature space through a projection function $\varphi(\cdot)$. It is not necessary to exactly know the equation of the projection $\varphi(\cdot)$, but one can use a kernel function $k(x_i, x_j) =< \varphi(x_i), \varphi(x_j) >$. This function is symmetric and satisfies the Mercer's conditions. The Mercer's conditions determine if a candidate kernel is actually an inner-product kernel. Let $k(x_i, x_j)$ be a continuous symmetric kernel defined in the closed interval $t_1 \leq t \leq t_2$, the kernel can be expanded into series $\Sigma_(n = 1)^\infty = \lambda_n \varphi_n(x_i) \varphi_n(x_j)$, where $\lambda_n > 0$ are called eigenvalues and functions φn are called eigenvectors in the expansion. The fact that all the eigenvalues are nonnegative means that the kernel is positive semidefinite [13].

To maximize the margin, H1 and H2 are pushed apart until they reach the support vectors on which the solution depends. To solve this optimization problem, the Lagrangian dual of equation is used as follows:

$$\max_\alpha \Sigma_{i=1}^N \alpha_i - \frac{1}{2} \Sigma_{i=1}^N \Sigma_{j=1}^N y_i y_j \alpha_i \alpha_j k(x_i, x_j)$$
$$s.t. 0 \leq \alpha_i \leq C, \Sigma_{i=1}^N \alpha_i y_i = 0, i = 1, ..., N$$

where, α_is are the Lagrangian multipliers in which just a few number of them are non-zero. These non-zero values are corresponding to the support vectors determining the parameters of the hyperplane $w = \Sigma_(i = 1)^N \alpha_i y_i x_i$. Therefore, the label of the test sample (y_z) is given by:

$$y_z = sgn(\Sigma_{i=1}^N \alpha_i y_i k(x_i, z)) + b$$

4.3 Experimental Result

This work has employ the EDA signals of 64 aubjects annotated based on the facial expressions from participants in order to evaluate the accuracy of the proposed wavelet-based feature extraction method on the emotion classification performance. The EDA dataset is classified based on different emotions perceived in the annotation step, which includes Joy, Boredom, and Acceptance emotions. The SVM classifier is applied on the dataset using three different kernel functions including the Linear function k (x, y) = x T y + c, Polynomial function k (x, y) = (x T y + c) d , and Radial Basis Function (RBF) k (x, y) = exp (γ|| x –y || 2), where x and y are two feature vectors, and γ, c, and d are constant values.

First thing need to be done before we proceed with such quantitative performance of the emotion classification method, which is to test the impact of various families of wavelet functions on the feature extraction stage as well as emotion classification ability.

4.3.1 Determination of Mother Wavelet

Table 4.1 shows the classification results given by different wavelet functions. For the sake of brevity, exclusively the results of the "db1", "coif1", "sym2", and "C-Morlet"

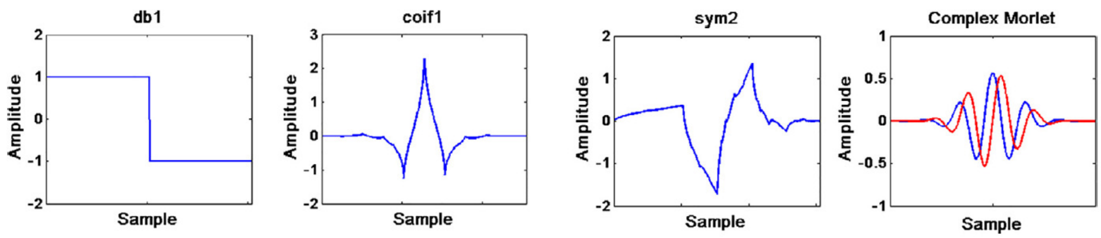


Figure 4.6: *Different mother wavelets used for feature extraction in this paper. The "db1", "coif1" and "sys2" wavelets are real-valued functions, while the "C-Morlet" function is complex-valued. The corresponding imaginary part of this wavelet function si highlighted in red insid the figure.*

wavelets and all three kernels with the SVM classifier are shown. As can be typically seen, the time-frequency features calculated by the C-Morlet ends up in following classification performance, in all probability thanks to the distinctive feature space provided by this classification perform. Figure 4.6 shows the difference between the mentioned wavelet functions. Note that the C-Morlet wavelet has successfully been applied on different types of bio-signals (e.g., EEG, LFP brain signals) and lead to promising results, specifically for feature extraction functions. One in every of the foremost characteristics of this wavelet function is its sophisticated nature that primarily tends to extract various features from the input time sequence.

4.3.2 Classification Result

The pre-processing stages performed to the raw EDA dataset are explained at the beginning in this section. After wards, the recognition results with various modalities under SVM and KNN classifiers are presented. Classification performance of

	Kernels	bd1	coif1	sym2	C-Morlet
ACC-BOR	Linear	61	56	61	75
ACC-JOY		50	46	50	69
BOR-JOY		51	69	57	90
BOR-JOY-ACC		51	35	39	66
ACC-BOR	Polynomial	51	56	58	64
ACC-JOY		54	54	57	81
BOR-JOY		55	64	69	86
BOR-JOY-ACC		43	46	50	61
ACC-BOR	RBF	59	60	58	74
ACC-JOY		55	44	57	84
BOR-JOY		68	51	69	89
BOR-JOY-ACC		45	35	50	69

Table 4.1: *Comparison of different wavelet functions on the feature extraction and emotion classification performance (%) of 2 and 3 classes using SVM classifier with different kernels. The abbreviations "Acc", "Bor", and "Joy" respectively stand for the emotions "Acceptance", "Boredom", and "Joy".*

the suggested wavelet-based feature extraction method with the raw EDA signal are compared here. Plus, statistical feature extraction methods [44, 48] used for EDA signal performance are also compared with proposed feature extraction method. Note that the extracted features with this method are mainly based on the statistical moments of the acquired EDA time sequence such as “the means of the raw signals”, “the standard deviations of the raw signals”, “the means of the absolute values of the first differences of the raw signals”, “the means of the absolute values of the first differences of the normalized signals”, “the means of the absolute values of the second differences of the raw signals”, and “the means of the absolute values of the second differences of the normalized signals”.

A median filter of size 10 are applied to the segments of the EDA signals which obtained from the annotation step in order to smooth the signal, eliminating some existing implusive noise that may happens due to the sudden move of the subjects during the experiments which conducted by other group. Then, the amplitude of the wavelet coefficients are calculated for the frequency range of (0.5, 50)Hz. The reason to use such wide frequency range was to secure all detailed components of the EDA signal are takien into account (See Figure 4.3).

