UNIVERSITY OF DENVER

XYLO-BOT: A THERAPEUTIC ROBOT-BASED MUSIC PLATFORM FOR CHILDREN WITH AUTISM

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

July 2020

UNIVERSITY OF DENVER

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

XYLO-BOT: A THERAPEUTIC ROBOT-BASED MUSIC PLATFORM FOR CHILDREN WITH AUTISM

Huanghao Feng

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

Abstract of a dissertation at the University of Denver.

Dissertation supervised by Dr. Mohammad Mahoor.
No. of pages in text 161.

Children with Autism Spectrum Disorder (ASD) experience deficits in verbal and nonverbal communication skills, including motor control, emotional facial expressions, and eye gaze / joint attention. This Ph.D. dissertation focuses on studying the feasibility and effectiveness of using a social robot, called NAO, and a toy music instrument, xylophone, at modeling and improving the social responses and behaviors of children with ASD. In our investigation, we designed an autonomous social interactive music teaching system to fulfill this mission. A novel modular robot-music teaching system consisting of three modules is presented. Module 1 provides an autonomous self-awareness positioning system for the robot to localize the instrument and make a micro adjustment for the arm joints to play the note bars properly. Module 2 allows the robot to be able to play any customized song per user's request. This design provides an opportunity to translate songs into C-major or a-minor with a set of hexadecimal numbers, allowing a person who has less music experience to complete this work. After the music score converted robot should be able to play it immediately. Module 3 is designed for providing real-life music teaching experience for the users. Two key features of this module are a) "music detection" and b) "smart scoring and feedback". Short-time Fourier transform and Levenshtein distance are adapted to fulfill the design requirements, which allow the robot to understand

music and provide a proper dosage of practice and oral feedback to users. A new instrument has designed to present better emotions from music due to the limitation of the original xylophone. This new programmable xylophone can provide a more extensive frequency range of notes, easily switch between the Major and Minor keys, extensively easy to control, and have fun with it as an advanced music instrument.

Because our initial intention has been to study emotion in children with autism, 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 to the recorded EDA signals. The dataset used in this research 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 primary 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 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.

NAO music education platform will be thought-about as a decent tool to facilitate improving fine motor control, turn-taking skills, and social activities engagement. 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 early events. More than half of the subjects could dominate proper turn-taking after few sessions. Music teaching could be a good example for accomplishing social skill tasks by taking advantage of customized songs selected by individuals. According to the session executioner and video annotators, this particular subject shows a high level of engagement for all activities, including free play. Based on the conversation and music performance with the robot, the subject showed a strong interest in challenging the robot in a friendly way.

To my beloved mother,

and

to my friends.

ACKNOWLEDGMENTS

No matter how many, words are pale compare to such friendship between Dr. Mahoor and I. Without his help and support, I might have been lost in my life and looking for something that I do not even know. However, with his help, I found my path and can roam in the scientific field, pursue novel ideas, and seek solutions. I cannot forget the very first day when Dr. Mahoor introduced me to NAO, the robot who I worked with. Thanks to Dr. Mahoor, who provided me with such and exceptional experience and helped me gain knowledge in the filed of Electrical and Computer Engineering and particularly Robotics!

More than everyone, I indebted to my mother for her enthusiastic encouragement, prayers, and unlimited support during all stages of my life. Understanding and tolerance helped us while argues and fights happened between us. I could not accomplish this work without her selfless love and assistance. Also, I must express my appreciation to my girlfriend, who stands beside me to conquer the very dark time.

In the past eight years, I have been working with these special children and I cannot stop thinking of their future life after ten years or even more. The honor goes to my late friend, Mr. Masoud Bahramisharif, without his help, I was not able to even begin this project. Mr. Hosein M Golshan provided his endless support for this fantastic work as well, which helps me to complete it on time. I also want to send my appreciation to all the committee members, Dr. Kimon Valavanis, Dr. Tim Sweeney, and Dr. Matthew Rutherford. I would thank them all for their support and serving on my defense committee.

I again would like to thank my advisor Dr. Mahoor, for providing me such a unique opportunity to work with NAO and supporting me for all these years working on my dissertation. Meanwhile, I want to thank the supports and help I have received from Dr. Anibal Gutierrez from the University of Miami and also Dr. Tim Sweeney and Ms. Francesca Dino from the 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.

Extended thanks to the faculty and staff members of the Electrical and Computer Engineering Department at the University of Denver, especially Ms. Molly Dunn and Ms. Drusell May, for their role in my education and various resources made available to me with my work. Also, I would like to thank my fellow lab members, especially, Dan Stoianovici, Hojjat Abdollahi, Rohola Zandie, Behzad Hasani, and Jian Sun, for their collaboration during the completion of this dissertation.

Furthermore, I want to thank Dr. Amin Khodaei, Dr. Gonçalo Martins, and Dean J.B. Holston for their support and help during my study at the University of Denver. Being a student and a researcher at DU is the best luck of my life!

Contents

List of Figures

List of Tables

List of Acronyms

1-D: One Demensional

ABA: Applied Behavior Analysis

ACG: Anime, Comic and Games

ADDM: Autism and Developmental Disabilities Monitoring

AI: Artificial Intelligence

ANOVA: Analysis of Variance

ANN: Artificial Neural Networks

API: Application Programming Interface

AS: Asperger's Syndrome

ASD: Autism Spectrum Disorders

AUC: Area under the ROC Curve

C-Morlet: Complex Morlet

CART: Classification and Regression Tree

CDC: Centers for Disease Control and Prevention

CSL: Child Study Lab

CWT: Continuous Wavelet Transform

DD: Developmental Delay

DTT: Discrete Trial Teaching

DWT: Discrete Wavelet Transform

ECG: Electrocardiography

EDA: Electrodermal Activity

EEG: Electroencephalography

EMG: Electromyography

FM: Fine Motor

GM: Gross Motor

HFA: High Functioning Autism

HPOT: Hippotherapy

HR: Heart Rate

JARS: Joint Action Routines

KNN: K-nearest Neighborhood

LDA: Linear Discriminant Analysis

LFP: Local Field Potential

LOOCV: Leave-One-Out Cross Validation

MLP: Multilayer Perceptron Network

MMDB: Multi-modal Dyadic Behavior

NIICT: National Institute of Information and Communications Technology

OH: Optimal Hyperplane

PCA: Principal Component Analysis

PDMS-2: Peabody Developmental Motor Scales, Second Edition

PPG: Photoplethysmogram

PRT: Pivotal Response Training

QP: Quadratic Programming

RBF: Radial Basis Function

RGB: Red,Green and Blue

RPMT: Responsive Education and Prelinguistic Milieu Teaching

SAD: Social Anxiety Disorder

SAR: Socially Assistive Robotics

SC: Skin Conductance

SIR: SOcially Interactive Robotics

SKT: Skin Temperature

STFT: Short Time Fourier Transform

SVM: support vector machine

TD: Typically Developing

TEACCH: Treatment in Education of Autistic and Related Communication Handicapped Children

Chapter 1

Introduction

1.1 Autism Spectrum Disorders (ASD)

Autism, defined on abnormal development of behavioral criteria such as social interaction, communication, and imagination, is considered as a neurodevelopmental disorder (Kanner syndrome) [1, 2]. Usually, autism could start at an early age, like infancy, at the latest, in the first three years of life. Not using words to communicate can be the first clue for the parents to be noticed, even the kid be able to repeat messages from videotapes or speaks the alphabet. Social deficits may not be seen immediately at an early age of childhood. However, it will gradually be noticed while other children become more socially sophisticated and more active. Children with autism usually do not have meaningful communication with others, even when they have to. As age increase, some of the repetitive behaviors begin to develop, for example, specific hand and finger movements, using peripheral vision to look at objects, or forward and backward body shaking [3]. Children with ASD could also experience deficits in inappropriate verbal and nonverbal communication skills, including motor control, emotional facial expressions, and eye gaze attention [4]. About 1 in 54 children have been identified with ASD according to estimates from CDC's Autism and Developmental Disabilities Monitoring (ADDM) Network and government statistics suggest the prevalence rate of ASD is increasing 10-17 percent annually [5].

It is often difficult for parents and professionals to recognize and judge the scientific validity of an intervention or treatment designed to be used with individuals with ASD. National Research Council includes a list of the features the committee believes to be successful educational intervention services for ASD children. The components include: early age entry into an intervention program; active engagement in intensive instructional programming for the equivalent of a full school day, including services that may be offered in different sites, for a minimum of five days a week with full-year programming; use of planned teaching opportunities, organized around relatively brief periods for the youngest children (e.g., 15-20 minute intervals); and sufficient amounts of adult attention in one-to-one or minimal group instruction to meet individualized goals. [6] Multiple treatments for ASD population can be categories as follows: (1) interpersonal relationship, (2) skill-based, (3) cognitive, (4) physiological/biological/neurological, and (5) other interventions and treatments. [7]

Some of the treatments have been proven that has significant and convincing support for ASD children, such as Applied Behavior Analysis (ABA) [8], Discrete Trial Teaching (DTT) [9], and Pivotal Response Training (PRT) [10]. Currently, ABA

[11, 12] has focused on teaching individuals with ASD appropriate social skills in an effort to make them more successful in social situations [13]. 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 on how to design and use modern technology that would result in clinically robust methodologies for autism intervention is vital. Assistive Technology [14], Joint Action Routines (JARS) [15], Cognitive Behavioral Modification [16], Structured Teaching [17], and Social Stories [18, 19, 20], such intervention and treatments also provide promising results for most of the cases, even though these methods still requires additional scientific support in the future. [7]

In human social 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 their 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 [21]. 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 in understanding better, model and improve the social skills of individuals on the autism spectrum.

With the rise in the prevalence of autism, the number of therapies for this condition has correspondingly increased. In general, practitioners accept the need for appropriate treatments. Effectiveness is usually thought to mean the use of reliable research with precise control over internal and external challenges to validity. Therefore, only therapies with constant clinical support that show effectiveness in alleviating negative autism symptomology should be widely disseminated for use. There are, however, many fad therapies that have no such evidence of efficacy. Use these therapies is wasting time and resources and preying on parents' and caregivers' emotional weakness. [22] Computer technology is expected to be increasingly used by a new generation of children in a variety of contexts (professional, educational and recreational), including interactive robotic toys, digitally enhanced objects , and tangible interfaces [23, 24, 25, 26]. Modern digital technologies and modern implementations are also vulnerable to affect therapy and recovery methods. The physical structure and behavior of socially intelligent agents, demonstrating facets of social intelligence in the human form [27], are likely to alter how we can teach social intelligence to people who have trouble recognizing and expressing social behaviour.

A robotic platform is hoped to provide the necessary stimulation to reinforce the child's responses according to Treatment in Education of Autistic and Related Communication Handicapped Children (TEACCH) treatment method. This should promote interaction by providing a pleasant stimulus, strengthening it by reacting in specific, non-threatening ways. A robot is expected to allow the child to relax and view the activity as play, reducing the amount of fear presented. It should, therefore, appear to be a new and interesting toy, while at the same time extending the interactive and communicative limits of the individual child through a playable medium. Bridging the gap between the inner world of autism and the unpredictable yet appropriate teacher, thereby offering a stable method of educating the child about the fundamentals of interaction in a gradual manner and adapting to the child's development should also be done by the robot platform [28].

This dissertation presents the methodology and results of a study that aimed to design a autonomous human-robot interaction education platform for capturing, modeling, and enhancing the social skills of children with autism. Such a platform should complete the following requirements: (1) fully autonomous to conduct an intervention session, (2) provide a life-like teaching-learning environment scenario, (3) in particular aiming motor control and turn-taking skills improvement, (4) stimulate emotional change in different social activities, and (5) be able to investigate how ASD and Typically Developing (TD) children react to such an education platform with a humanoid robot. In the following section, a brief introduction of existing assistive robots that have been used in autism applications will be introduced.

1.2 Socially Assistive Robotics (SAR)

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 [29]. SAR includes all the characteristics of the SIR mentioned in it [21], as well as a few additional attributes such as 1) user populations (e.g., elders; individuals with physical impairments; kids diagnosed with ASD; students); 2) social skills (e.g., speech ability; gestures movement); 3) objective tasks (e.g., tutoring; physical therapy; daily life assistance); 4) robot function (depends on the task the robot has been assigned for) [29]. Companion robots [30] is one type of SAR that is widely used for older adults for health care supports. Research shows that this type of social robot can reduce the stress and depression of individuals in the elderly stage [31]. Service social robots are able to accomplish a variety of tasks for individuals with physical impairments [32]. 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 [4, 11, 12].

Nowadays, there are very few companies that have designed and developed socially beneficial robots. The majority of existing SARs are not yet commercialized, and because they are expensive and not well-designed user interfaces, they are mostly used for research purposes. Honda, SoftBank Robotics and Hanson Robokind are the leading companies that are currently developing humanoid robots. Ideally, socially helpful robots can have fully automated systems for detecting and expressing social behavior while interacting with humans. Some of the existing robot-human interfaces are semi-autonomous and can recognize some basic biometrics (e.g., user visual and audio commands) and behavioral responses. In addition, most of the existing robots are very complicated to work with. As a result, in the last few years, several companies have begun to make these robots more user-friendly and responsive to both user needs and potential caregivers' commands. [29]. In all, service social robots are able to do a variety of tasks for individuals with physical impairments. SAR can be used in therapy sessions for those individuals who have autism. SAR provides an efficient, helpful medium to teach certain types of skills to these groups of individuals.

Intelligent SARs strive at being able to understand visual or auditory instructions, objects, and basic human movements. Any of these robots have the power to identify human faces or simple facial expressions. For example, ASIMO, a robot created by Honda, the company, has a sensor for detecting movements of multiple objects using visual information obtained from two cameras on its head. Besides, its "eyes" will determine the distance between objects and robots. [33] Another example is Softbank Robotics, which builds small-scale humanoid robots called the NAO. The NAO robot has two cameras mounted to it that are used to take single photographs and video sequences. This capture module enables NAO to see and recognize the different sides of an object for future use. Besides, NAO has a remarkable ability to recognize faces and to detect moving objects. More details will be discussed in the following chapters. The speech recognition system has been embedded in both of the aforementioned robots, which provide a strong voice communication ability to accomplish more natural social interaction with human beings. NAO is able to understand words and sentences which have been pre-programmed in the memory for running specific commands. However, ASIMO is able to distinguish between voices and other sounds. This feature empowers ASIMO to perceive the direction of a human's speaker or recognize other companion robots by tracking their voice [34]. Several language packages can be installed into NAO, which feature gives the robot a strong social communication functionality to interact world widely.

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 the engagement of users as communicating with robots, and elicit novel social behaviors through their interaction. One of the goals in SAR is to use social robots either individually or in conjunction with caregivers to improve the 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 inappropriate 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 [13]. 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 [35, 36, 29, 21]. 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 practices. These investigations propose that interaction with robots may be a promising approach for rehabilitation of children with ASD.

Several research groups 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., [21]), and to utilize robotic systems to develop novel interventions and enhance existing treatments for children with ASD [33, 34, 37]. Mazzei et al. [38], for example, designed the robot "FACE" to show the details realistically of human facial expressions. A combination of hardware, wearable devices, and software algorithms measured the 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 [29, 21] 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, explaining 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 exciting 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 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 [39], haptic interaction [40], and adaptive game interactions based on affective cues inferred from physiological signals [41] 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 limited explored and focused on their core deficits (i.e., Facial expression, eye gaze, and joint attention skills). Bekele and colleagues [42] 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 response in joint attention stimulation. 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 significant social deficits that individuals with autism may have, we have designed and conducted robot-based therapeutic sessions that are focused on different aspects of the social skills of children with autism. In this thesis, we employed NAO, which can autonomously communicate with the children. We conducted two different designs to examine the music social skills of children with autism and provide feedback to improve their behavioral responses.

1.3 Music Therapy in ASD Treatment

Early pioneers in the 1940s, music therapy were used in psychiatric hospitals, institutions, and schools for children with autism. Back in that time, since both autism diagnosis and the music therapy profession were emerging simultaneously, there was no official documentation in such a field can be found. In the 1950s, the apparent unusual musical abilities of children with autism intrigued many music therapists. By the end of the 1960s, music therapists started delineating goals and objectives. The beginning of the 1970s encountered the emergence of theoretically grounded music therapists working toward a more clearly defined approach to improving the lives of children with autism. "A great deal of research needs to be done in many directions. For the present, we have to use whatever approach has some value, and from our experience, there is no doubt, music therapy has value" [43]. However, for decades, music therapists are not using a consistent assessment method with autism spectrum disorder clients. The lack of a quality, universal assessment tool has caused difficulty for music therapists. Music therapists are in danger of activity-based, non-goal driven treatment. Without a common language, it is difficult for music therapy to be recognized as a valid, evidence-based approach [44]. Music therapists have continued to implement many of the techniques of the preceding few decades in recent years, such as music games and singing music as a reinforcement [45, 46]. The spectrum of therapeutic strategies has since been expanded to involve family-based music therapy prescriptive songs and to include clients and parents with music therapy services for use beyond music therapy [47, 48, 49].