Principal component analysis (PCA) [1] is then applied on the extracted wavelet-based features to decrease the dimensionality of the data, and, therefore, reduce the compu- tational burden. PCA is a well-known dimensionality reduction approach which is extensively used for data analysis before classification. Therefore, it can decrease the chance of overfitting, which may happen beacuse of enormous size of the faeture vectors. Note that, in our experiments, 95% of the eigen-values corresponding to the maximum variance directions are kept. Since the spectrogram of the raw EDA data (see Figure 4.4) is calculated for 100 scales (e.g., Frequency range (0.5, 50)Hz with a resolution of 0.5 Hz), for a fair comparison, we first down-sample the spectrogram by a factor of 100 to make the length of the wavelet-based features equal to the length of the raw data. Then, PCA is applied on it. As a result, on average, the length of the wavelet-based feature vector before and after PCA is respectively 1000 and 35 samples, while these lengths are 1000 and 15 samples for the raw data.

In order to generate the training and test sets for the classification step, leave-one-out cross validation (LOOCV) approach has adopted. To achieve the best accuracy of classification on the validation set, the parameters of the hyperplan are fixed in terms of the SVM classifier (LibSVM library [10]). The following parameters are used for each kernel function: 1. Linear kernel C = 0.01, 2. RBF kernel C = 0.01, γ= 0.001, and 3. Polynomial kernel C = 0.01, d = 2. These values are experimentally set so as to obtain the best classification performance.

Classification accuracy for SVM and KNN classifier applied with different kernel functions is shown in Table 4.2, and the dataset acquired from 64 annotated subjects. In terms of the binary classification cases (i.e., "Acceptance vs Boredom", "Acceptance vs Joy", and "Boredom vs Joy"), besidesthe classification accuracy, the quantitative measures precision (true positive / (true positive + false positive)), recall (true positive / (true positive + false negative)), and AUC (area under the receiver operating characteristic curve), which are also given in this table. In order to calculate the precision and recall, first of all, a positive class is choosen from one of the emotions and then the precision and recall values will be calcuated. After that, order will be changed and the other emotion is used as the positive class and calculate the precision and recall values. Then the final step is to calculate the average value of all precision and recall which is given in the table. It is obvious be seen, compare to other feature extraction methods, the proposed wavelet-based features lead to a higher classification performance among almost all cases (both SVM and KNN classifier). In SVM classifier, for example, with the linear kernel SVM and raw EDA signal, the classification rate of 3-class case is about 38%. However, the introduced feature space reaches an accuracy of 68%. From Table 4.2, note that competitive classification performance for polynomial kernel shown in the raw EDA data and the combination of the raw data and the statistical features, while the proposed wavelet-based features lead to a stable performance among all kernel functions. Looking at the KNN classifier, 3 different valuses for K = 1, 3, 5 are used in this study. As can be seen, the proposed method outperforms in most of the cases. The results obtained by the combination of the raw data and the statistical features surpass the proposed method for some classification tasks. For instrance, in K = 1 and "Acceptance vs Boredom" task, it obtains a accuracy of 73% compare to the proposed method which reaches 70% for the same task. However, the proposed method shows better classification performance in most of the other cases. For 3-class classification, proposed method achieves 64% of accuracy for all K values on average, while other two feature extraction approaches reslut in 57% and 44% respectively. This result indicates the superiority of the proposed method for the complex classification missions. Note that one major problem when analyzing physiological signals is noise interference. In particular, the EDA signal is nonstationary and may include random artifacts, which makes it unsuitable to use the raw time sequence for practical signal process- ing approaches. Prior studies have represented stochastic physio- logical signals using statistical features to classify emotional states [48]. Unfortunately, information can be lost with such features as simplifying assumptions are made, including knowledge of the probability density function of the data. Furthermore, there may be signal features that have the potential to improve emotion classification accuracy, but are not yet identified [76].

	SVM	Wavelet-based				Statistics-based feature + Raw data				Raw data			
	Kernels	Accuracy	AUC	Precision	Recall	Accuracy	AUC	Precision	Recall	Accuracy	AUC	Precision	Recall
ACC-BOR	Linear	75	75	81	71	56	55	57	56	56	55	58	43
ACC-JOY		69	84	75	79	47	56	47	47	49	57	49	39
BOR-JOY		90	87	82	88	55	55	54	55	53	54	52	67
ACC-JOY-BOR		66				35				36			
ACC-BOR	Polynomial	64	68	70	64	74	82	83	74	70	78	79	70
ACC-JOY		81	83	78	81	77	90	84	77	60	68	72	60
BOR-JOY		86	87	85	86	60	66	59	60	57	62	57	57
ACC-JOY-BOR		61				55				46			
ACC-BOR	RBF	74	79	84	74	53	58	54	53	57	59	60	57
ACC-JOY		84	89	85	84	42	61	37	42	42	65	30	42
BOR-JOY		89	90	85	89	51	54	51	51	50	55	50	50
ACC-JOY-BOR		69				34				34			
	KNN K values	Accuracy	AUC	Precision	Recall	Accuracy	AUC	Precision	Recall	Accuracy	AUC	Precision	Recall
ACC-BOR	K = 1	70	50	70	70	73	53	72	73	68	52	68	68
ACC-JOY		76	58	76	76	76	63	77	76	62	60	62	62
BOR-JOY		80	64	76	80	57	66	57	57	56	68	56	56
ACC-JOY-BOR		60				56				43			
ACC-BOR	K = 3	70	63	71	70	81	77	79	81	71	68	75	71
ACC-JOY		82	81	80	82	82	82	81	82	61	50	61	61
BOR-JOY		85	84	78	85	60	53	61	60	55	56	55	55
ACC-JOY-BOR		65				56				46			
ACC-BOR	K = 5	64	64	67	64	75	78	74	75	71	67	75	71
ACC-JOY		77	84	73	77	77	82	77	77	57	50	58	57
BOR-JOY		85	86	77	85	61	58	63	61	46	57	46	46
ACC-JOY-BOR		66				58				43			