In order to deliver a solid music therapy intervention solution with a consistent assessment method with ASD children, a humanoid social assistive robot could be a perfect choice. Many researches show that children with autism have less interest in communicating with humans due to sensing overwhelming issues. A robot with a still face could be a good agent with less intimidating characteristics for helping children with autism. There are also researches show that kids with autism are more attracted to interact with humanoid social robots in daily life [50, 51, 52, 19]. That makes the socially assistive robot a perfect medium for delivering certain therapy methods, such as music therapy. A significant amount of reports suggest that using music as an assistive method, also known as music therapy, for helping individuals with autism can be beneficial. Composed songs and improvisational music therapy have been used as performance strategies in these practices. However, there was limited evidence to support the use of music interventions to conduct social, communicative, and behavioral skills in children with autism at an early age under certain conditions. By listening, singing, playing instruments, and moving, patients can get a feeling for the music. Children's music therapy is performed either in a one-on-one session or a group session. It can help children with communication, attention, and motivation problems as well as behavioral issues [53]. Motivation and emotion are essential to music education, and 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 [54].

1.4 Contributions

The major contributions of this dissertation 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. Afterward, 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 is applied to music teaching/playing therapy intervention for a better understanding of emotional 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 to prevent a minor change of relative position between a musical instrument and robot; module 2: joint trajectory generator to play any meaningful customized melody; and module 3: real-time performance scoring feedback using short-time Fourier transform and Levenshtein distance to provide an autonomous real-time music learning experience.

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

- Proposing 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 Organization

This dissertation is organized as follows: Chapter 2 presents related work related to autism spectrum disorders, emotions classification, music therapy in autism treatment, and social robots in autism therapy. Chapter 3 introduces a wavelet-based feature extraction approach for emotion classification as a pre-study for music interaction emotion recognition. Chapter 4 explains a novel approach in designing the autonomous social interactive robot music teaching system with experimental session design. Chapter 5 illustrates all the experimental results. Finally, Chapter 6 presents X-Elophone, a new instrument for music playing and Chapter 7 concludes the dissertation with some discussions, remarks, and proposed future work.

7

Chapter 2

Related Works in Autism and Robots

2.1 Autism

Verbal and non-verbal communication impairments have often been associated with individuals autism spectrum disorder, who has experience specific deficits including language delay, social communication issues, emotion recognition, and eye gaze attention, etc. Autism is a disorder that appears in infancy [55]. Some of the kids were diagnosed as high-functioning autism, even though, some of the social areas sill can be difficult to them such as (1) fine motor control (e.g., unable to perform precisely handy work), (2) having difficulty in understanding emotions from others (e.g., no empathy skills or not be able to read/perform proper facial expressions) and more remarkably, (3) joint attention (e.g., less eye contact and eye gaze attention)[55]. It is known that no single accepted intervention, treatment, or cure for ASDs; however, a successful treatment and better recovery would be performed if intervention been delivered in early diagnosis stages. Almost no clue can be found at a very early stage of individuals with autism; however, signs may emerge after trying to interact with them for a certain period of time. The first thing that may be noticed is not responding by calling their names, while communicating eye contact may not present as well. Repetitive abnormal body movement may also appear, for example, body rocking overtimes or head banging against the wall, which some of the gestures may hurt them permanently. In the early 1990s, researchers in the University of California at San Diego aimed to find out the connections between autism and the nervous system (i.e., mirror neurons). Mirror neuron [56] 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 a mirror neuron in individuals with ASD [56]. There are several studies that focus on the neurological deficits of individuals with autism and studying their brain activities. Figure 2.1 demonstrates brain activity difference between groups in forward speech [57].

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 effec-

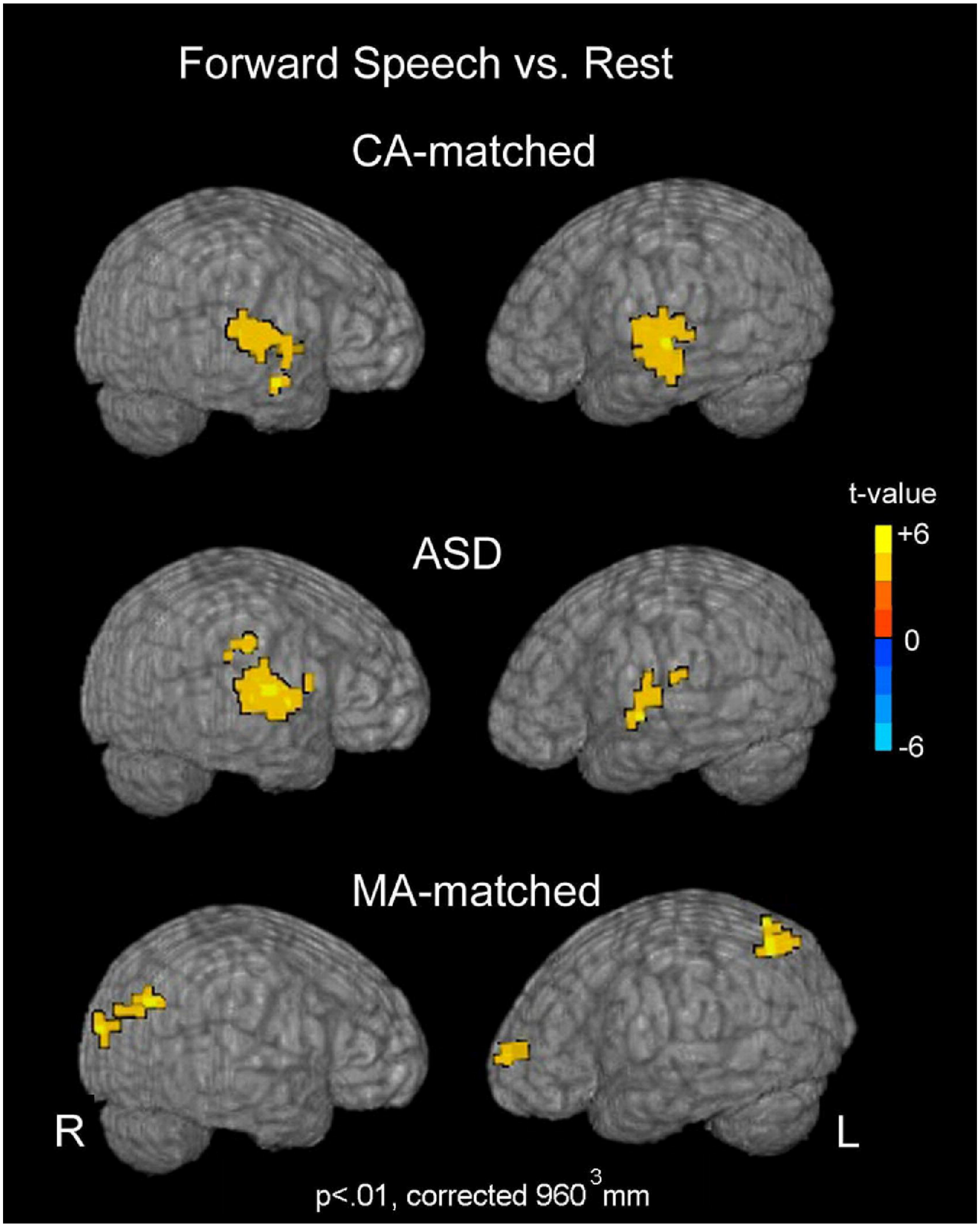


Figure 2.1: *Chronological age-matched, ASD and Mental age-matched brain activities in forward speech.*

tively and recognize emotions behind specific facial expressions or the tone of voice with difficulties [58]. In contrast to TD individuals, children with autism (e.g., 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 other [59].

2.1.1 Motor Control

According to previous researches, some of the impairments seem not defined as core features in ASD, such as motor control or turn-taking skills. However, it has been widely accepted that these skills are nevertheless high prevalent and can have a significant impact on improving social life for individuals with autism [60]. Recent studies show that individuals with autism can be observed with abnormal motor skills in the early age stage [61, 62, 63]. This deficit sticks with them throughout childhood and even in their adulthood as well as [64, 65, 66]. It has been reported that the prevalence of motor skill deficits is between 21 and 100 % [67, 68, 69], which highlighted that motor control problem is a significant but potentially variable aspect of ASD. Research showed that motor ability is correlated with daily living skills in children with autism [70], and in order to decrease the severity of ASDs in their future life which requires better motor control skills for more practice in early age [71]. To this end, increasing the understanding of the etiology of motor deficits in ASD is, therefore, a crucial step towards treating this potential developmental cascade and preventing that [60].

Motor control is systematic movement regulation in organisms that have a nervous system. The motor regulation involves aspects of movement that can be related to reflex [13]. Motor control as a field of study is essentially a psychology or neurology sub-discipline. Recent motor control psychological theories present it as a process through which humans and animals use their brain/cognition to activate and coordinate the muscles and limbs involved in performing a motor skill. Through this mixed psychological viewpoint, motor control is simply the integration of sensory input, both about the environment and the actual state of the body, to decide the correct collection of muscle forces and joint activation to produce any desired movement or motion. This process involves mutual cooperation between the central nervous system and the musculoskeletal system and is , therefore, a question of information processing, communication, dynamics, physics, and cognition [35, 72]. Effective motor control is essential for communicating with the environment, not only deciding capabilities for intervention but also controlling equilibrium and stability. Although the modern motor control analysis is an increasingly interdisciplinary field, research issues have been traditionally described as either physiological or psychological, depending on whether the emphasis is on physical and biological properties, or organizational and systemic rules [36]. Research areas related to motor control include motor synchronization, motor learning, signal processing and the theory of perceptual function.

Hippotherapy (HPOT) [73] as a treatment strategy that uses the horse's movement as a tool to affect functional outcomes in autism therapy. While offering necessary support in challenging the cognitive-sensorimotor system, HPOT also considers the context of the therapy sessions, which makes it a unique treatment strategy for children with ASD [74]. Also, active engagement helps to improve in adaptation and increased willingness to participate in daily activities after the therapy sessions [75], and similar effectiveness in using the horse's movements in the HPOT treatment tool [73]. In [73], children with ASD showed improved postural stability and improvements in receptive communication, coping, and daily activity participation after 12 weekly HPOT sessions. Researchers also claim that ASD kids may possibly pick up automatic postural mechanisms to better adapt to the therapeutic activities due to randomly changed stability from the horse [73]. Gross motor (GM) and fine motor (FM) development in children with ASD has also been studied in [76]. By comparing with children with developmental delay (DD) without ASD, useful results were found. In a total of 38 children, half ASD half DD was assessed using the Peabody Developmental Motor Scales, Second Edition (PDMS-2). Each participant was requested to complete one motor control activity based on his/her levels of skills. Research showed that most of the ASD children had similar levels of GM and FM development, the analogous result also reflects the two groups DD and ASD, they both show identical motor skills development [76].

2.1.2 Turn-Taking Skills

Social communication can be initiated by typically developing kids in their infant stage [77]. Eye contact, initiate turn-taking communicative exchanges , and play tricks with someone who familiar with, such social skills are frequently used by kids as well [78, 79, 80]. However, these skills can be impossible for children with autism, and can be noticed at 7 to 9 months old, according to [81, 82]. Research also found that it is impaired to utilizing appropriate turn-taking interaction skills or fluent verbal interchanges and play turns between partners while communicating for children with ASD [83]. Moreover, these communicative behaviors have been linked to important developmental outcomes in children with ASD [84, 85, 86]. Some researchers have argued that improving turning and initiating joint attention can reduce ASD's severity since social reciprocity is one of the core deficits of autism [87, 88]. It is also found that in difficulty utilizing proper turn-taking behaviors for preschoolers with ASD due to the lack of core social communication skills. This can cause interactive issues with their communicative partners in fluency interchanges or verbal and play turns in daily life [83, 89].

Turn-taking is a form of conversational and dialogue organization, where members talk in alternating turns one at a time. In practice, it includes processes to create inputs, respond to previous comments, and move to another speaker using a variety of linguistic and non-linguistic indications [13]. While the arrangement is substantially uniform, that is, simultaneous talk is usually avoided, and silence is reduced between turns, rules for turn-taking differ by culture and society. In specific ways, norms vary, such as how rolls are handled, how shifts are indicated, or how long the typical distance between turns is [35, 72]. Conversation turns are a desirable way of engaging in social life in many ways, and thus subject to rivalry [36]. Turn-taking approaches are also thought to vary by class; thus, turn-taking has become a topic of intensive analysis in gender studies. Although early work backed gendered assumptions, such as men interrupting more than women and women chatting more than men, recent research has found inconsistent evidence of gender-specific conversational approaches, and few consistent trends have emerged [29, 21].

A core component for targeting turn-taking behaviors of children with autism is early intervention treatment. Few reasons can explain this: (1) the back-and-forth shared structure is considered as a critical framework for early studying, (2) social acceptance in preschoolers are highly connected with turn-taking behaviors [90, 91, 92, 93]. Nonetheless, behavioral approaches to enhance turn-taking habits have scarcely been tested empirically and quantitatively through experiment monitoring as opposed to measures to develop abilities in communicative, emotional, and behavior [94, 95, 93, 96].

One of the therapeutic treatment is called Responsive Education and Prelinguistic Milieu Teaching (RPMT). Unfortunately, it is hard to find publications related to turn-taking on the efficacy of RPMT. However, the RPMT directly teaches object exchange as a means of turn-taking. Past work suggests that the RPMT is successful in promoting interventions of non-autistic children with mixed etiology developmental delays and in encouraging the introduction of joint treatment in children with developmental delays with originally unfortunate introduction of joint attention [97]. While this would appear to bode well for children with ASD, it should be remembered that the children used, even more, facilitating shared focus in this previous study initially than most young children with ASD do. It is not clear that the RPMT facilitates the initiation of joint attention in deficient motivated children to communicate for care or social connection [98]. The effectiveness of LEGO© therapy has been examined by a group of researchers recently [99, 100, 101]. Participants are encouraged to use both verbal and visual input to learn social skills by constructing LEGO © constructs within a community or adult environment. One of the drawbacks of the current literature is that studies have focused on developing cognitive competence in high functioning autism (HFA) or Asperger's syndrome (AS) in school-age children and teenagers. No scientific data confirm the efficacy of therapy for young children with autism spectrum disorders on turn-taking behaviors [102]. Another research has found in examined the turn-taking behaviors in young children with autism is in [93], which tested the effect of different types of turn-taking on language and play skills. Based on the Pivotal Response Training (PRT), Four types of turn-taking skills have been examined. This was found that the turn-taking actions of the educator favorably impacted children with autism 's sensitivity. Specifically, the guiding tools of the teacher and having a predetermined reaction from the subject child were the two main factors that dictated play and language skills development.

2.1.3 Music Therapy

Music is an 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 suffered from mental and behavioral disabilities [103, 104]. Nowadays, at least 12% of all treatment of individuals with autism consists of music-based therapies [105]. Specifically, teaching and playing music to children with autism spectrum disorders (ASD) in therapy sessions have shown a great impact on improving social communication skills [106]. Recorded music or human played back music are used in single and multiple subjects' intervention sessions from many studies [105, 107]. Different social skills are targeted and reported (i.e., eye-gaze attention, joint attention and turn-taking activities) in using music-based therapy sessions [108, 109]. Noted that improving gross and fine motor skills for ASD through music interventions is a missing part of this field of studies [105].

Early affective activity evolves into relationships in the regular boy, where games dominate [110]. An adult who is familiar with the child attempts to engage with him or her through play during the musical interaction therapy. The purpose of musical engagement therapy is to create and improve any sociability the child may have by making music that offers fun opportunities for the child and familiar adults to come together and experience a mutual interest by developing a musical conversation. The accompanying live music strengthens both the actions of the carer and the understanding of that by the infant. Jordan and Libby comment that 'Music is usually helpful to children with autism in that it seems to add both interest and meaning to social situations where they would otherwise be lacking' [111]. The musician is prepared to fill in, support, or enhance the role of either partner in what begins as a preverbal discourse. Within this case, the use of 'service' suggests that the music is part of the connection of both making it more explicit and keeping the series together [112]. The caregiver and musician aim to build a gift-and-take communication experience between the caregiver and the child. Such knowledge may allow the child to communicate with willfulness. The caregiver tries to adapt the volume and pacing of the feedback to the degree of responsiveness of the infant by being attentive to non-verbal signals and facial gestures of the infant [113]. Wimpory and Nash (in press) identified three themes at every stage of their active process that runs through musical interaction therapy. These topics include the scaffolding [114] of caregiver interaction that affords communicative control to the child. The contributions and child's efforts (whether intended or not) provide artistic encouragement from the artist who deals with them on a clinical basis. The musician also provides scaffolding, but in time does so fewer [112].

2.2 Human Robot Interaction in Autism

Children with ASD experience deficits in inappropriate 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 [115]. However, only a few groups used humanoid robots for teaching or practicing social communication skills [116, 38, 117, 118, 119, 120].

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 pieces of evidence suggest that robots can trigger them more effectively than human [121]. Researchers observed that individuals with ASD have more interest in a robot therapeutic partner than a human. In most cases, participants showed better speech and movement imitation compared with the response to a human partner [122]. Although a recent case study [115], 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 [115]. 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 [117], while others have created robots with very mechanical forms [38], and others have developed robots with a cartoonish or animal form [119]. Generally speaking different categories of the robot that has been used for autism research can be placed either into Non-Humanoid and Humanoid robots group [115], which will be explained in the following sections.

Non-Humanoid Robots

Non-humanoid robots are those robots that do not have the same body joint and facial appearance as a human does. It contains those animals 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 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 a bubble from the lower part of robot body), for instance, was not a human form robot and can be built simply [116]. Another non-humanoid robot used by researchers from the University of Hertfordshire called Labo-1 [38], 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.) At Yale University, researchers were using a mobile robotic dinosaur named Pleo, who can show emotions and desires by using its sounds and body movements. 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 more accessible 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. The advanced humanoid robot would be able to move different parts of its body to walk or dance (NAO). Some of the humanoid robots also have 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 in improving their social behaviors.

Robins from the University of Hertfordshire, who is one of the pioneers which employed a study to evaluate the importance of the robot's appearance for autism research. A doll-like robot called Robota was asked to interact with children with autism [117]. 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 fewer details of humans but still hold the humanoid form. So, a robot called KASPAR has been developed by Robins to fit this design criteria [118]. 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 to the mechanical parts of the robot's body than communicating with the robot itself [115]. A small soft snowman-shaped robot, called Keepon, was designed to represent as a simple, repeatable, mechanical robot regarding the reason mentioned above [119]. 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 that 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 humans react as the FACE displays different expressions [120]. (During the sessions, the child (IQ around 85) with autism did not show any interest in FACE at the beginning. However, with the verbal suggestion, the kid replied to the expression by using the 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 skills 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 them in the next session.

2.2.2 Different Therapeutic approaches for Individuals with ASD

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

Self-Initiated Interactions

The difficulty of initiating a social conversation or interaction is one of the impaired social skills of children with ASD. This problem may represent a difficulty in conveying what they want and why they want it. For example, when a child at an early age intends 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, the researcher extended this idea using robots to encourage the children to engage the robot proactively. The robot has built at USC, which has a large button on its back, and it was programmed to encourage social interaction with children. For example, the robot nods its head and makes a sound to encourage the kid to approach it; when the kid walks away, it moves its head down and make a 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 blows 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 into two groups: 'Group A' (children like the robot) and 'Group B' (children do not like the robot), a 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 the robot, play with himself) from avoiding the robot. This study might not be compelling because it is free to play with the robot; the experimental settings haven't kept the same, and the limited numbers of participants. Also, without a control group like typically developing, they could not compare the differences between ASD and TD children, within the robot games. However, it shows the capability of encouraging children to communicate with a robot and lead the conversation [116].

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 [116, 38]. It is easy to found out that children with ASD have a hard time allowing 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 forces them to alternate between engaging and avoiding the robot [38].

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 how to switch between different games/sessions considering children appear (i.e., different reactions of children like tired or less interested in robot). From their initial trials, children were overall happy to play with robots. At the beginning of the game, the robot showed several simple behavior patterns, such as going forward and backward. Kids showed a positive response to these behaviors and enjoyed 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 saw 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 a 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 was 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 [38]. During the interactions, it is obvious to notice that robots 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 needs to be considered.

Expression/Emotion Recognition and Imitation

Another critical 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 that contained faces or eye contact can result in overwhelming or sensory overload. For example, a person could smile twice, and the child with ASD might pick two entirely different expressions from those two smiles. The robot can provide more constancy repeatable, consistent behaviors than a 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 The University of Hertfordshire has been used to show bodily expressions by move head and arms. KASPAR was operated via wireless remote. Sessions are designed to allow the children to have free play interaction with the robot. Some behaviors had been pre-programmed in the robot, those behaviors allows KASPAR 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 with various places of KASPAR's arms, hands, face, and shoulders. By identifying different levels of touching, KASPAR would provide different movements or expressions to tell the children the emotions or feelings 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 robots, which is another weakness of this study [118]. FACE is a robot designed at the University of Pisa point to closely approximate a real human look and show detailed facial expressions. Children would be asked to imitate those expressions to practice their ability in facial expression recognition and imitation. Specific 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 easier for a therapist because of the automate repeatable of the robots process. However, still, a minimal number of kids participated in the study that made the results somehow not wholly untenable [120].

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 the different gesture, moving separate 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 the 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, necessary language skills). The frequency of child looking at the robot and the duration of each occurrence of communication has been reported. After all, they concluded that those seven modules could be applied to develop human-robot integration therapy sessions for children with autism [123]. The same year, these researchers use 5 of those seven modules did a case study, with the same setting, they recruited one high-functioning (with IQ 107) to complete those five tasks. They aimed to discover whether that child can provide a better exposure behavior with a robot compared with the activity in the class. After running the five tasks for only one instance, they concluded that the child behavior had been improved significantly with the robot than in the class, they also suggested that humanoid robot NAO can be used as a significant platform to support and initiate interaction with children with ASD [124]. After this case study, they recruited five other children with ASD (low IQ, average around 50) and did the same experimental interaction sessions with them. Out of five children showed better performance during robot interaction compared with daily in-class performance [125]. Further research has been done by this group, and they added the 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. After finished the session, researchers pointed out that NAO has the inherent capability to teach head and bod posture related to social emotions for children with autism without provided any statistical analysis only based on observations [126]. This group has been initiated working with NAO for autism therapeutic sessions 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 is 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 task 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 individuals and compare it with the TD control group. The details of our experiment and the results will be discussed in Chapter 4.

2.3 Music Therapy in robot

Socially assistive robots are widely used in the young age of autism population interventions these years. Some studies are focusing on eye contact and joint attention [19, 127, 20], showing that at some point, the pattern of ASD group in perceiving eye gaze is similar to typically developed (TD) kid, and eye contact skills can be significantly improved after intervention sessions. Plus, these findings also provide strong evidence of ASD kids are easy to attract to humanoid robots in various types 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 [128, 129, 130]. 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 a 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 [131]. Although limited research has found in such areas, especially using bio-signals for emotion recognition for ASD and TD kids [132] in understanding the relationship between activities and emotion changes.

To this end, in current research, an automated music-based social robot platform with an 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. Furthermore, by using bio-signals with Complex-Morlet (C-Morlet) wavelet feature extraction [132], emotion classification, and emotion fluctuation are analyzed based on different activities. TD kids have participated as a control group to see the difference from ASD group.

2.4 Summary

The current chapter reviewed the research-related work in Autism Spectrum Disorder, Socially Assistive Robotics, and the interdisciplinary for both topics. Went through the history of researches, the author described details in the problems of autism, including motor control, turn-taking behavior, eye-gaze, and joint attention. Some of the treatments have also been discussed in this chapter, such as music therapy. As time moving forward, robotic solutions starting to become popular nowadays. Several types of research have been listed in this chapter, and all these studies focused on different aspects of autism spectrum disorder. In the end, NAO has been briefly introduced and will have more details in further chapters.

Chapter 3

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.

3.1 Related Work

3.1.1 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 multi-modal 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 [133, 134, 135, 136]. Note that the sympathetic nervous burst changes the skin conductance, which can be traced by analyzing the EDA signals[137, 138, 139]. 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 [140]. 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 [139].

EDA signals are non-stationary and noisy; hence, wavelet-based analysis of EDA signals has been considered in the literature [141, 142] either as a pre-processing step or a feature extraction approach for emotion classification. [141] 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. [143] 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

3.1.2 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 [133] 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 [144, 145, 146, 147]. 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 [146] 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 [144] 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. [148] 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. [149] 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 [150]. 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 [151, 152], the authors combined various types of bio-signals such as ECG, skin temperature (SKT), HR, and Photoplethysmogram (PPG) for emotion classification purposes. [153] 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 [145, 154] 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.

3.2 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 [155]. 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 [156].

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 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. Figure3.2 shows the above-described procedure diagrammatically. Besides, the distribution of different emotions across all subjects and events is given in Figure 3.2.

3.3 Proposed 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

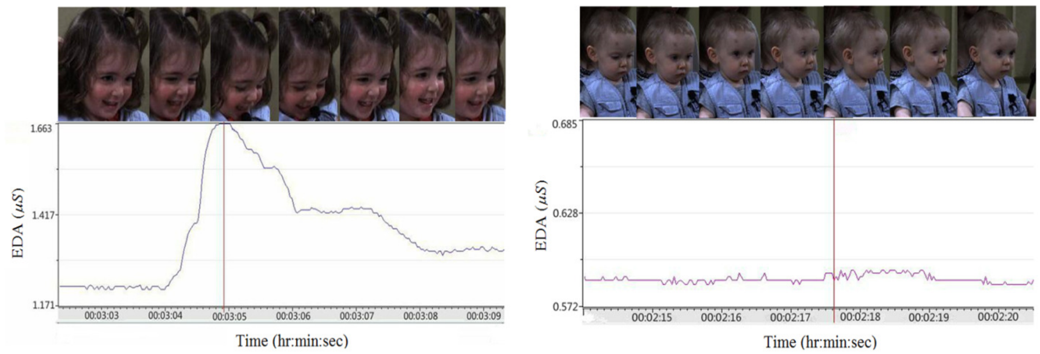


Figure 3.1: *Two samples of the annotation process, left shows some video frames associated with event 5 "tickling" and the right one shows the video frames of event "using the book as a hat". The corresponding EDA signals are shown under each case. While for the event 5 the EDA signal contains meaningful information, the EDA signal of event 4 does not contain useful information, likely due to the disengagement of the subject.*

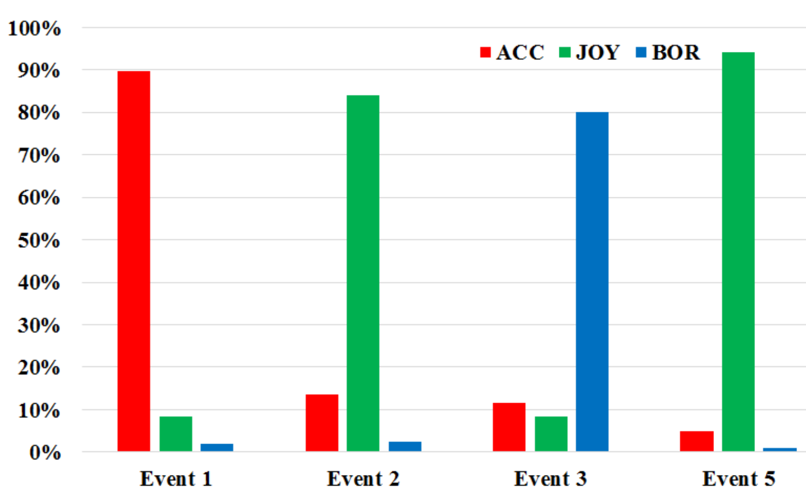


Figure 3.2: *The distribution of the emotions across all subjects and events. The abbreviations "ACC", "BOR", and "JOY" respectively correspond to the emotions "Acceptance", "Boredom", and "Joy".*

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.

3.3.1 Continuous Wavelet Transform

The EDA data recorded using the SC sensors are categorized as non-stationery signals[157, 141]. Hence, multiresolution analysis techniques are essentially suitable to study the qualitative components of these kinds of bio-signals [157]. Note that continuous wavelet transform (CWT) is one of the strongest and most widely used analytical tools for multi-resolution analysis. CWT has received considerable attention in processing signals with non-stationary spectra [158, 159]; 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[159, 160] . 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 multi-scale 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 [158]:

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

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:

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

3.3.2 Wavelet-Based Feature Extraction

The time-frequency analysis of varied bio-signals has been addressed in many related literature [161, 162, 163, 164]. 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 [165, 166], 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 [158]. As a result, a lot of elaborate representation of the signal is provided as compared with the raw time-domain signal.

Figure 3.3 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 [161]. Figure 3.4 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:

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

3.3.3 Support Vector Machine

The SVM classifier tends to separate data

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

by drawing an optimal hyperplane <w,x>+b=0 between classes such that the margin between them becomes maximum [150]. With reference to Figure 3.5, 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.

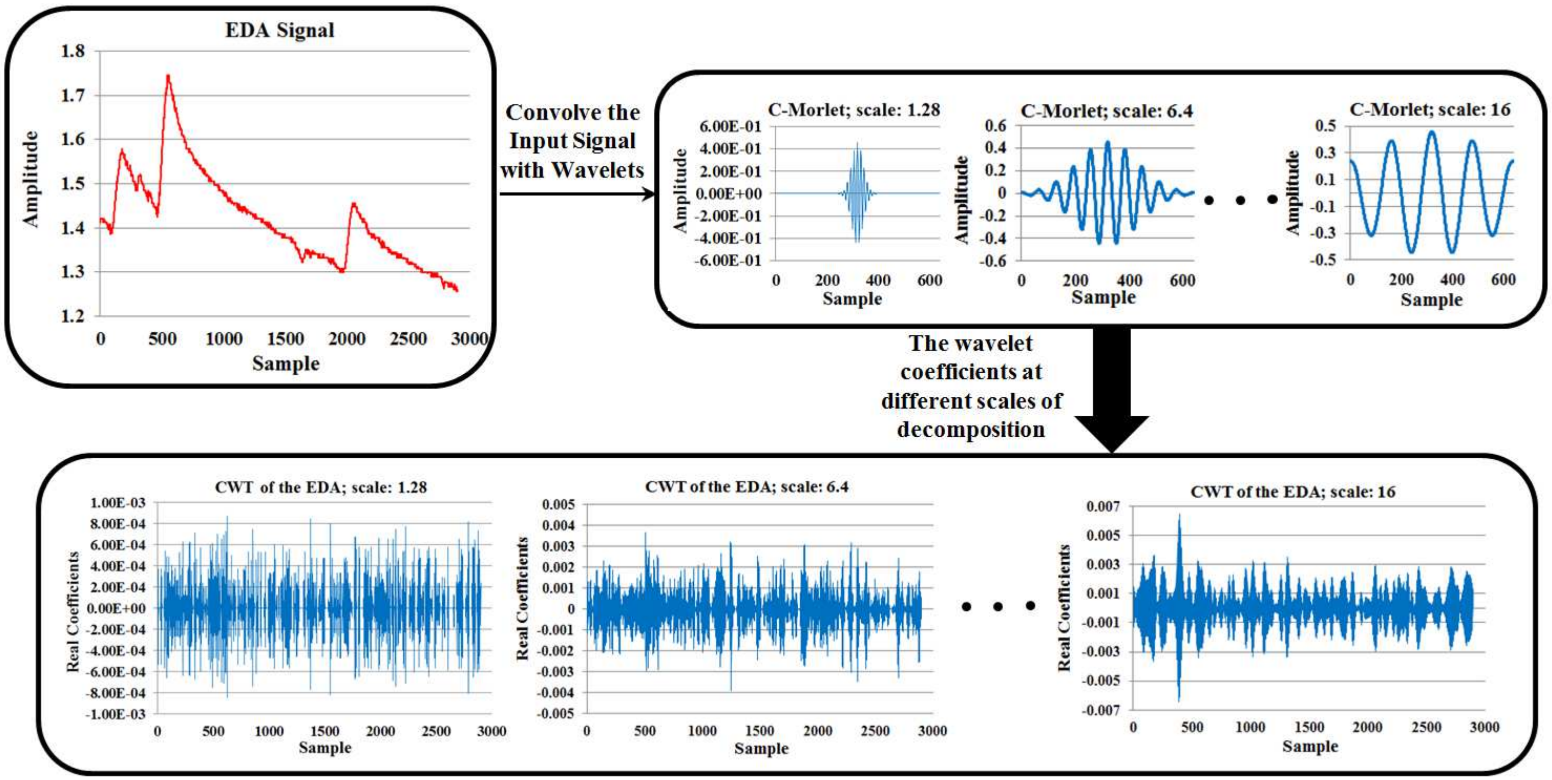


Figure 3.3: *The CWT of a typical EDA signal using the C-Morlet mother wavelet. Different scales of the wavelet functions are convolved with the original EDA signal to highlight different features of the raw data. As can be seen inside the bottom box, when the scaling parameter of the wavelet function 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.*