Table 4.2: *First half of the table is the comparison of classification accuracy (%) of SVM classifier with different kernel functions using the presented wavelet-based feature extraction, the raw EDA data, and the raw EDA data + statistical features. The results of 64 subjects and 2 and 3-class classification cases are reported. The abbreviations "ACC", "JOY", and "BOR"respectively stand for the emotions "Acceptance", "Joy", and "Boredom". The best value is highlighted in each case. Bottom half of the table is the comparison of classification accuracy (%) of KNN classifier with different K values using the presented wavelet-based feature extraction, the raw EDA data, and the raw EDA data + statistical features. The results of 64 subjects and 2 and 3-class classification cases are reported. The abbreviations "ACC", "JOY", and "BOR" respectively stand for the emotions "Acceptance", "Joy", and "Boredom". The best value is highlighted in each case.*

4.4 Summary

Three basic emotions were recognized within the annotation step as well as Acceptance, Joy, and Bordom. Numerous experiments were dole out on the dataset mistreatment either the raw segmented EDA signal or its corresponding time-frequency illustration as options. The quantitative results show that the emotion classification performance is remarkably improved once the planned wavelet-based options are used with the SVM classifier. Apart from the "C-Morlet", we've got additionally evaluated the impact of various wavelet functions such as "Symlets", "Daubechies", and "Coiflets" on the feature extraction stage, and thus, the classification performance. The experimental results confirmed the prevalence of the "C-Morlet" wavelet function. Developing an automatic system capable of real-time operation of the info will be a remarkable extension to the current work. This permits us to observe emotions and give feedback to the participants throughout the experimental sessions. Moreover, due to the limitation of obtainable datasets, making a more comprehensive dataset would be necessary for the longer term analysis. The quantitative results show that the emotion classification performance remarkably improves compared to other methods when the proposed wavelet-based features are used. This pre-study also provided a possibility of using C-Morlet wavelet function as feature extraction method for emotion recognition in music social interaction for children with autism.

Chapter 5

Experimental Result

Several questions will be answered in terms of social skills in this chapter: 1) Turn-taking: How well kids with autism hehave during the music activities;

2) Motor control: How well ASD kids play the xylophone in terms of volume, pitch and accuracy after intervention sessions. e.g., a good multiple strikes should be recognized by STFT as a sequence of frequencies;

3) Social Engagement: How kids engage with different level of music teaching events;

4) Emotion Fluctuation: How emotion change among activities? How emotion change in single event? What is the different target and control groups?

5.1 Social Behavior Result

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

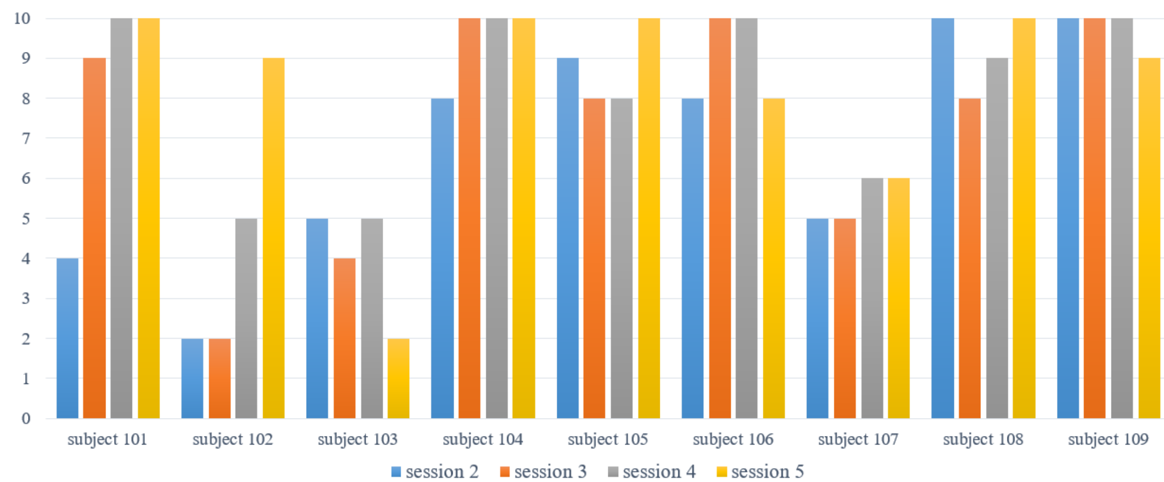


Figure 5.1: *Motor Control Accuracy Result*

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 5.1, 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 5.2 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 5.2, 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.

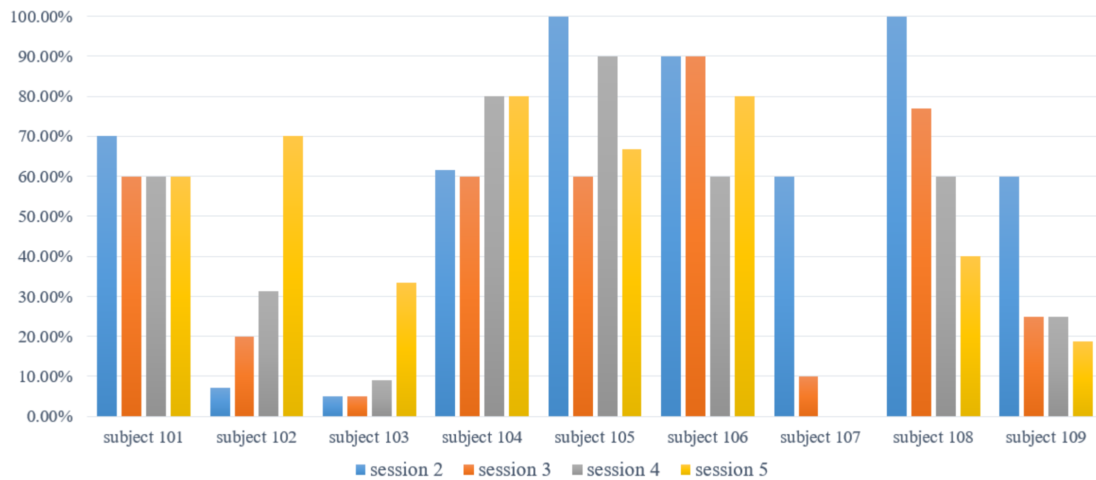


Figure 5.2: *Main Music Teaching Performance Accuracy*

5.2 Music Emotion Classification Result

Since we developed our emotion classification method based on the time-frequency analysis of the EDA signals, the main properties of the continuous wavelet transform assuming complex Morlet wavelet is first presented here. Then, the pre-processing steps, as well as the wavelet-based feature extraction scheme, are discussed. Finally, we briefly review the characteristics of the support vector machine as the classifier used with our approach.

EDA signal was also collected in this study. By using the annotation and analysis method from pre-study [18], 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 5.1 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 5.2 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

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

Table 5.1: *Emotion change in different events using wavelet-based feature extraction under SVM classifier.*

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 5.3 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 5.1.