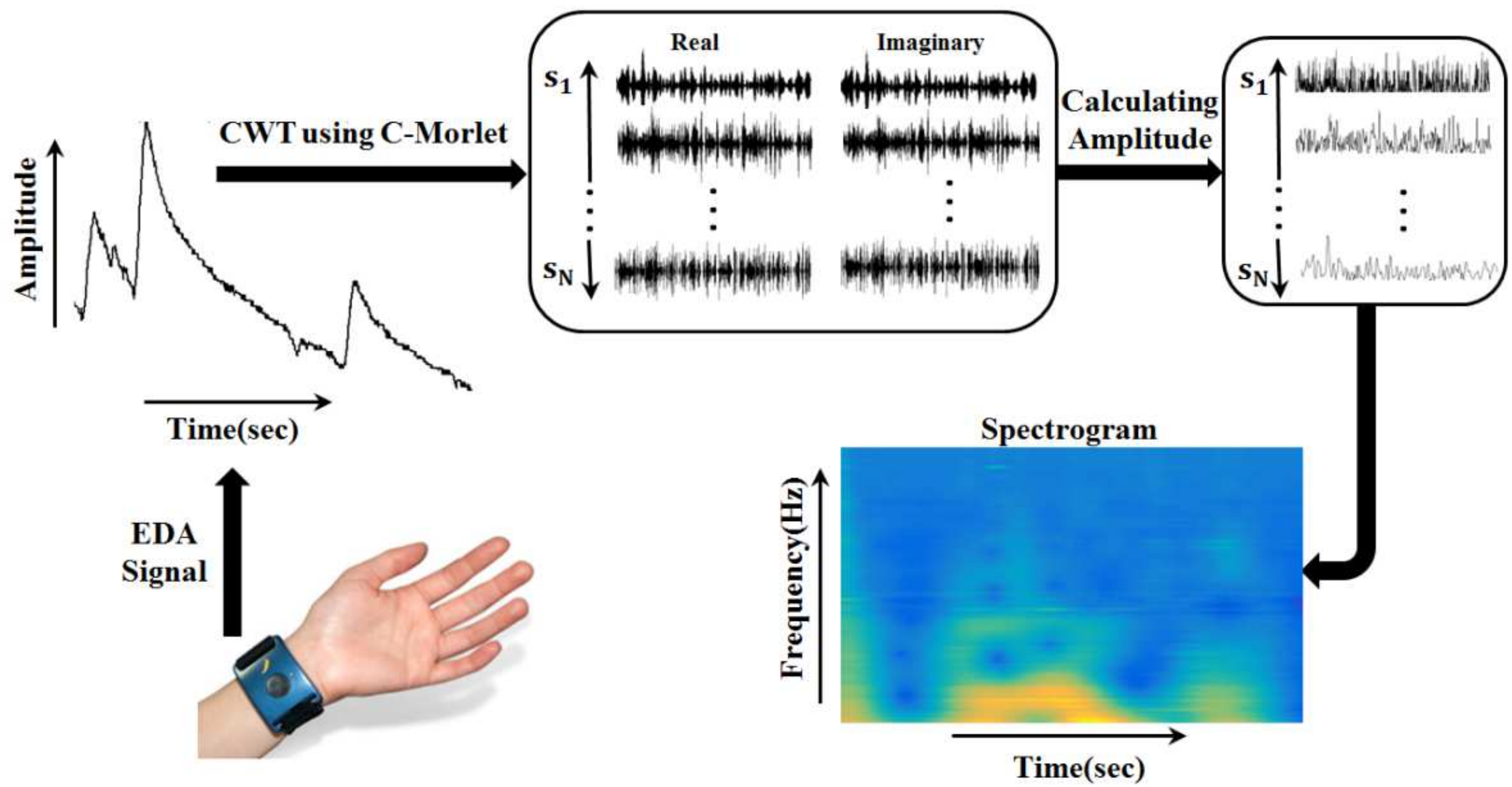


Figure 3.4: *The wavelet-based feature extraction. Using the C-Morlet mother wavelet, the real and imaginary wavelet coefficients are calculated at different scales. Then the amplitude of these coefficients is calculated to provide the corresponding spectrogram. This spectrogram is then used as the feature space.*

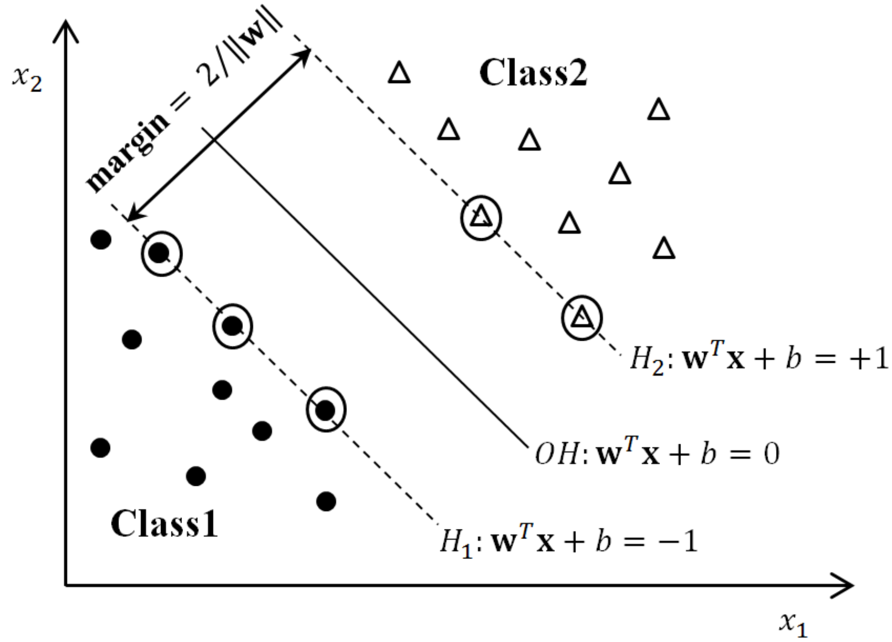


Figure 3.5: *Canonical SVM for classifying two linearly separable classes. The decision boundary is shown by OH. Two hyperplanes H1 and H2 pass the support vectors that are circled inside the figure.*

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 non-linearly separable 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 eigen vectors in the expansion. The fact that all the eigenvalues are non-negative means that the kernel is positive semi-definite [150].

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:

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

3.4 Experimental Result

This work has employ the EDA signals of 64 subjects 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.

3.4.1 Determination of Mother Wavelet

Table 3.1 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” 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 3.6 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.

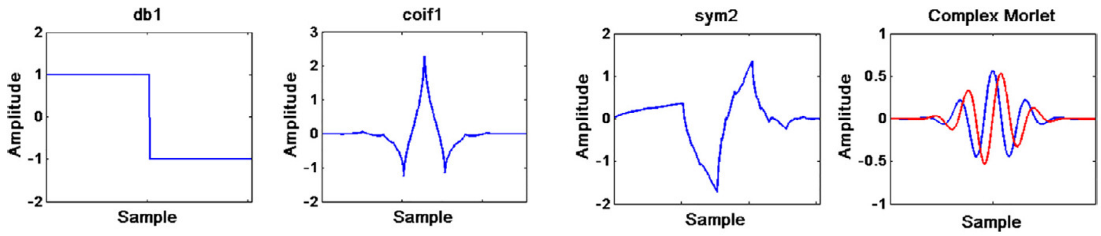


Figure 3.6: *Different mother wavelets used for feature extraction in this paper. The "db1", "coif1" and "sys2" wavelets are real-valued functions, while the "C-Morlet" function is complex-valued. The corresponding imaginary part of this wavelet function is highlighted in red inside the figure.*

Table 3.1: *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".*

	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

3.4.2 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 the suggested wavelet-based feature extraction method with the raw EDA signal are compared here. Plus, statistical feature extraction methods [167, 168] 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 impulsive 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 taken into account (See Figure 3.3).

Principal component analysis (PCA) [169] is then applied on the extracted wavelet-based features to decrease the dimensionality of the data, and, therefore, reduce the computational 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 because of enormous size of the feature 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 3.4) 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 [170]). 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 3.2, 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”), besides the 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 chosen from one of the emotions and then the precision and recall values will be calculated. 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 3.2, 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 values 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 instance, 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 result 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 non-stationary and may include random artifacts, which makes it unsuitable to use the raw time sequence for practical signal processing approaches. Prior studies have represented stochastic physio- logical signals using statistical features to classify emotional states [168]. 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 [171].

3.5 Summary

Three basic emotions were recognized within the annotation step as well as Acceptance, Joy, and Boredom. Numerous experiments were dole out on the dataset mistreatment either the raw segmented EDA signal or its corresponding time-frequency

Table 3.2: *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.*

	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			

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 4

Xylo-Bot: An Interactive Music Teaching System

As mentioned in Chapter 3, music can be considered as an effective method for emotion and non-verbal communication. Individuals with ASD are interested in interacting with a social robot. Therefore, implement music-based intervention sessions using humanoid social robot becoming possible. A novel interactive human-robot music teaching system design is presented in this chapter. Hardware and software design will be discussed in the following sections including experiment room setup, platform design, robot selection, and instrument accessories.

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, we need to find a proper xylophone with perfect timber; second, we have to arrange the xylophone in a proper position in front of the robot to make it be visible and be reachable to play; after that, a set of challenge-based experimental sessions need to be constructed including: baseline session, intervention sessions and an exit session with various levels of activities; finally, the module-based interactive music teaching system is designed and programmed which can be implemented into the experimental sessions.

4.1 Hardware Selection and Design

4.1.1 NAO: A Humanoid Robot

NAO, the well-known humanoid robot, was selected in the current research which sold by SoftBank company. NAO is 58 cm (23 inches) tall, who has 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 * 960 for online observation. As shown in Figure 4.1, these utilities are located in the middle of the forehead and mouth area. NAO's computer vision module includes facial and shapes 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-hand self-calibration system which allows the robot to have real-time micro-adjustment of its arms' joints in case of off positioning before music play.

The robot arms have a length of approximately 31 cm. The Position feedback sensors are equipped in each joint of the robot 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 following sections.

Four microphone embedded on the toy or NAO's head can be seen in Figure 4.2. 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 is 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 the later section.

4.1.2 Accessories

The purpose of this study is to have a toy-size humanoid robot to play with and teach music to children with autism. Some necessary accessories needed to be purchased

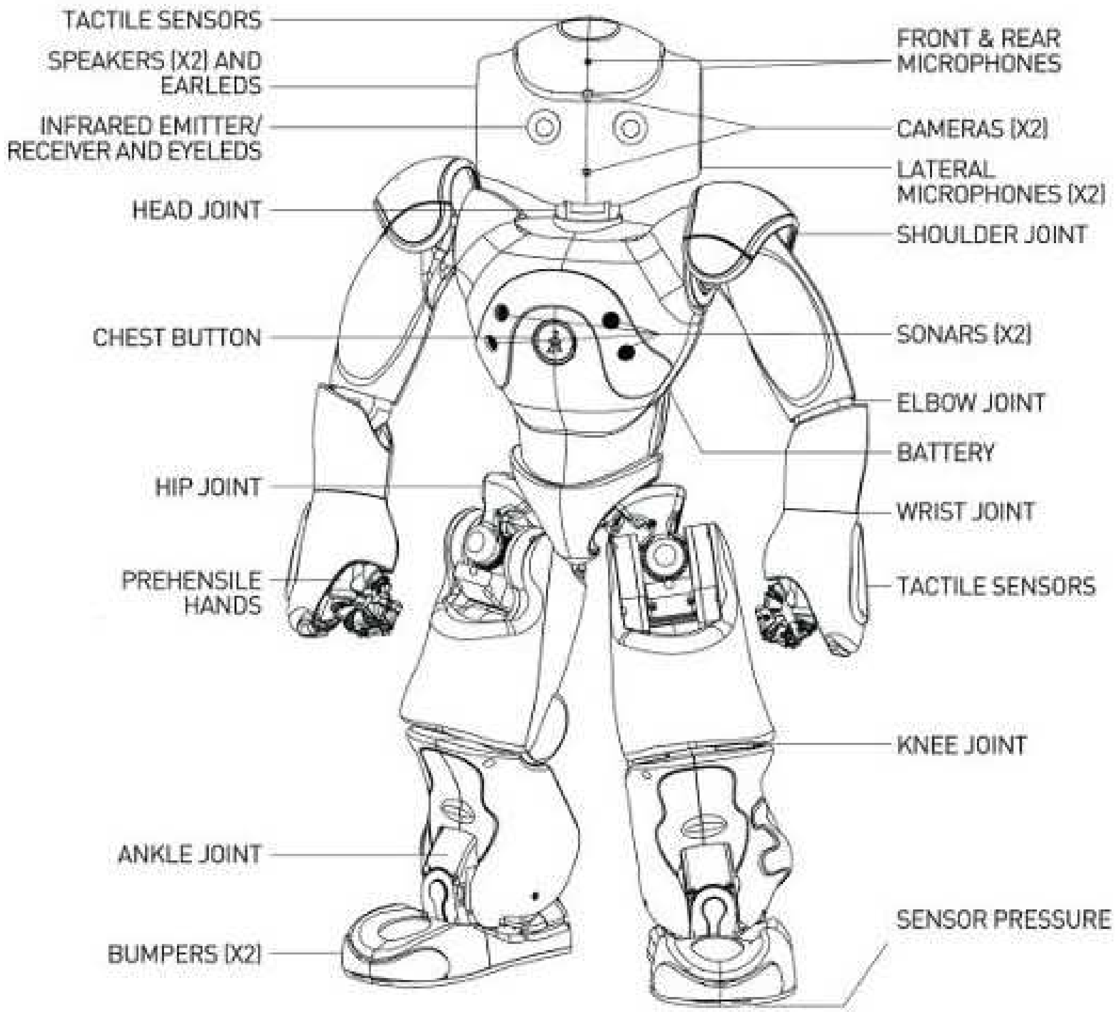


Figure 4.1: *A Humanoid Robot NAO: 25 Degrees of Freedom, 2 HD Cameras and 4 Microphones*

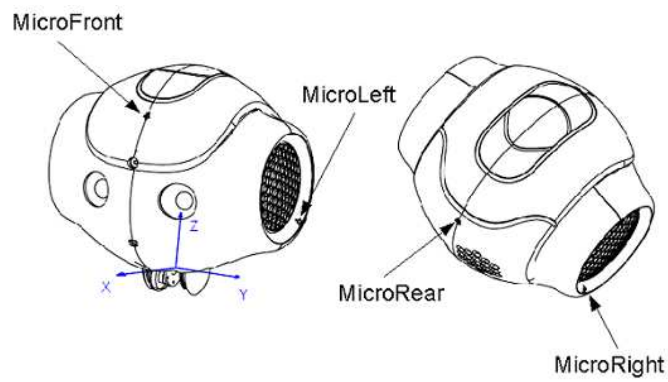


Figure 4.2: *Microphone locations on NAO's head*

and made before the robot able to complete this task. All accessories will be discussed in the following paragraphs.

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 was selected and purchased. The instrument has a size of 31 cm * 9.5 cm * 4 cm, including the resonating body. The smallest sound bar is playable in an area of 2.8 cm * 2 cm, the largest in an area of 4.8 cm * 2 cm. The instrument is diatonically tuned in C-Major/a-minor. See Figure 4.3. The 11 bars of the xylophone represent 11 different notes (11 frequencies), which covers one and a half octave scales starting from C6 to F7.

In order to provide a music teaching environment system for children with autism, the xylophone is one of the best choices for such a study. Xylophone, as well known as marimba, has categorized as a percussion instrument consisting of a set of metal/wooden bars struck with mallets to produce delicate musical tones. Other

Figure 4.3: *Actual Xylophone and Mallets from NAO's Bottom Camera*

than the keyboard or drum, for playing xylophone properly, a unique technique needs to be applied. A proper strike movement is required to produce a beautiful note coming out of xylophone. This action is perfect for a practice motor control, and the melody played by the 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 gripper, which allows the robot hands to hold them properly. The mallets are approximately 21 cm in length and include head of a 0.8 cm radius. Comparing to other designs, this mallet gripper prides nature holding position for the robot, and

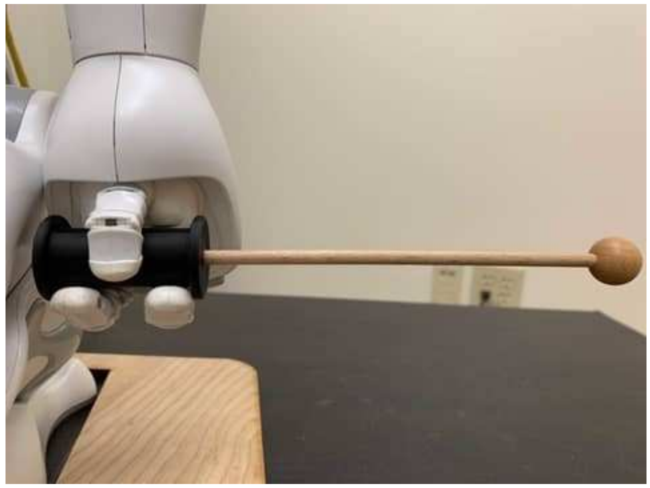
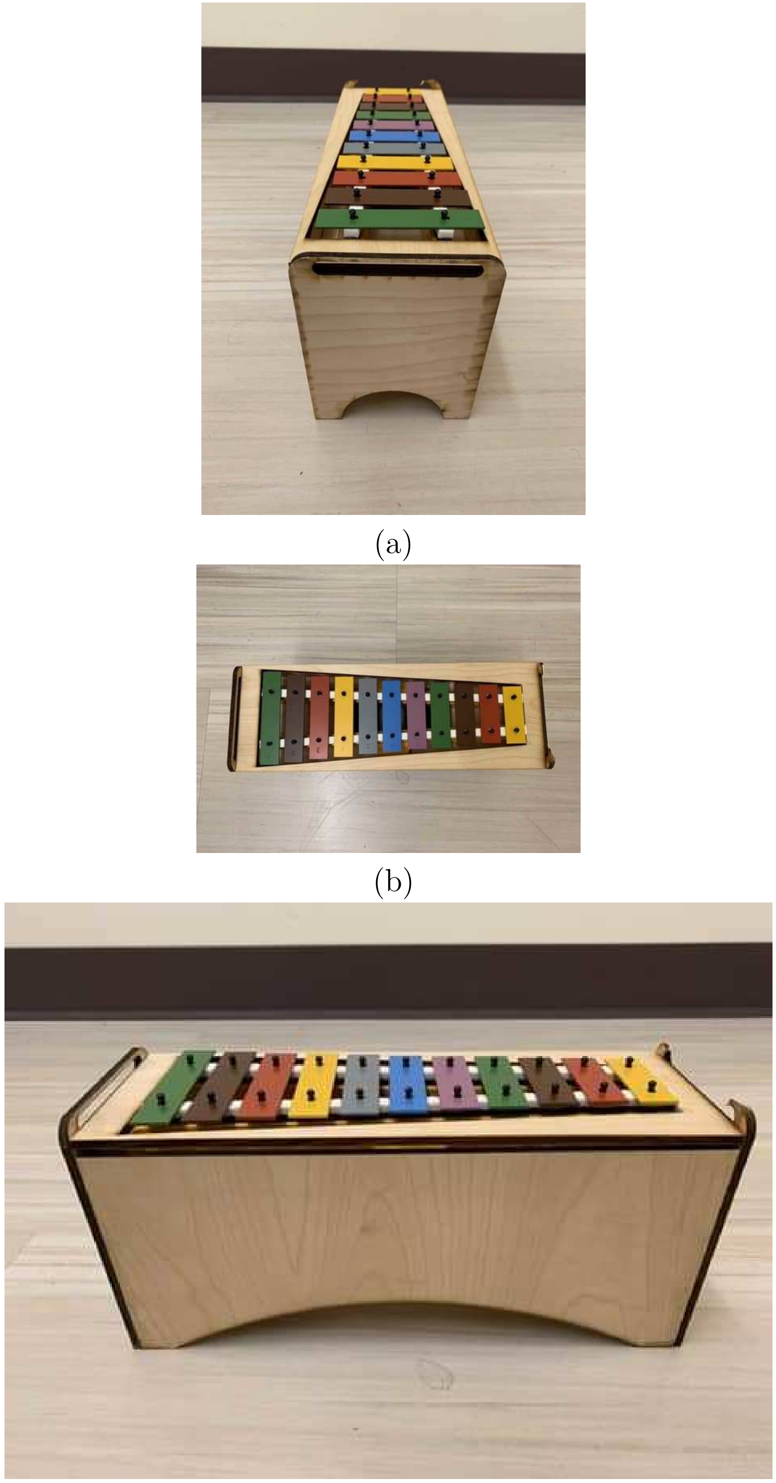


Figure 4.4: *Mallet Griper*

set up a proper modal for participants in holding the mallet stick in a right way. See Figure 4.4.

Instrument Stand Design

By carefully measured dimensions, a wooden base was designed and laser cut to hold the xylophone in a proper height for the robot crouching position. Using this fine adjusted position, robots can easily be fixed in a location and somehow have the same height level as the participants, which makes it more natural in teaching activities for the entire time. See Figure 4.5.



(a)

(b)

(c)

Figure 4.5: *Instrument Stand: (a) Left View (b) Top View (c) Front View.*

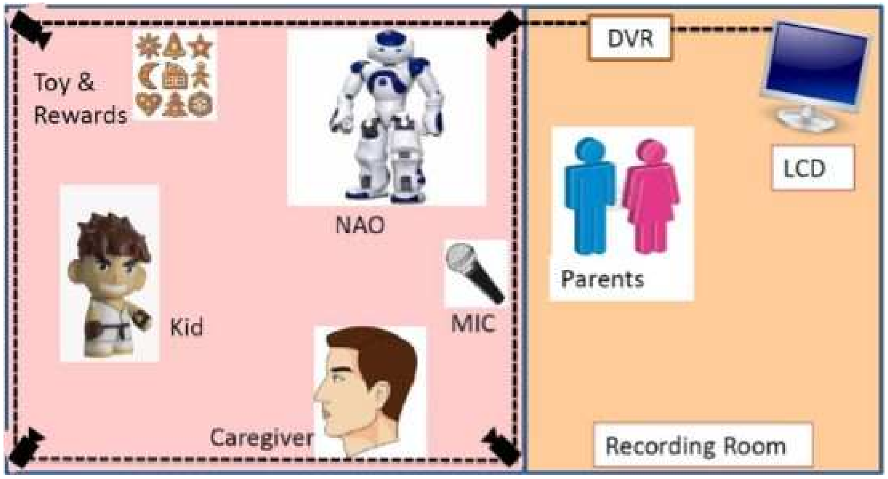


Figure 4.6: *Schematic robot-based therapy session and video capturing setting*

4.2 Experimental Sessions Designs

4.2.1 Experiment Room

All the sessions were held in an 11ft * 9.5ft * 10ft room located in the Ritchie School of Engineering Room 248, University of Denver. 6 HD surveillance cameras installed at corners, sidewalls, and ceiling of the experimental room see Figure 4.6. One mini hidden microphone attached at the ceiling camera for sending real-time audio to the observation room in order to let the care giver listen to. An external hand-held audio recorder was set in front of the participant during sessions for collecting high-quality audios for the future process.

As shown in Figure 4.7, 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 ongoing. Real-time video and audio were broadcast to the observation room during sessions, which allowed researchers to observe and to record in the meantime.

Figure 4.7: *Experiment Room*

Parents behind the mirror may also call off the session in case of an emergency.

4.2.2 Q-Sensor

One Q-sensor was used in this study. Participants were required to ware this device while session in running. It was allowed to take off the sensor during the break upon subjects request. EDA signal (frequency rate 32Hz) were collected from the Q-sensor attached to the wrists (left or right wrist determined by the participants). Taking breaks were frequently required from the participants. Due to the fact of this, 2 to 3 pieces of EDA files were recorded after each session. These files need to be annotated by comparing the time stamps with the videos.

4.2.3 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 subject pool with help from the Psychology department. For each participant in the ASD group, 6 sessions were delivered, including baseline session, intervention sessions, and exit session. As for the 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, the baseline and exit sessions length can be comparable for the same subject. As for the intervention sessions, the duration should gradually increase one after another due to the challenge level uprising.

4.2.4 Session Detail

Baseline and exit sessions contain 2 activities which are 1) music practice and 2) music gameplay. Participants were asked to complete a challenge like full song play, starting from single note strike with a 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 the practice session is done, a freshly designed music game will be presented to participants which contain three novel entertaining game modes in it, participants were allowed to communicate with the robot regarding which mode to play with. Mode 1): the robot will randomly pick a song from its song bank and play for kids, after each play, a music feeling will be requested from participants in order to find out whether music emotion can be recognized from the early age of children; Mode 2): a sequence of melody will be randomly generated by a robot with consonance (happy or comfortable feeling) or dissonance (sad or uncomfortable feeling) style and oral emotion feeling from participants will be requested and physical playback afterward; Mode 3): allows participants to have 5 seconds of free play and challenge the robot to imitate from the participants what just played by them, after complete robot playing, performance will be rated by the participants provide 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 a 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 each individual chose participants and a customized song for exit session 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, inputting 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 Taylor Swift, "Spongebob Square Pants" from Spongebob cartoon, and "You are my Sunshine." by Johnny Cash etc. Music styles are not only kid's songs like "Twinkle Twinkle" but also covered in classic, pop, ACG, folk, and more can be played by using such a 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 a color hint); and S3) music gameplay. Starting from intervention sessions, a user-customized song was used in the following sessions and have them engage more in multiple repetitive activities. The purpose of having a warm-up section is to have the motor control skill been practiced and meanwhile to help participants implement the motor skills in following activities with a fresh memory. The single activity was based on music practice from the baseline/exit session, other then those sessions, the single activity will only have one type of music practice each session, for instance, the single-note play was delivered in the first intervention session. The next time this practice will become multiple notes play, and the level of difficulty for music play was gradually increased session after session. This was in order to make a challenge based engagement activity for ASD groups for better motivation and emotion stimuli. As for music, the gameplay remained the same as the baseline/exit session. See details in Table 4.1

4.3 Module-Based Acoustic Music Interactive System Design

In this section, a novel module-based robot-music teaching system is presented. Several goals need to be accomplished here: a) make the robot play sequence of notes or melody fluently; b) have the robot play music notes accurately; c) be able to adopt multiple songs easily; d) be able to have social communication between the robot and the participants; e) be able to deliver learning and teaching experiences to the participants; f) fast response and accurate decision making. In order to meet these goals, a module-based acoustic music interactive system was designed in this work. Three modules have been built in this intelligent system including Module 1: eye-hand self-calibration micro-adjustment system; Module 2: joint trajectory generator; and Module 3: real-time performance scoring feedback system. See Figure 4.8.

Table 4.1: *Session Details: Baseline session, Intervention sessions, and Exit session. Different activities also included in this table.*

#	Session Type	Outlines	Session Details	Purpose
1	Baseline Session	Include all music activities in one session, play song "Twinkle Twinkle Little Star".	Music Practice: Single note strike Multiple notes strike Half song practice Whole song practice Music Game: Mode 1) robot play song(s) based on user's request Mode 2) robot randomly plays consonance and dissonance sequence of notes, ask user to feel and playback. Mode 3) user play random melody for 5 seconds, robot playback, user grade robot performance.	Introduce the novel platform to participants for the first time. Record music play and social behavior baseline. Introduce music game to user for the first time.
2	Intervention Session	Single music activity practice, all notes coming from customized song.	Warm-Up: Single bar strike regardless color Single Note Strike: Single bar strike correspond to correct color. Music Game: Same as previous session	Mostly focus on motor control skill practice. First time introduce music pitch to user. Using teaching scenario to practice turn-taking behavior.
3		Multiple notes music practice, all notes coming from customized song.	Warm-Up: Same as previous session Multiple Notes Strike: 2-3 multiple bars strike correspond to correct colors. Music Game: Same as previous session	Review motor control skill. Simple music short memory practice. Using teaching scenario to practice turn-taking behavior.

Table 4.2: *Continued Table for Session Details.*

Continued table				
4	Intervention Session	First half customized song practice.	Warm-Up: Same as previous session Half Song Practice: Randomly select a chunk of melody from first half of the song (5-7 notes), ask user to memorise the music and playback, with color hits. Music Game: Same as previous session	Review previous music and social behaviors and skills. Implement all skills to a harder challenge.
5		Second half customized song practice.	Warm-Up: Same as previous session Half Song Practice: Randomly select a chunk of melody from second half of the song (5-7 notes), ask user to memorise the music and playback, with color hits. Music Game: Same as previous session	Review previous music and social behaviors and skills. Implement all skills to harder challenge.
6	Exit Session	Include all music activities in one session, play customized song.	Music Practice: Single note strike Multiple notes strike Half song practice Whole song practice Music Game: Same as previous session Music Game: Same as previous session	Compare music and social performance with baseline session.

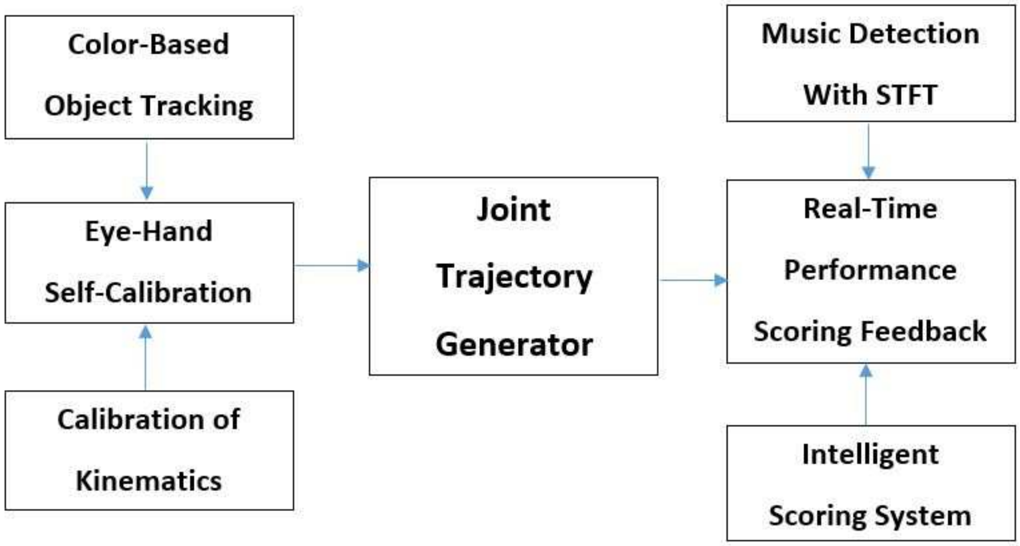


Figure 4.8: *Block Diagram of Module-Based Acoustic Music Interactive System*

4.3.1 Module 1: Eye-hand Self-Calibration Micro-Adjustment

Knowing about the parameters of the robot's kinematic model is essential for high precision tasks, 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 that the targeted angle chain of arms never actually equals the returned chain. We, therefore, used a calibration method to eliminate these errors accurately.

Color-Based Object Tracking

To play 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 mallets it is holding. To estimate these poses, adopted in this thesis, we used a color-based technique.

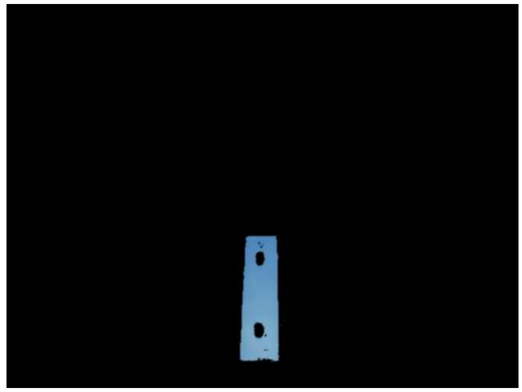
The main idea is based on the RGB color of the center blue bar, given a hypothesis about the instrument's pose, one can project the contour of the object's model into the camera image and compare them to actually observed contour. In this way, it is possible to estimate the likelihood of the pose hypothesis. By using this method, it allows the robot to track the instrument with very low cost in real-time. See Figure 4.9

4.3.2 Module 2: Joint Trajectory Generator

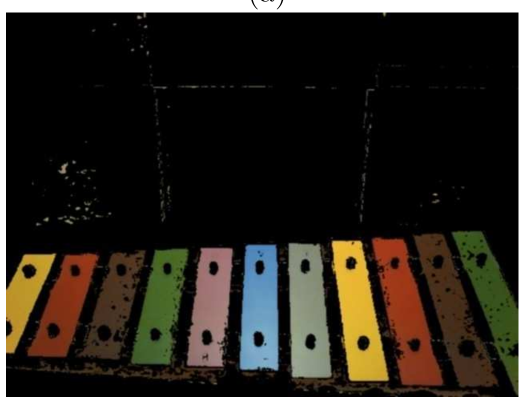
Our system parses a list of hex-decimal numbers (from 1 to b) to obtain the sequence of notes to play. It converts the notes into a joint trajectory using the beating configurations obtained from inverse kinematics as control points. The timestamps for the control points will be defined by the user to 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.

4.3.3 Module 3: Real-Time Performance Scoring Feedback

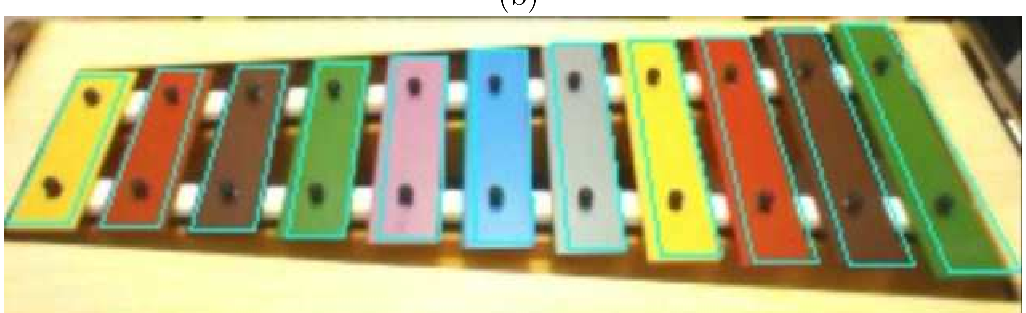
The purpose of this system is to provide a real-life interaction experience using music therapy to teach kids social skills and music knowledge. In this scoring system, two core features were designed to complete the task: 1) music detection; 2) intelligent



(a)



(b)



(c)

Figure 4.9: *Color Detection From NAO's Bottom Camera: (a) Single Blue Color Detection (b) Full Instrument Color Detection (c) Color Based Edge Detection.*

scoring-feedback system.

A. Music Detection

Music, in the understanding of science and technology, can be considered as a combination of time and frequency. 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 of 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:

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

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 the 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 the user; 2) used as a new input to have robot playback in the game session.