The types of activities and process of the session between baseline session for

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 5.2: *Emotion change classification performance in single event with segmentations using both SVM and KNN classifier.*

	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

Table 5.3: *classification rate in robot demo, kids play and robot feedback across warm up (S1) and music practice (S2) sections*

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

Table 5.4: *TD vs ASD Emotion Changes from Baseline and Exit Sessions*

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 5.3. 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 better as possible while they made mistakes.

5.3 Each Participant Has A Story

In this section I will list all details which need to be mentioned during the session for all asd kids. still working on it, not sure if it is a good idea?

5.3.1 Subject 101

5.3.2 Subject 102

5.3.3 Subject 103

5.3.4 Subject 104

5.3.5 Subject 105

5.3.6 Subject 106

5.3.7 Subject 107

5.3.8 Subject 108

5.3.9 Subject 109

5.4 Summary

All the experimental results are presented in this chapter, some of the answers can be found out of them. According the report from annotators, most of the kids (both ASD and TD groups) shown well turn-taking communication behavior among all sessions.

However, difference can also be found comparing both groups. All TD participants could initiate the activities from the beginning of the session, while some of the ASD kids may need some help although after several visits of intervention sessions, most of them can perform well turn-taking skills as well as TD group. In terms of motor control skill, as can be seen from Figure 5.1, most of the ASD participants can master this skill after first few visits. For the ones who may not play xylophone properly, a improvement also can be found in this figure. Based on the recorded videos and Figure 5.2, more than half of the ASD kids shown well engagement during the intervention sessions. Few of them need help from the researcher to complete tasks at first one or two sessions. Since the music was chosen by each individual, this could somehow motivate them more engage even with repetitive activites. From Chapter 4 we learned that emotion classification using EDA signal can be possible, and a wavelet-based feature extraction method was developed and applied in such research with younger age of children. In this chapter we adopted proposed approach in order to decode the emotion fluctuation with music social activites for children with autism. Multiple experiment has been established in this chapter, emotion changes were compared across different events, within one activites and between target and control group. Detailed information has shown in tables from previous sections, and more detailed discussion and conclusion can be found in next chapter.

Chapter 6

Discussion and Conclusion

The results indicates that the conferred music education platform will be thought-about as a decent tool for facilitate improving fine motor control, turn-taking skills and social activities engagement. The automated music detection system created a self-adjusting surroundings for participants in early sessions. Most of the ASD youngsters began to develop the strike movement when initial 2 intervention sessions, some even will master the motor ability throughout the very firt warm-up event. Though the robot might provide verbal directions and demonstrations by voice command input from participants whenever they need it. However, majority of the participants didn't request such service whereas playing with NAO. This finding suggests that fine motor control ability will 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 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 5.2 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 5.2. Comparing emotion patterns from baseline and exit sessions between TD and ASD groups in Table 5.3, 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.

Chapter 7

Link The Future: X-Elophone, A New Instrument Experiment Design Concept

In this chapter, a novel instrument will be described. This design is needed due to the limitation of keys. The purpose of this design provides more possibility for different timber and major/minor keys. Increased amounts of possible notes allows the system to play more customized songs the kids love.

7.1 Xylophone Modification

7.1.1 Components Selection

A. Piezo Vibration Sensor: The LDT0-028K is a flexible component comprising a 28 μm thick piezoelectric PVDF polymer film with screen-printed Ag-ink electrodes, laminated to a 0.125 mm polyester substrate, and fitted with two crimped contacts. As the piezo film is uprooted from the mechanical unbiased pivot, bowing makes high strain inside the piezopolymer, in this manner high voltages are created. At the point when the get together is avoided by direct contact, the gadget goes about as an adaptable "switch", and the created yield is adequate to trigger MOSFET or CMOS arranges legitimately. If the assembly is supported by its contacts and left to vibrate "in free space" (with the inertia of the clamped/free beam creating bending stress), the device will behave as an accelerometer or vibration sensor. Increasing mass, or adjusting the free length of the component by clipping, can change the thunderous recurrence and affectability of the sensor to suit explicit applications. Multi-axis reaction can be accomplished by situating the mass askew. The LDTM-028K is a vibration sensor where the detecting component contains a cantilever bar stacked by an extra mass to offer high affectability at low frequencies. Figure 7.1 shows the schematic of a piezo vibration sensor and Figure 7.2 shows how it looks like.

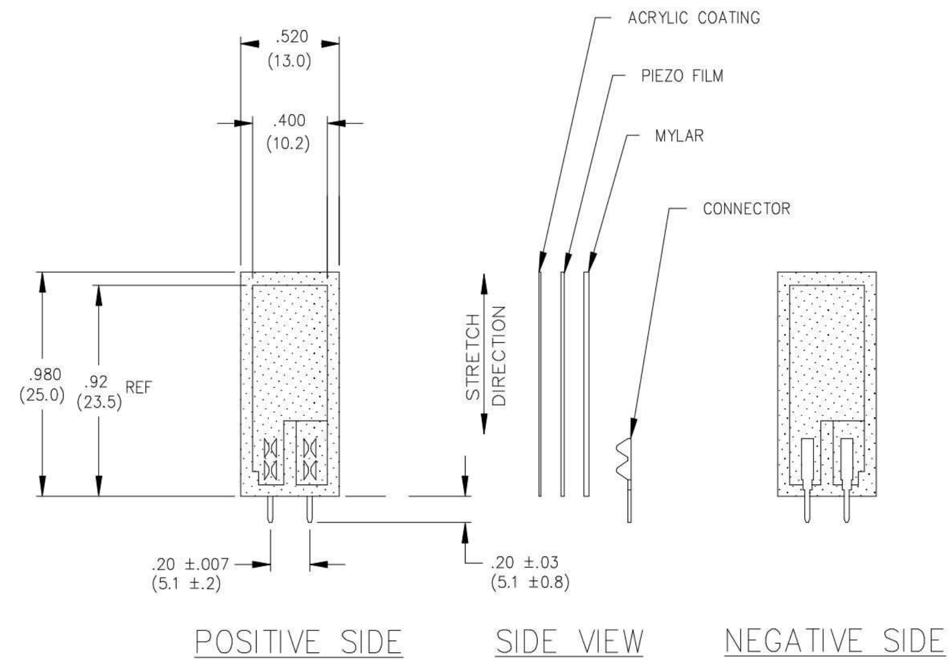


Figure 7.1: *Piezo Sensor Schematic*

B. Op-Amp: An operational amplifier (often op-amp or opamp) is a DC-coupled high-gain electronic voltage amplifier with a differential input and, usually, a single-ended output. In this configuration, an op-amp produces an output potential (relative to circuit ground) that is typically hundreds of thousands of times larger than the potential difference between its input terminals. Operational amplifiers had their origins in analog computers, where they were used to perform mathematical operations in many linear, non-linear, and frequency-dependent circuits.