B. Intelligent Scoring-Feedback System

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

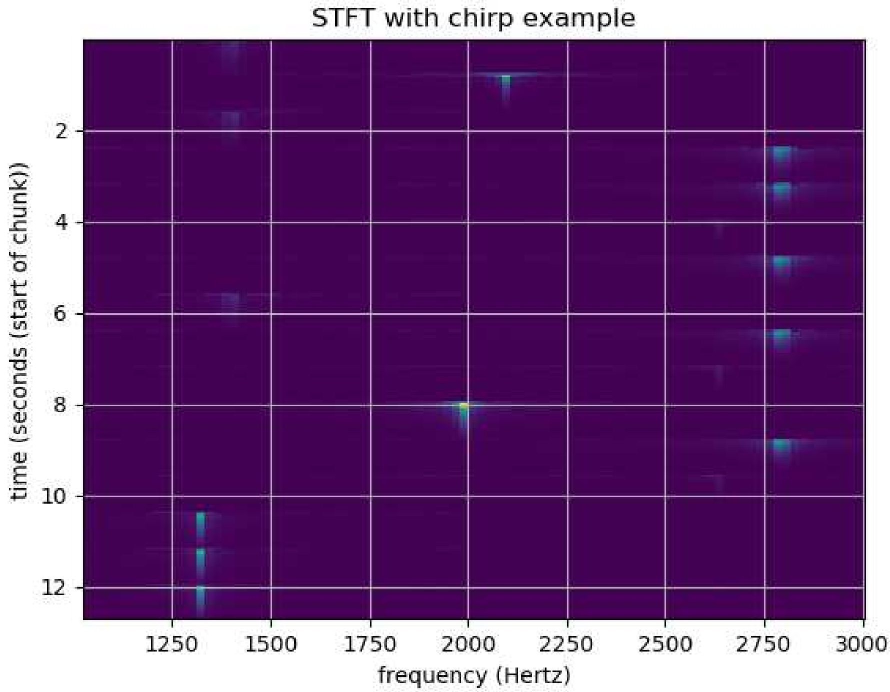


Figure 4.10: *Melody Detection with Short Time Fourier Transform*

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

where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$ and equal to 1 otherwise, and $\mathrm{lev}_{a,b}(i,j)$ 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),

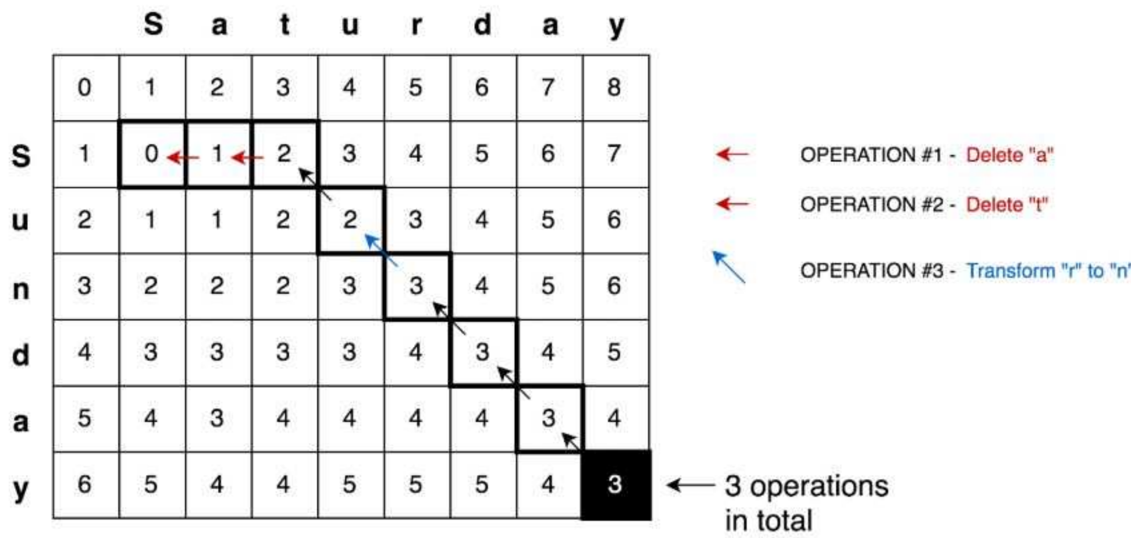


Figure 4.11: *An Example of Compute Levenshtein Distance for "Sunday" and "Saturday"*

the second to insertion and the third to match or mismatch, depending on whether the respective symbols are the same. Table 4.11 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 more significant than 66% (not including single note practice since it is only correct or incorrect in that case), 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 (e.g., 6 correct out of a total of 10 trials). More details will be discussed in the next chapters as it relates to the experiment design.

4.4 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 pure music and deliver social content simultaneously. 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. Based on the size and position, a wooden based xylophone stand was customized and assembled in use. 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 adopt 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.

The 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 the single note to the whole song. Music game was designed to keep the session less tedious comparing to the learning part, and it also provides an opportunity for us to learn more about the difference in learning and to teach music from the participants.

Module-based acoustic music interactive system design was challenging. Several problems need to be solved. In order to improve the robot strike xylophone accuracy, Module 1 provides an autonomous self-awareness positioning system for the robot to localize the instrument and make the micro adjustment for arm joints that helps NAO plays the note bar correctly. Multiple songs were required to be able to play by NAO, program each song with specific arm movements sounds ridiculous. An easy music score inputting method needs to be completed before the 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 adapted to fulfill the requirement which allows the robot to understand music and provide a proper dosage of practice and oral feedback to users.

Chapter 5

Results of Social Behavior Intervention and Emotion Classification

In this chapter several questions will be addressed: 1) Turn-taking: How well do children with ASD behave in the music activities?

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

3) Social Engagement: How did children engage with different level of music teaching events?

4) Emotion Fluctuation: How did emotion change among different activities? How did emotion change in a single event? What is the difference between the target and

control groups?

5.1 Social Behavior Result

5.1.1 Motor Control

Nine ASD and 7 TD participants finished this study in 8 months. All the ASD subjects completed 6 sessions including the intervention sessions and the TDs participated in only 2 sessions (a baseline and an exit session). By using a Wizard of Oz control style, a well-trained researcher conducted the baseline and exit sessions. With the well designed and fully automated intervention sessions, NAO was able to initiate music teaching activities with the participants.

Since the music detection method was sensitive to the audio input, that required clear and long-lasting sound from a xylophone. From Figure 5.1, it was evident that the majority of subjects were able to strike or play xylophone properly after one or two sessions. Notice that subject 101 and 102 had a significant improvement curve during the 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 some ups and downs. Two subjects (103 & 107) had a difficult time playing the xylophone and following the turn-taking cues with the robot. This fact affected the performance in the following activities for both subjects.

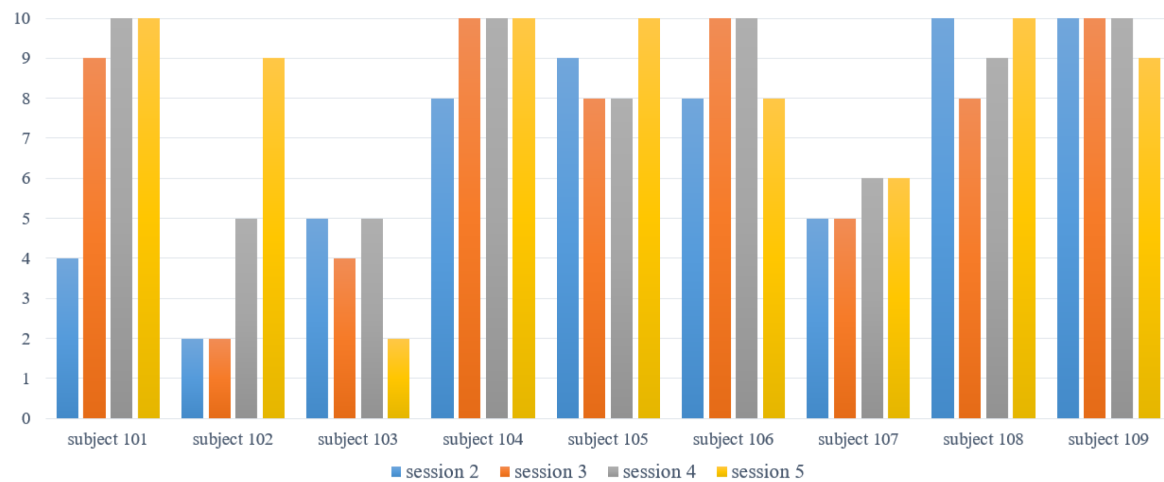


Figure 5.1: *Motor Control Accuracy Result*

Figure 5.2 shows the accuracy of the leading music teaching activity for intervention sessions across all the participants. Learning how to play one's favorite song can be considered as a motivation for ASD kids understanding and learning turn-taking skills. As described in the previous section, the difficulty level of this activity was designed uprising. By this fact, the accuracy of the performance from participants was expected to decrease or maintain at the same level. This activity required participants to concentrate and utilize joint attention skills in the robot teaching stage and also respond properly afterward. Enough waiting time was given after the robot says: 'Now, you shall play right after my eye flashes,' participants have also received an eye color change cue from the robot in order to complete a desired music-based social interaction. Different from the warm up section, notes played in the correct sequence of order can be considered as a good-count strike. From Figure 5.2, we can say that most of the participants were able to complete single/multiple notes practices with an average accuracy rate of 77.36%/69.38%, although even with color

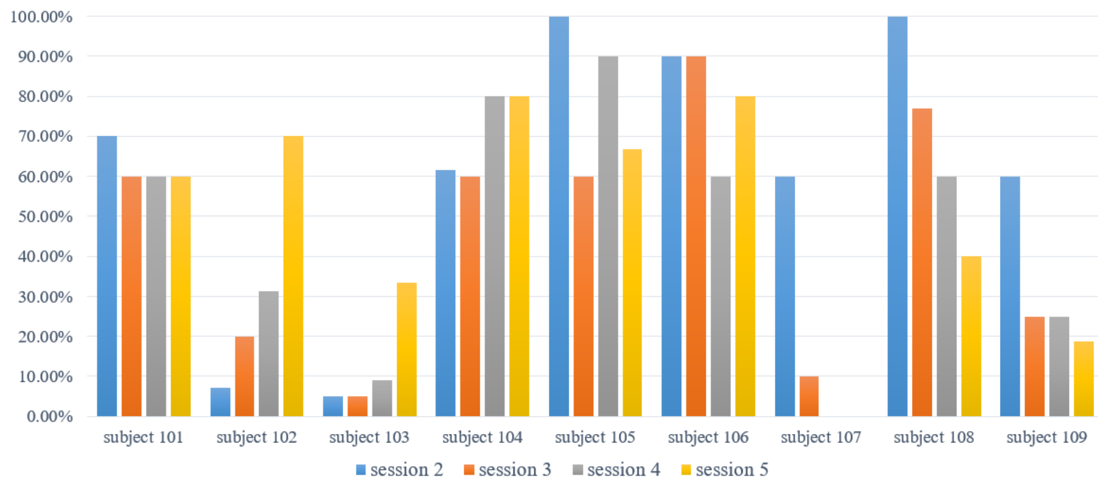


Figure 5.2: *Main Music Teaching Performance Accuracy*

hints, the notes' pitch difference still can be a significant challenge for them. Due to the difficulties of sessions 4 and 5, a lower performance compared to the previous two sessions was acceptable. However, more than half of the participants showed consistently an acceptable performance (i.e., accuracy) or even better results than previous sessions. Combining the report from video annotators, 6 out of 9 subjects showed stable and engaging behavior in playing music, especially after the first few sessions. Better learn-and-play turn-taking rotation was performed over time and a significant increase in performance by 3 subjects, reveal that the turn-taking skills were picked up from this activity.

5.1.2 Turn-Taking Behavior

Measuring turn-taking behavior can be subjective. In the current study, a grading system were designed in order to have convincing result. Music teaching activity can be considered as "conversation" between instructor and student. 4 different behaviors are defined in the grading system: (a) "well-done", this level is consider as a good behavior, participant should be able to finish listen to the instruction from NAO, start playback after receive the command, and wait for the result without interrupting, 3 points for each "well-done"; (b) "lite-interrupt", in this level, participant may show slightly impatient in different stages for example did not wait for the proper moment to play or did not pay attention to the result, 2 points will be given in this level; (c) "heavy-interrupt", more interruptions may accrue in this level, participant may interrupt the conversation at anytime but still willing to playback to the robot, 1 point for this level; (d) "indifferent", participant shows less interest in music activity including but not limited to following behaviors, not willing to play, not listen to robot or play irrelevant music in one conversation, this level will score a 0 point. The higher the score the better turn-taking behavior the participant has. All scores are normalized into percentage due to the difference of total "conversation" numbers. Figure 5.3 shows the total result among all subjects. Note that, 6 out of 9 participants can perform consistent turn-taking behavior during intervention sessions.

5.2 Music Emotion Classification Result

Since we have developed our emotion classification method based on the time-frequency analysis of the EDA signals, we first present the main properties of the continuous wavelet transform assuming complex Morlet wavelet. Then, we discuss the preprocessing steps as well as the wavelet-based feature extraction scheme. Finally, we

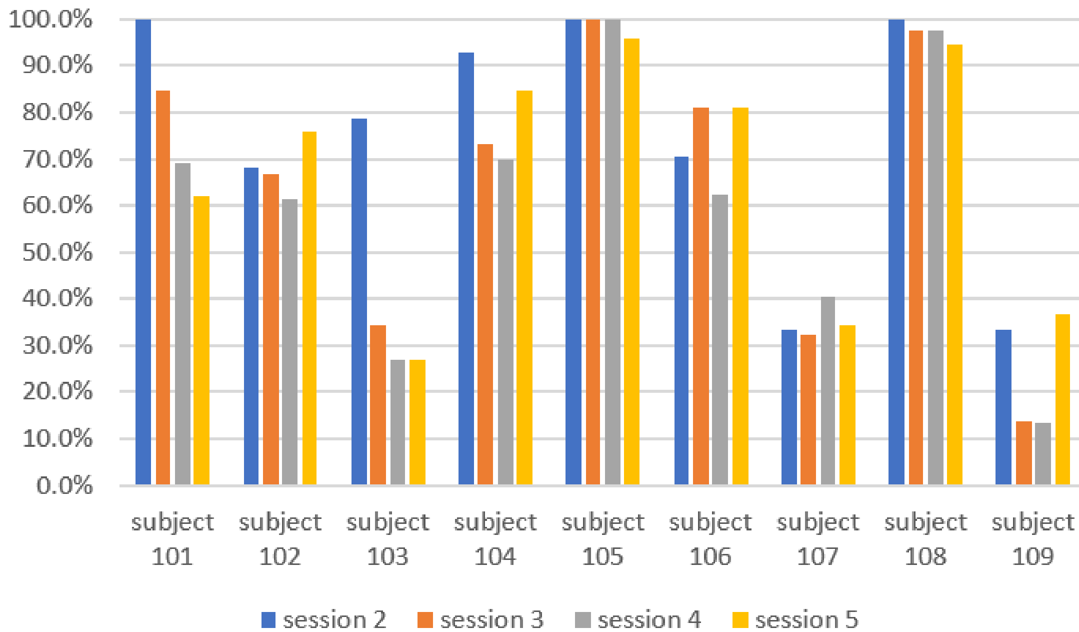


Figure 5.3: *Normalized Turn-taking behavior result for all subjects during intervention sessions.*

briefly review the characteristics of the support vector machine as the classifier used in our approach.

EDA signal was collected in this study. By using the annotation and analysis method from our pre-study [132], a music-event-based emotion classification result will be presented below. In order to find out the emotional experiences of the ASD group, multiple comparisons were made after annotating the videos. Different activities may cause emotional arousal change. As presented above, the warm up section and single activity practice section have the same activity in different levels of intensities, and gameplay has the lowest difficulty and more relax as designed in purpose.

The annotation was administrated supported the temporal relation between the video frames and the recorded EDA sequences of every subject. In different words, the annotator 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 perceived emotion. The music activities were designed to stimulate different emotions: (S1): "warm-up", (S2): "music practice", and (S3): "music game". During the first part annotation, it is not obvious to conclude the facial expression changes from different activities, however, in S1 participants showed more "calm" for most of the time due to the simplicity of completing the task; in S2, the annotator could not have precise conclusion by reading the facial expression from participants as well, most of the subjects intent to play music and complete the task, "frustrated" can be the best to describe them; a very similar feeling can be found in S3 comparing to S1,

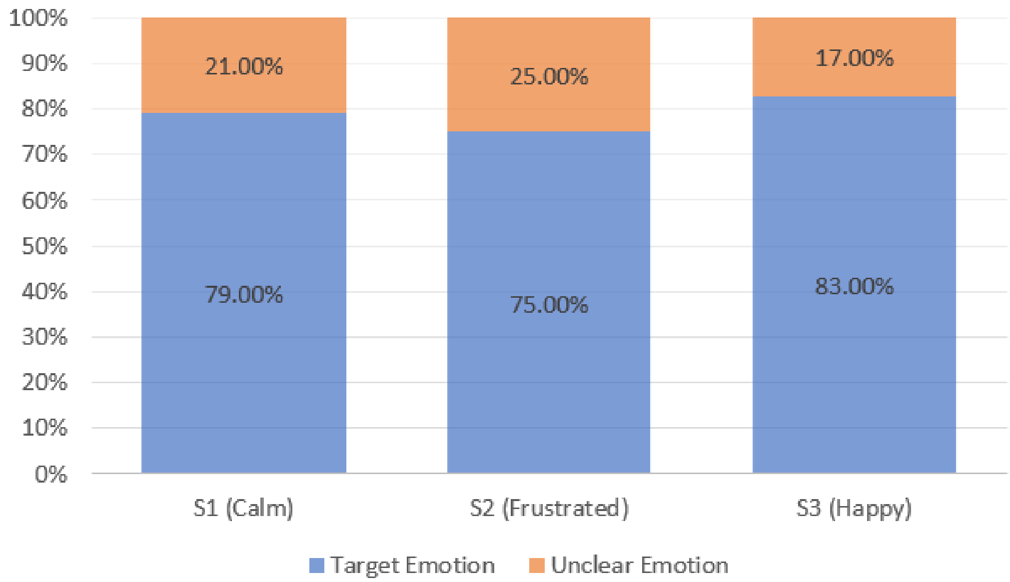


Figure 5.4: *The distribution of the targeted emotions across all subjects and events.*

in S3, all activities were designed to create a role changing environment for users in order to stimulate a different emotion, most of kids showed "happy" during the music game section. Figure **??** shows the above-described procedure diagrammatically. Due the fact that it is hard to conclude these emotions with specific facial expressions, event numbers will be used in following analysis representing emotion comparison. Note that there is total of 21% on average of emotions are not clear during the first annotation stage. It is necessary to have them as unclear rather than label them with specific categories.

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, and an SVM classifier was then employed to classify "conversation" segmentation among 3 sections using the wavelet-based features. Table 5.1 shows the classification accuracy

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

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

for the SVM classifier with different kernel functions. As can be seen, emotion difference between warm-up (S1) and music practice (S2), S2 and music game (S3) can be classified using wavelet-based feature extraction SVM classifier with an average accuracy of 76% and 70%. With the highest 64% of accuracy for S1 and S3, that may indicate fewer emotion changes between the warm-up and game sections.

In the second part of the analysis, EDA signals were segmented into small event-based pieces according to the number of "conversations" in each section as mention before. One "conversation" was defined with 3 segments: 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 lasts about 45 seconds. In order to discover the emotion fluctuation inside one task, each "conversation" section has been carefully divided into 3 segments, as described before. Each segment lasts about 10

- 20 seconds. Table 5.2 shows the full result of emotion fluctuation in the warm-up (S1) and music practice (S2) sections from the intervention session. Notice that all of the segments cannot be appropriately classified using the existing method. Both SVM and KNN show stable results. This may suggest that the ASD group may have less emotion fluctuation or arousal change once the task starts even with various activities in it. Stable emotion arousal in a single task could also benefit from the proper activity content, including robot agents play music and language used during the conversation. Friendly voice feedback was based on the performance delivered to participants who were well written and stored in memory, both favorable awards while receiving correct input and encouragement while play incorrectly. Since emotion fluctuation can affect learning progress, less arousal change indicates the design of intervention sessions are robust.

Cross-section comparison is 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 processes of the session between the baseline session for both groups were the same. By using the "conversation" concept above, each of them has been segmented. Comparing with target and control groups using the same classifier, 80% of accuracy for detecting different groups. See Table 5.3. Video annotators also reported "unclear" in reading facial expressions from the ASD group. These

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

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

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

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

combined messages suggest that even with the same activities, different bio-reaction was completely opposite between TD and ASD groups. It has also been reported that significant improvement of music performance was shown in the ASD group, although both groups have a similar performance at their baseline sessions. Furthermore, the TD group were shown more willing to try to make their performance as better as possible while they made mistakes.

5.3 Summary

All the experimental results are presented in this chapter; answers to the questions mentioned in the beginning of the chapter can be found out from them. According to the report from the annotator, most of the kids (both ASD and TD groups) showed well turn-taking communication behavior among all sessions. However, differences can also be found when comparing both groups. All TD participants could initiate the activities from the beginning of the session, while some of the ASD kids needed some help, although, after several visits (i.e., intervention sessions), most of them could perform the turn-taking skills well. In terms of motor control skills, as can be seen from Figure 5.1, most of the ASD participants can master this skill after the first few visits. For the ones who may not play xylophone properly, an improvement also can be observed in this figure. Based on the recorded videos and Figure 5.2, more than half of the ASD kids well engaged during the intervention sessions. Few of them needed help from the researcher to complete the tasks in the first or second sessions. Since each individual chose their favorite music, this could provide a certain level of motivation which is more engaging even with the repetitive activities. From Chapter 3, we learned that emotion classification using the EDA signal could be possible, and a wavelet-based feature extraction method was developed and applied in a similar research with a group of younger children. In this chapter, we adopted the proposed approach in order to decode the emotion fluctuation with music social activities for children with autism. Multiple experiments were established in this chapter and emotion change was compared across different events, within one activity and between the target and control group. It can be found that, from Table 5.1 different titled music activities can stimulate different emotion changes. Warm-up as a single note play activity without pitch correctness which makes it the most easiest activity compare to the music practice activities. To this fact, less stress can be caused in the warm-up section during intervention sessions. Similar to the music game section, participants were not require to consider how well they play but only challenge the robot to mimic what they have played. This can explain the reason why S2 can be classified from S1 and S3. Intra-activity emotion are also discussed in this chapter see Table 5.2. By comparing different segments of one activity, it is hard to tell the emotion changes between segments in one "conversation" among all activities. At some point, emotion in learning how to play xylophone may have less difference between playback to the instructor of what have just learned. This may suggests that well social behaviors may be benefit from less emotion fluctuation in music activities combining the results from Figure 5.1 and Figure 5.2. More further discussion and conclusion can be found in the last chapter.

Chapter 6

X-Elophone, A New Instrument

After the first phase of the experiments done, some limitations for the current design were noticed. A limitation of the music representation can be found out of the acoustic instrument, especially when certain songs with rich music contents are played. Some of the participants requested rock or electric songs for practices, thus an acoustic xylophone barley meet the requirements due to the sound quality. In this chapter, the design and prototyping of a novel instrument based on Xylophone are described. The purpose of this design is to add more possibilities for playing different timber and major/minor keys. The increased amounts of possible notes would allow the system to play more customized songs for kids.

6.1 Xylophone Modification

6.1.1 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 Piezo polymer, 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 the MOSFET or CMOS arranges legitimately. If the assembly is supported by its contacts and left to vibrate "in a free space" (with the inertia of the clamped/free beam creating bending stress), the device will behave as an accelerometer or vibration sensor. Increasing the 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 6.1 shows the schematic of a Piezo vibration sensor and Figure 6.2 shows how it looks like.

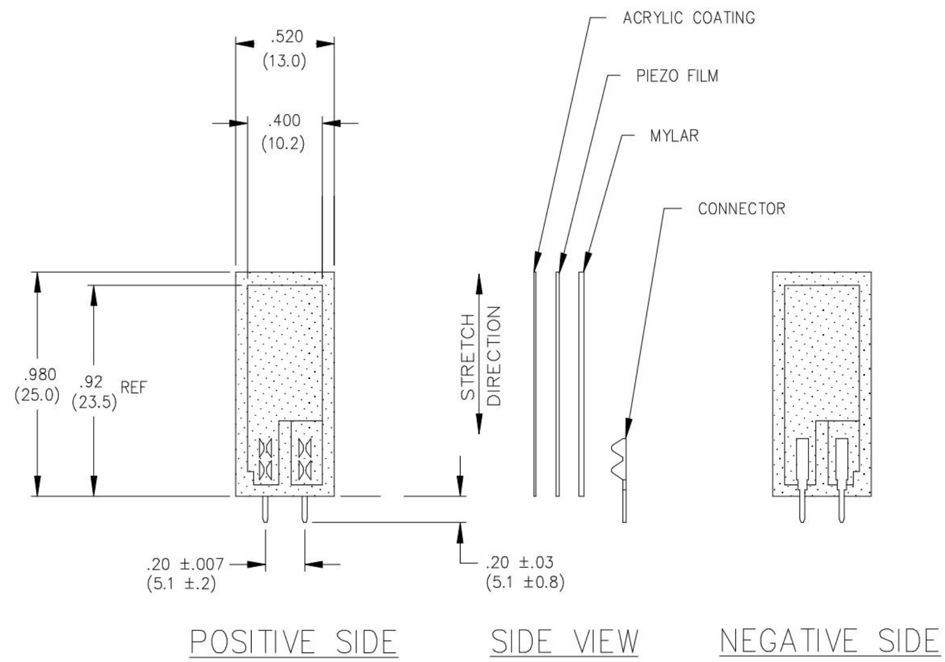


Figure 6.1: *Piezo Sensor Schematic*

B. Op-Amp: An operational amplifier (often called 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 or 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 dependency on temperature coefficients or engineering

Figure 6.2: *Piezo Sensor VS A Quarter*

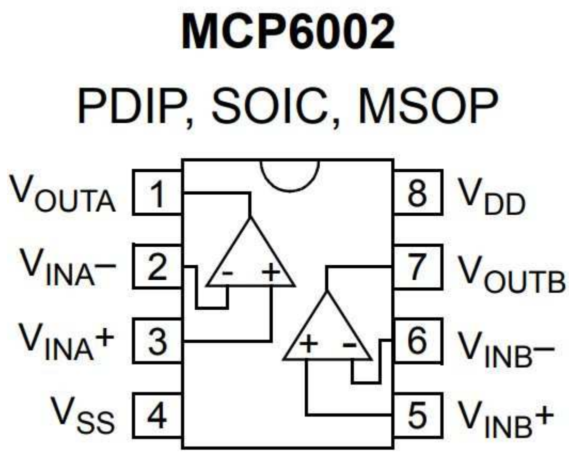


Figure 6.3: *Schematic of Op-Amp MCP 6002*

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 6.3 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 6.4

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

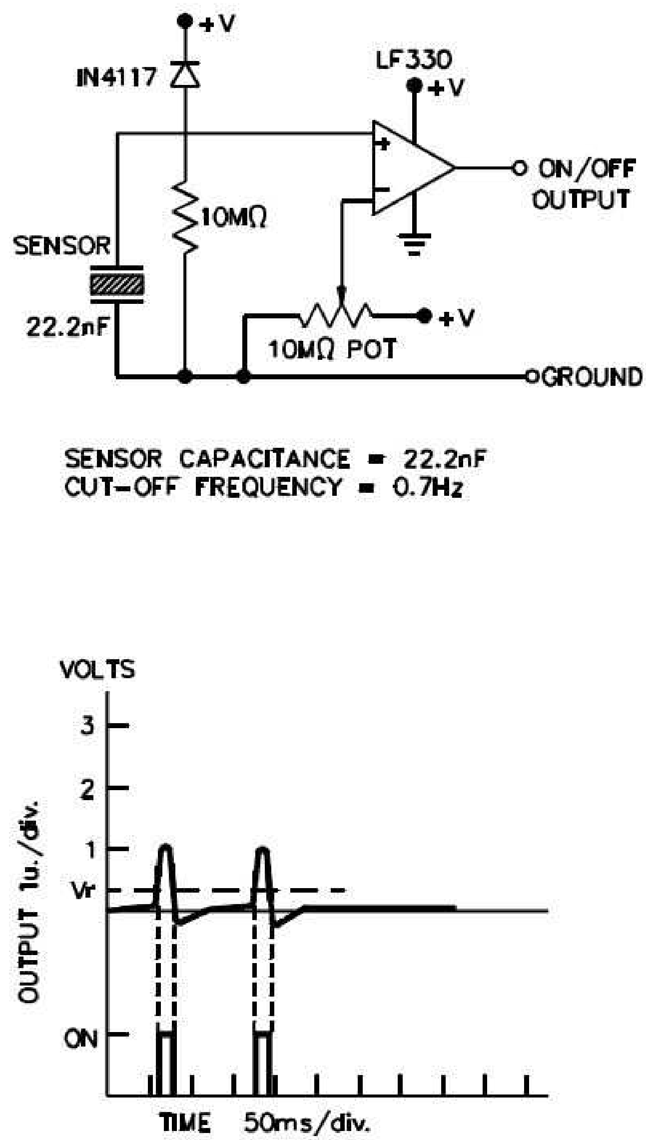


Figure 6.4: *Schematic of Piezo Sensor Application as A Switch*

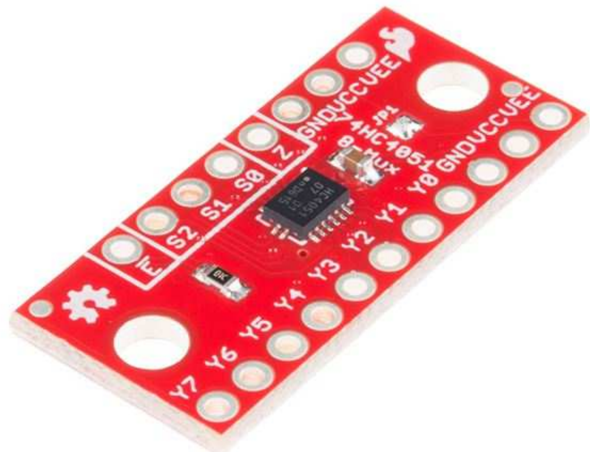


Figure 6.5: *SparkFun Multiplexer Breakout 8 Channel (74HC4051)*

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 *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 6.5 shows the multiplexer used in this study.

D. Arduino UNO: The Arduino Uno see Figure 6.6 is an open-source micro-controller board based on the Microchip ATmega328P micro-controller and developed by Arduino.cc. The board is equipped with sets of digital and analog input/output

Figure 6.6: *Arduino Uno Microprocessor*

(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 pre-programmed 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.

6.1.2 ChucK: An On-the-fly Audio Programming Language Based on C++

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 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 [172].

6.2 Hardware and Software Design

In order make X-Elophone produce different sound than normal xylophone. Two major problem need to be done: 1) circuit design, in order to collect vibration from the xylophone and convert analog signal to digital signal, and 2) software control, in order to transfer digital signals into melody and playback. These design will be discussed in following sections.

6.2.1 Circuit Design

Piezo sensors are attached at the back of each metal bar of the xylophone in order to pick up the vibration. Voltage generated from the Piezo sensor will be compared with the voltage across a potentiometer using the MCP6002 Op-Amp in order to filter out the noises from the signal such as slight move of the instrument. All potentiometers were carefully adjusted individually using oscilloscope to make sure strikes or touches of the note bars could create clear and perfect peaks. One Piezo sensor and one potentiometer are connect in parallel, then series with one Op-Amp. This setup is considered as one line for single note bar which contains 11 lines for the whole system. Output wire from each Op-Amps are connect to multiple input channels (labeled Y0 to Y7) of the multiplexers. 11 lines are connected with two multiplexes in parallel. C6 to B6 are connected in mux-1 correspond to channel Y0 to Y6,and the rest notes are in mux-2 from channel Y0 to Y3. Common output Z1 and Z2 are connect to the analog inputs A0 and A1 of UNO board. Six digital select inputs are connect with the

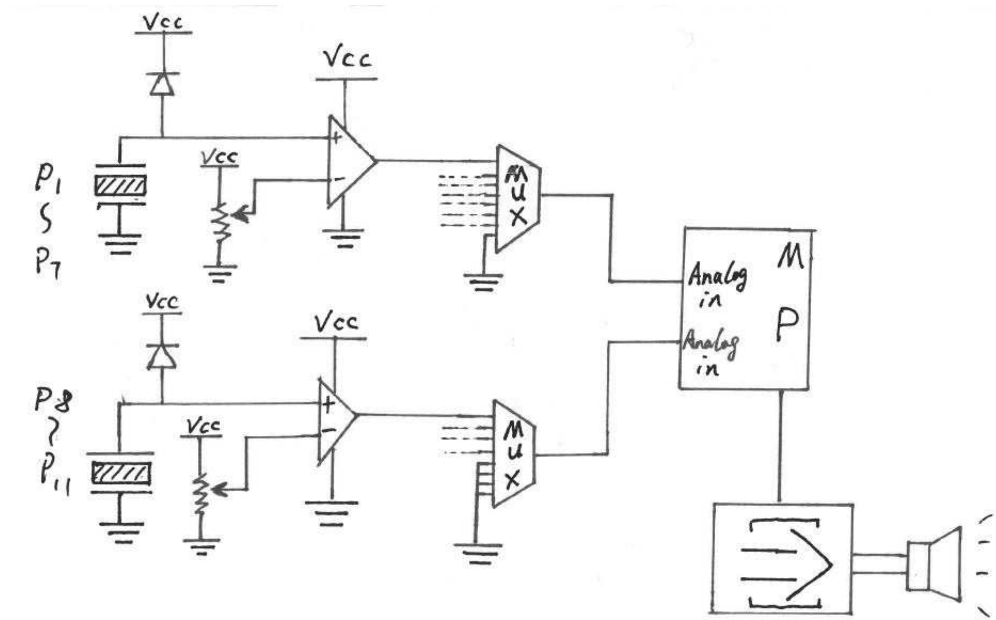


Figure 6.7: *The schematic design for the circuit.*

digital ports from UNO for analyzing the analog inputs from the instrument. Figure 6.7 shows the schematic of circuit, Figure 6.8 shows the actuarial circuit on bread board, and Figure 6.9 shows a signal representation from the oscilloscope of multiple hits through the newly designed xylophone. Yellow signal represents the filtered voltage change of seven strikes from the xylophone, and the green pulse signal represents the output digital signals converted by the mux. Note that the input voltage level were around 2.5v and the output signals are amplified to 5.0v.