The popularity of the op-amp as a building block in analog circuits is due to its versatility. By using negative feedback, the characteristics of an op-amp circuit, its gain, input and output impedance, bandwidth etc., are determined by external components and have little dependence on temperature coefficients or engineering

Figure 7.2: *Piezo Sensor VS A Quarter*

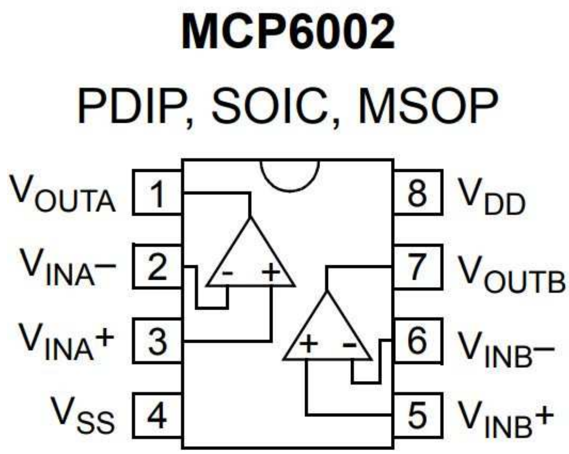


Figure 7.3: *Schematic of Op-Amp MCP 6002*

tolerance in the op-amp itself.

Op-amps are among the most widely used electronic devices today, being used in a vast array of consumer, industrial, and scientific devices. Many standard IC op-amps cost only a few cents in moderate production volume; however, some integrated or hybrid operational amplifiers with special performance specifications may cost over US 100 in small quantities. Op-amps may be packaged as components or used as elements of more complex integrated circuits. Figure 7.3 shows the schematic of MCP6002 IC.

The op-amp is one type of differential amplifier. Other types of differential amplifiers include the fully differential amplifier (similar to the op-amp, but with two outputs), the instrumentation amplifier (usually built from three op-amps), the isolation amplifier (similar to the instrumentation amplifier, but with tolerance to common-mode voltages that would destroy an ordinary op-amp), and negative-feedback amplifier (usually built from one or more op-amps and a resistive feedback network). Figure 7.4

C. Multiplexer: In electronics, a multiplexer (or mux) is a device that selects between several analog or digital input signals and forwards it to a single output line. A multiplexer of $2^n 2^n$ inputs has n n select lines, which are used to select which input line to send to the output. Multiplexers are mainly used to increase the amount of data that can be sent over the network within a certain amount of time and bandwidth. A multiplexer is also called a data selector. Multiplexers can also be used to implement Boolean functions of multiple variables.

An electronic multiplexer makes it possible for several signals to share one device or resource, for example, one A/D converter or one communication line, instead of having one device per input signal.

Conversely, a demultiplexer (or demux) is a device taking a single input and selecting signals of the output of the compatible mux, which is connected to the single input, and a shared selection line. A multiplexer is often used with a complementary demultiplexer on the receiving end.

An electronic multiplexer can be considered as a multiple-input, single-output

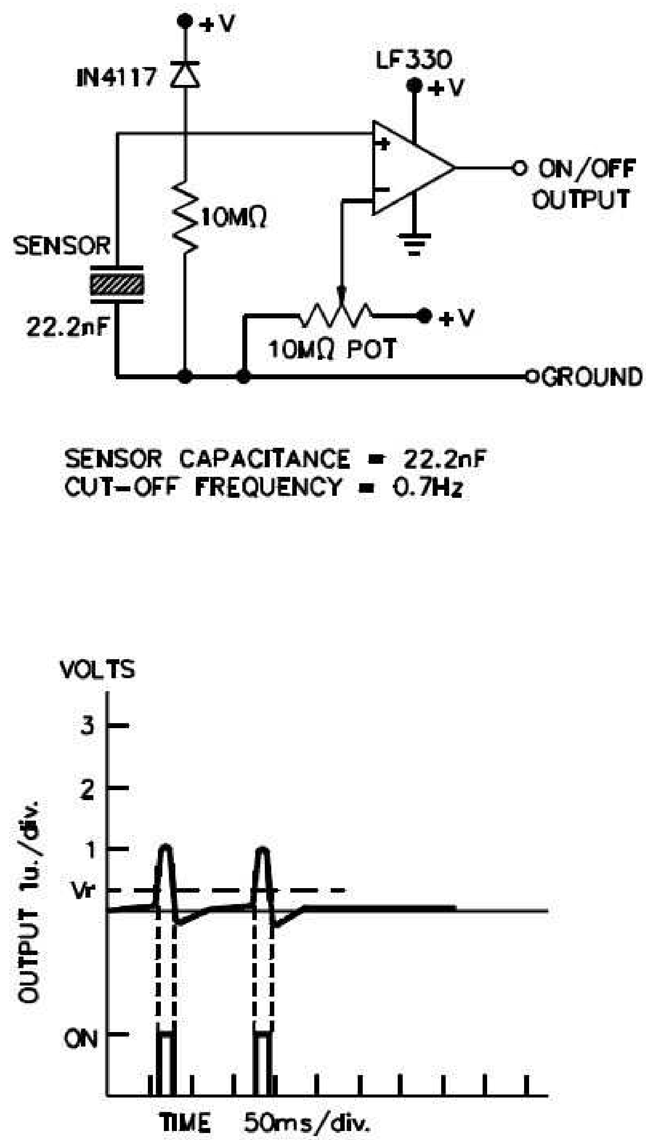


Figure 7.4: *Schematic of Piezo Sensor Application as A Switch*

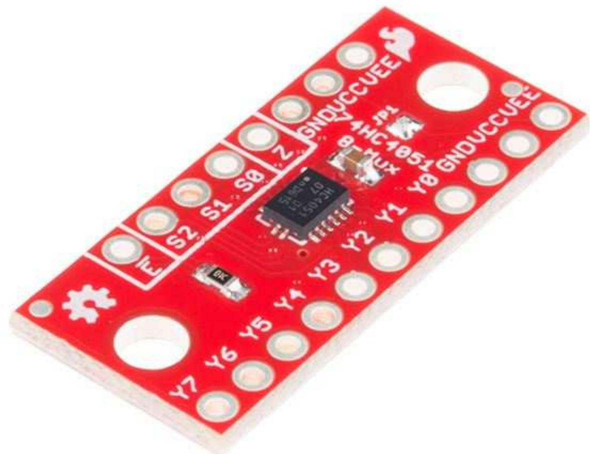


Figure 7.5: *SparkFun Multiplexer Breakout 8 Channel (74HC4051)*

switch, and a demultiplexer as a single-input, multiple-output switch. The schematic symbol for a multiplexer is an isosceles trapezoid with the longer parallel side containing the input pins and the short parallel side containing the output pin. The schematic on the right shows a 2-to-1 multiplexer on the left and an equivalent switch on the right. The *sel* sel wire connects the desired input to the output. The 74HC4051; 74HCT4051 is a single-pole octal-throw analog switch (SP8T) suitable for use in analog or digital 8:1 multiplexer/demultiplexer applications. The switch features three digital select inputs (S0, S1 and S2), eight independent inputs/outputs (Yn), a common input/output (Z) and a digital enable input (E). When E is HIGH, the switches are turned off. Inputs include clamp diodes. This enables the use of current limiting resistors to interface inputs to voltages in excess of VCC. Figure 7.5 shows the multiplexer used in this study.

Figure 7.6: *Arduino Uno Microprocessor*

D. Arduino UNO: The Arduino Uno see Figure 7.6 is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc. The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. The board has 14 Digital pins, 6 Analog pins, and is programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable. It can be powered by the USB cable or by an external 9-volt battery, though it accepts voltages between 7 and 20 volts. It is also similar to the Arduino Nano and Leonardo. The hardware reference design is distributed under a Creative Commons Attribution Share-Alike 2.5 license and is available on the Arduino website. Layout and production files for some versions of the hardware are also available.

The word "uno" means "one" in Italian and was chosen to mark the initial release of the Arduino Software. The Uno board is the first in a series of USB-based Arduino boards, and it and version 1.0 of the Arduino IDE were the reference versions of Arduino, now evolved to newer releases. The ATmega328 on the board comes preprogrammed with a bootloader that allows uploading new code to it without the use of an external hardware programmer.

While the Uno communicates using the original STK500 protocol, it differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it uses the Atmega16U2 (Atmega8U2 up to version R2) programmed as a USB-to-serial converter.

7.1.2 Circuit Design

Figure 7.7 shows the schematic of circuit.

7.2 Sound Design

7.2.1 ChucK: A On-the-fly Audio Programming Language

The computer has long been considered an extremely attractive tool for creating, manipulating, and analyzing sound. Its precision, possibilities for new timbres, and potential for fantastical automation make it a compelling platform for expression and experimentation - but only to the extent that we are able to express to the computer what to do, and how to do it. To this end, the programming language has perhaps

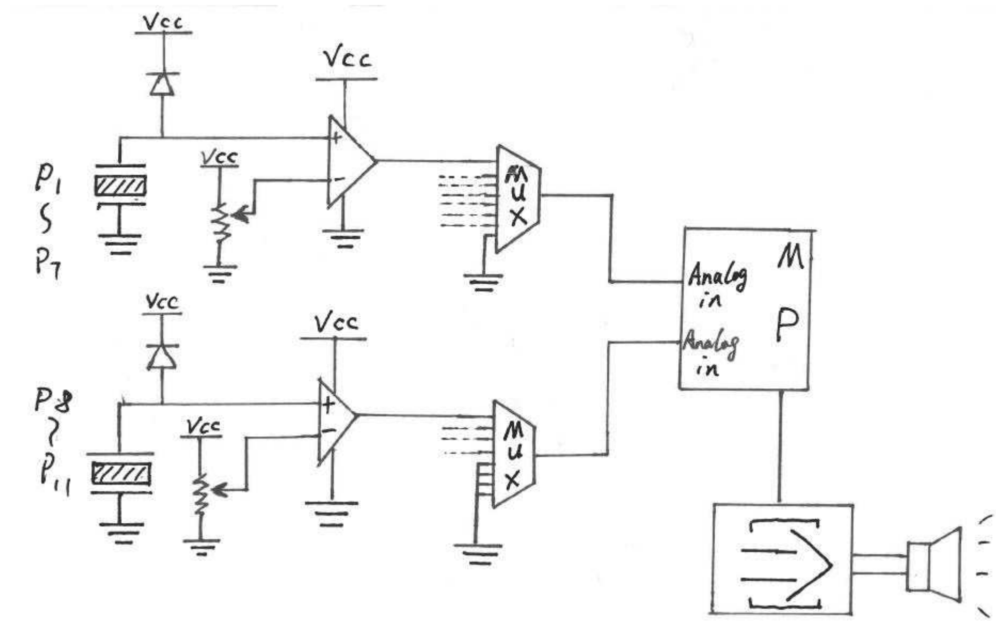


Figure 7.7: *The complete design of the new xylophone circuit.*

served as the most general, and yet most precise and intimate interface between humans and computers. Furthermore, "domain-specific" languages can bring additional expressiveness, conciseness, and perhaps even different ways of thinking to their users.

Figure 7.8 shows the flow chart for music software design. This design allows user to switch keys between different scales and Major/minor in order to create emotional music.

7.3 New Style Session Proposal

This newly designed electric xylophone has two major difference comparing to the original accoustic one. The most obvious improvment is the sound. Various timbers,

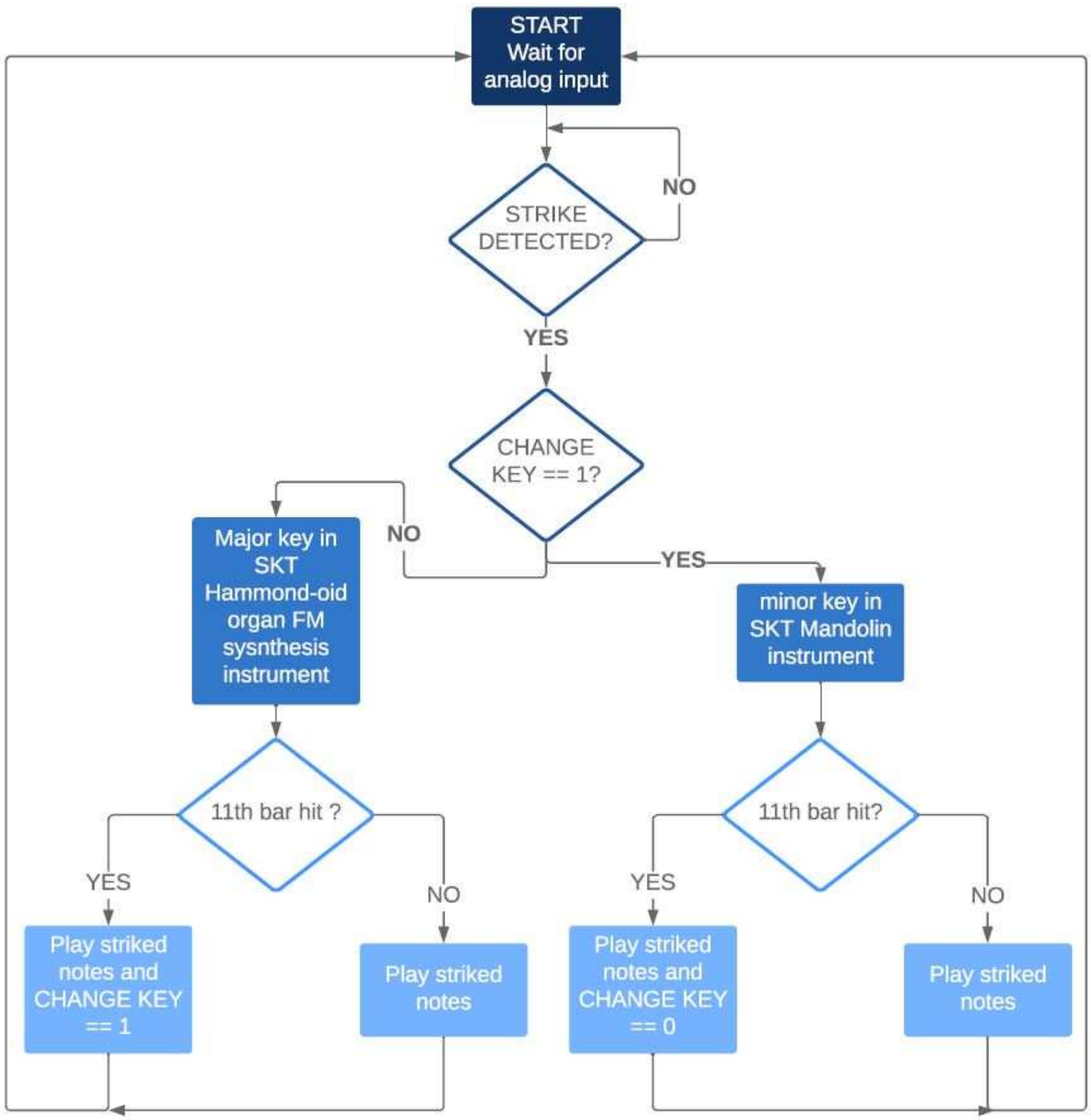


Figure 7.8: *ChucK Flowchart*

keys and scales can be programmed on the board and switched in real-time which provides the infinite possible with limited note bars. Gental touch also been embeded in the play style. Previously only proper motor control can strike a nice melody, by using this new design, one soft touch of the fine tuned bar will also provide a nice and clean note out of the speaker. These two designed provides unlimited possible of music play feature in this music teaching platform we proposed.

Bibliography

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] Conati C. Amershi, S. and H. McLaren. Using feature selection and unsupervised clustering to identify affective expressions in educational games. In *In Workshop in Motivational and Affective Issues in ITS, 8th International Conference on Intelligent Tutoring Systems*, 2006.

[3] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: Dsm-iv. 2000.

[4] K. Dautenhahn B. Robins and J. Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies*, 7(3), 2006.

[5] Rutter M. Bartak, L. and A Cox. A comparative study of infantile autism and specific developmental receptive language disorder. *British Journal of Psychiatry*, 126, 1975.

[6] 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.

[7] 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.

[8] 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.

[9] J. Heemsbergen H. Jordan L. Mawhood C. Lord, M.Rutter S. Goode and E. Schopler. Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, 9(2), 1989.

[10] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011.

[11] D. W. Churchill and C. Q. Bryson. Looking and approach behavior of psychotic and normal children as a function of adult attention and preoccupation. *Comparative Psychiatry*, 13, 1972.

[12] 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.

[13] C. Cortes and V Vapnik. Support-vector networks. *Machine learning*, 20(3), 1995.

[14] 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.

[15] K. Dautenhahn and I. Werry. Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmatics and Cognition*, 12(1), 2004.

[16] N. Edwards and A. Beck. Animal-assisted therapy and nutrition in alzheimer's disease. *Western Journal of Nursing Research*, 24(6), October 2002.

[17] D. Feil-Seifer. and M. J. Mataric. Defining socially assistive robotics. 2005.

[18] 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.

[19] 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.

[20] 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.

[21] Conway R. Leonard M. Meagher D. Godfrey, A. and G. M Ólaighin. A continuous wavelet transform and classification method for delirium motoric subtyping. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(3), 2009.

[22] Hosein M Golshan, Adam O Hebb, Sara J Hanrahan, Joshua Nedrud, and Mohammad H Mahoor. A multiple kernel learning approach for human behavioral task classification using stn-lfp signal. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1030–1033. IEEE, 2016.

[23] Hosein M Golshan, Adam O Hebb, Sara J Hanrahan, Joshua Nedrud, and Mohammad H Mahoor. An fft-based synchronization approach to recognize human

behaviors using stn-lfp signal. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 979–983. IEEE, 2017.

[24] Hosein M Golshan, Adam O Hebb, Sara J Hanrahan, Joshua Nedrud, and Mohammad H Mahoor. A hierarchical structure for human behavior classification using stn local field potentials. *Journal of neuroscience methods*, 293:254–263, 2018.

[25] Valenza G. Citi L. Greco, A. and E. P Scilingo. Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*, 17(3), 2017.

[26] C. Hutt and M. J. Vaizey. Differential effects of group density on social behavior. *Nature (London)*, 209, 1966.

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

[28] Park B. J. Kim S. H. Jang, E. H. and J. H. Sohn. Emotion classification based on physiological signals induced by negative emotions: Discriminantion of negative emotions by machine learning algorithm. In *In Networking, Sensing and Control (ICNSC), 2012 9th IEEE International Conference*, 2012.

[29] Park B. J. Kim S. H. Chung M. A. Park M. S. Jang, E. H. and J. H Sohn. Emotion classification based on bio-signals emotion recognition using machine learning algorithms. In *In Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference*, 2014.

[30] 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.

[31] T. Saito K. Wada, T. Shibata and K. Tanie. Analysis of factors that bring mental effects to elderly people in robot assisted activity. October 2002.

[32] L Kanner. Autistic distrubances of affective contact. *Nervous Child*, 2, 1943.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] Belda-Lois J. M. Medina E. Campos N. Laparra-Hernández, J. and R Poveda. Emg and gsr signals for evaluating user's perception of different types of ceramic flooring. *International Journal of Industrial Ergonomics*, 39(2), 2009.

[38] Messinger D. S. Kermol E. Legisa, J. and L Marlier. Emotional responses to odors in children with high-functioning autism: autonomic arousal, facial behavior and self-report. *Journal of autism and developmental disorders*, 43(4), 2013.

[39] Jie Li, Liqing Zhang, Dacheng Tao, Han Sun, and Qibin Zhao. A prior neurophysiologic knowledge free tensor-based scheme for single trial eeg classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(2):107–115, 2008.

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

[41] 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.

[42] Wang C. H. Wu T. L. Jeng S. K. Lin, Y. P. and J. H. Chen. Multilayer perceptron for eeg signal classification during listening to emotional music. In *In TENCON 2007-2007 IEEE Region 10 Conference*, 2007.

[43] Liu G. Liu, F. and X Lai. Emotional intensity evaluation method based on galvanic skin response signal. In *In Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium*, 2014.

[44] Mingyang Liu, Di Fan, Xiaohan Zhang, and Xiaopeng Gong. Human emotion recognition based on galvanic skin response signal feature selection and svm. In *2016 International Conference on Smart City and Systems Engineering (ICSCSE)*, pages 157–160. IEEE, 2016.

[45] V Lotter. Epidemiology of autistic conditions in young children. *Social Psychiatry*, 20, 1966.

[46] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on pattern analysis and machine intelligence*, 11(7), 1989.

[47] 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.

[48] Kazuya Mera and Takumi Ichimura. Emotion analyzing method using physiological state. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 195–201. Springer, 2004.

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

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

[51] 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.

[52] 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.

[53] Aminian K. Paraschiv-Ionescu A. Loew F. Bula C. J. Najafi, B. and P. Robert. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *IEEE Transactions on biomedical Engineering*, 50(6), 2003.

[54] Alvarez K. Lisetti-C. L. Nasoz, F. and N Finkelstein. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology and Work*, 6(1), 2004.

[55] Winkielman P. Oberman, L. M. and V. S Ramachandran. Slow echo: Facial emg evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Developmental Science*, 12(4), 2009.

[56] L. Ann Obringer and S. Jonathan. Honda asimo robot. July 2011.

[57] OY Ousley, R Arriaga, GD Abowd, and M Morrier. Rapid assessment of social-communicative abilities in infants at risk for autism. *Technical Report, Georgia Tech*, 2012.

[58] 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.

[59] Mari M.-Lusher D. Pierno, A. C. and U Castiello. Robotic movement elicits visuomotor priming in children with autism.

[60] Mari M.-Lusher D. Pierno, A. C. and U Castiello. Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46.

[61] K. Popper. The logic of scientific discovery. 1959.

[62] Vilayanur S. Ramachandran and Lindsay M. Oberman. Broken mirrors: A theory of autism. *Scientific American*, 295(5), 2006.

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

[64] 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.

[65] James Rehg, Gregory Abowd, Agata Rozga, Mario Romero, Mark Clements, Stan Sclaroff, Irfan Essa, O Ousley, Yin Li, Chanho Kim, et al. Decoding children's social behavior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3414–3421, 2013.

[66] 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.

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

[68] A. Sano and R. W. Picard. Toward a taxonomy of autonomic sleep patterns with electrodermal activity. In *In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011.

[69] S. Schmidt and H Walach. Electrodermal activity (eda)-state-of-the-art measurement and techniques for parapsychological purposes. *the Journal of Parapsychology*, 64(2), 2000.

[70] 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.

[71] Prakash N. R. Sharma, V. and P. Kalra. Eda wavelet features as social anxiety disorder (sad) estimator in adolescent females. In *In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference*, 2016.

[72] Omitz E. M. Brown N. B. Sorosky, A. D. and E. R. Ritvo. Systematic observations of autistic behavior. *Archives of General Psychiatry*, 18, 1968.

[73] 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.

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

[75] M. Swangnetr and D. B. Kaber. Emotional state classification in patient-robot interaction using wavelet analysis and statistics-based feature selection. *IEEE Transactions on Human-Machine Systems*, 43(1), 2013.

[76] Manida Swangnetr and David B Kaber. Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection. *IEEE Transactions on Human-Machine Systems*, 43(1):63–75, 2012.

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

[78] 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.

[79] 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.

[80] Diehl J.J. Villano. M. Wier K. Thomas B. Shea N. et al Tang, K. Enhancing empirically-supported treatments for autism spectrum disorders: A case study using an interactive robot. 2011.

[81] 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.

[82] Crowell C. R. Wier K. Tang K. Thomas B. Shea N. et al. Vallano, M. Domer: A wizard of oz interface for using interactive robots to scaffold social skills for children with autism spectrum disorders. 2011.

[83] M. Vetterli and C Herley. Wavelets and filter banks: Theory and design. *IEEE transactions on signal processing*, 40(9), 1992.

[84] 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.

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

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

[87] S. Wolff and S Chess. A behavioral study of schizophrenic children. *Acta Psychiatrica of Child Psychology and Psychiatry*, 40, 1964.

[88] Fu Y. Jiang Y. G. Li B. Xu, B. and L Sigal. Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization. *IEEE Transactions on Affective Computing*, 2016.

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

VITA

Huanghao Feng was born in Suzhou, Jiangsu Province, China, on August 1, 1986. He received his elementary education at Pingzhi Central Elementary School, his secondary education at Suzhou Lida Middle School, and his high school education at Suzhou NO.1 High School. In September 2005, he was admitted to the Suzhou University of Science and Technology in Suzhou, China from which he was graduated with the B.S. degree in Electrical and Communication Engineering with Best Senior Design Honor in June 2011. He continued his graduate studies in University of Denver and was awarded M.S. degree in Electrical and Computer Engineering in June 2014.

In August 2015, he was admitted to the Graduate School of the University of Denver, where he was granted the degree of Doctor of Philosophy in Electrical and Computer Engineering in August 2020.

Permanent Address: 118 Jiuhua Rd. Haishang Yipin Community, Building #25, Room #601, Suzhou, Jiangsu, 215027, China