6.2.2 Software Design

As mentioned above, well designed circuit has been tested and implemented. However, to make xylophone sound different, software control and design also plays important

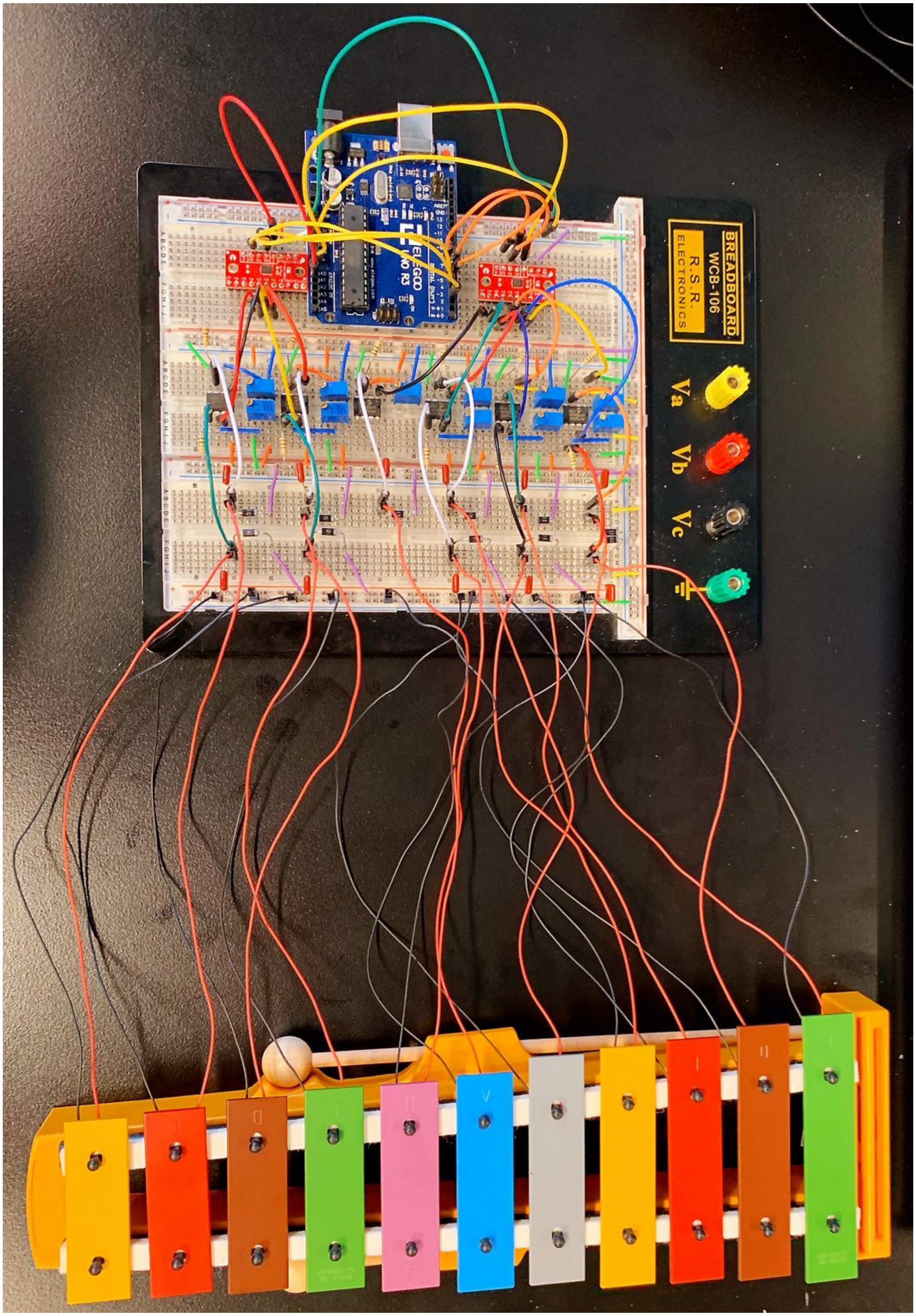


Figure 6.8: *Circuit board in real size.*

Figure 6.9: *A sample input and output: green channel comes from the output of the mux, yellow channel comes from the output of the Op-Amp.*

```
const int selectPins1[3] = {5,6,7}; // mux1 S0~5, S1~6, S2~7
const int selectPins2[3] = {2,3,4}; // mux2 S0~2, S1~3, S2~4
const int zOutput = 5;
const int zInput1 = A1; // Connect common z to A1 for mux1
const int zInput2 = A2; // Connect common z to A2 for mux2
int seq[16];
int hasInput = 0;

void setup() {
  Serial.begin(9600);   // Initialize the serial port
  // Set up the select pins as outputs:
  for (int i=0; i<3; i++)
  {
    pinMode(selectPins1[i], OUTPUT);
    pinMode(selectPins2[i], OUTPUT);
    digitalWrite(selectPins1[i], HIGH);
    digitalWrite(selectPins2[i], HIGH);
  }
  pinMode(zInput1, INPUT);  // Set up z1 as an input for mux1
  pinMode(zInput2, INPUT);  // Set up z2 as an input for mux2
}
```

Figure 6.10: *Arduino code for pin assignment of mutiplexers.*

role. Figure 6.10 to Figure 6.13 shows the code detail from Ardurino in collecting sensor signal and filter out small vibration noise from accident gestures.

One of the most important reason select using ChucK as the music design tool is due to its real-time sound synthesis and music creation. Figure 6.14 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. From Figure 6.15 to Figure 6.18 shows the Chuck code in detail.

```
void loop() {
  // loop through all 11 pins
  for (byte pin=0; pin<=7; pin++)
  {
    selectMuxPin1(pin);   // Select one at a time from mux1
    int inputValue1 = analogRead(A1);  // read z1
    int hitBar1 = checkSensor(inputValue1, pin);
    seq[pin] = hitBar1;
  }
  for (byte pin=0; pin<=7; pin++)
  {
    selectMuxPin2(pin);   // Select one at a time from mux2
    int inputValue2 = analogRead(A2);  // read z2
    int hitBar2 = checkSensor(inputValue2, pin);
    seq[pin+7] = hitBar2;
  }
  hasInput = checkInput(seq);
  if (hasInput == 1){
    for (int j=0; j<16; j++)
    {
      Serial.print(seq[j]);
    }
    Serial.println();
    delay(300);
  }

}
```

Figure 6.11: *Loop in checking all pins for getting signals from the instrument.*

```
// The selectMuxPin function sets the S0, S1, and S2 pins
// accordingly, given a pin from 0-7.
void selectMuxPin1(byte pin)
{
  for (int i=0; i<3; i++)
  {
    if (pin & (1<<i))
      digitalWrite(selectPins1[i], HIGH);
    else
      digitalWrite(selectPins1[i], LOW);
  }
}

void selectMuxPin2(byte pin)
{
  for (int i=0; i<3; i++)
  {
    if (pin & (1<<i))
      digitalWrite(selectPins2[i], HIGH);
    else
      digitalWrite(selectPins2[i], LOW);
  }
}
```

Figure 6.12: *Selecting proper mux by decoding the signals.*

```
int checkSensor(int inputValue, int pin) {
  if (inputValue > 1013) {

    return 1;
  }
  else{
    return 0;
  }
}

int checkInput(int seq[])
{
  for(int i=0; i<16; i++)
  {
    if(seq[i] != 0){
      return 1;
    }
    else
      continue;
  }
  return 0;
}
```

Figure 6.13: *Control the input level in order to filter out small viberation noise to make sure getting meaningful input.*

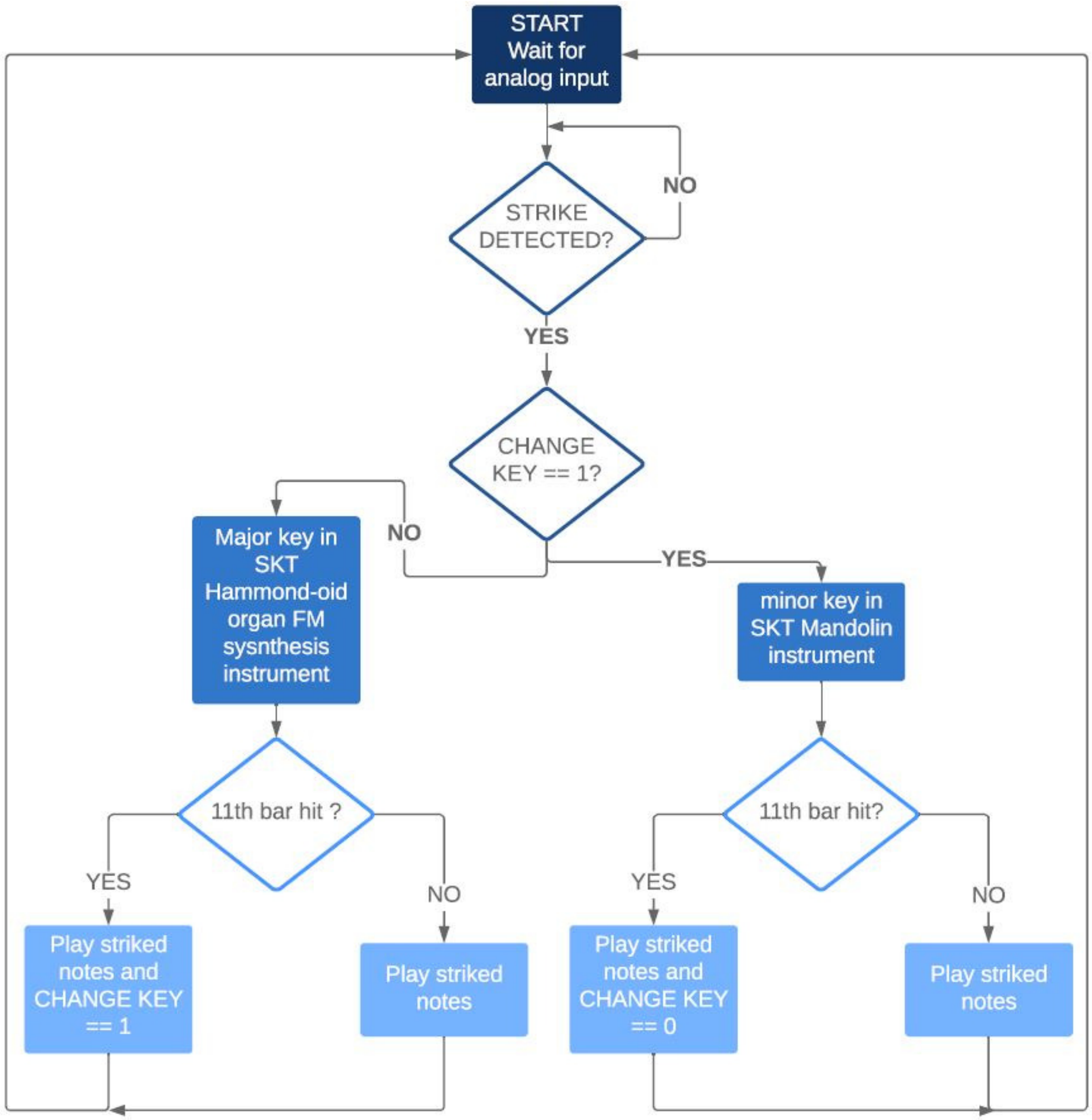


Figure 6.14: *A Flow Chart of Using ChucK in Designing Sound Control System to X-Elophone*

```
// connect to the serial port
SerialIO.list() @=> string list[];
for(int i; i < list.cap(); i++)
{
    chout <= i <= ": " <= list[i] <= IO.newline();
}
// parse first argument as device number
0 => int device;
if(me.args()) {
    me.arg(0) => Std.atoi => device;
}
if(device >= list.cap())
{
    cherr <= "serial device #" <= device <= " not available\n";
    me.exit();
}
SerialIO cereal;
if(!cereal.open(device, SerialIO.B9600, SerialIO.ASCII))
{
    chout <= "unable to open serial device '" <= list[device] <= "'\n";
    me.exit();
}
```

Figure 6.15: *ChucK Code Part I: Allows ChucK program to accept text information as strings via a serial interface specifically for Arduino board at port 9600. Each time Arduino sends a new string, events will be triggered, allowing following code get the string and process.*

```
// Set up the music generators
// synchronize to period
.5::second => dur T;
T - (now % T) => now;
// setup 2 instrument and differnt sound effect to them
BeeThree s => NRev n => Echo e => dac;
Mandolin ss => NRev ee => dac;

.1 => n.mix;
.2 => e.mix;
.2 => ee.mix;

0 => int change;
0 => int cnt;
0 => int select;

<<<"Starting with minor lower key">>>;
// assign different frequencies to two xylophone arrays
[60, 62, 64, 65, 67, 69, 71, 72, 74, 76, 77] @=> int xylo1[];
[48, 50, 51, 53, 55, 56, 58, 60, 62, 63, 65] @=> int xylo2[];
```

Figure 6.16: *ChucK Code Part II: Create 2 STK instrument Mandolin and BeeThree with 2 sets of scales. Different sound effect also been assigned to instrument.*

```
while(true)
{
    // wait for event
    cereal.onLine() => now;
    cereal.getLine() => string line;
    // instrument selection
    if (select == 0)
    {
        1 => select;
    }
    else if(select == 1)
    {
        0 => select;
        continue;
    }
```

Figure 6.17: *ChucK Code Part III: Using 1 and 0 to change instrument sound for broadcasting.*

```
    if(change == 1)
    {
        if (line$Object != null)
        {
            chout <= "read line: " <= line <= IO.newline();
            StringTokenizer tok;
            tok.set(line);
            Std.atoi(line) => int pos;
            chout <= "pos: " <= pos <= IO.newline();
            //playChord(xylo[pos], "major", T)
            Std.mtof(xylol[pos]) => float f;
            chout <= "Freq: " <= f <= IO.newline();
            s.freq(f); // Change sin wave frequency
            .5 => s.noteOn;
            // advance time
            // note: Math.randomf() returns value between 0 and 1
            if( Math.randomf() > .25 ) .25::T => now;
            else .5::T => now;
            if (pos == 10 && cnt == 1)
            {
                0 => change;
                0 => cnt;
                <<< "Play lower key minor next round" >>>;
            }
            else
            {
                1 => change;
                1 => cnt;
            }
        }
            .02 => s.noteOn;
            .0 => s.noteOff;
            0.0 :: second => now;
            chout <= "play done" <= IO.newline();
    }
```

Figure 6.18: *ChucK Code Part IV: Once in instrument 1, when the 11th note gets hit, 'change' value starts to switch the sound to the other instrument. Each note will be played between .25 to .5 second. All play information will be displayed on computer.*

Chapter 7

Discussion, Conclusion and Future Work

7.1 Each Participant Has A Story

In this section, a brief general report from each participant will be given. This would the readers better understand how each child behaved in our study sessions and how different one child is from another when it comes to social behaviors and interaction with our robot and music therapy method.

7.1.1 Subject 101

As the first ASD participant who joined our study, subject 101 provided excellent support for the future intervention sessions for him and other participants. There is one kid song fascinated him greatly, which is called "Baby Shark" taken off since 2016 from Pinkfong, a South Korean education company. Due to the popularity of this music all over the world, it made it a perfect song for children in such music intervention sessions. During the session, subject 101 showed special love in "Baby Shark". Every time the robot asks what he wanted to play, that song was chosen by him. In the game session, when the free play part is one, a "Baby Shark" song was played all the time. To make this more challenging for the subject, three different versions of "Baby Shark" were pre-programmed in the system. Thanks to the simplicity of this song, which allows three different keys to be rearranged to it with the availability of the bars on the current xylophone. With the familiarity of this song, subject 101 provided a stable constant level of performance even with the three keys of that song.

Subject 101 is a music lover and showed a strong passion for playing the instrument. Based on the recorded videos, motor control has been appropriately taught during the first few sessions and nice and clean notes were played by him. However, sometimes subject 101 may hit the bar a bit too hard, which potentially caused the damage to the instrument or the base stand. And there was one time that one of the base handles broke. Fortunately, this issue did not affect the rest of the sessions. Such a high level of engagement supports the confidant of this platform in use of the practice in the future.

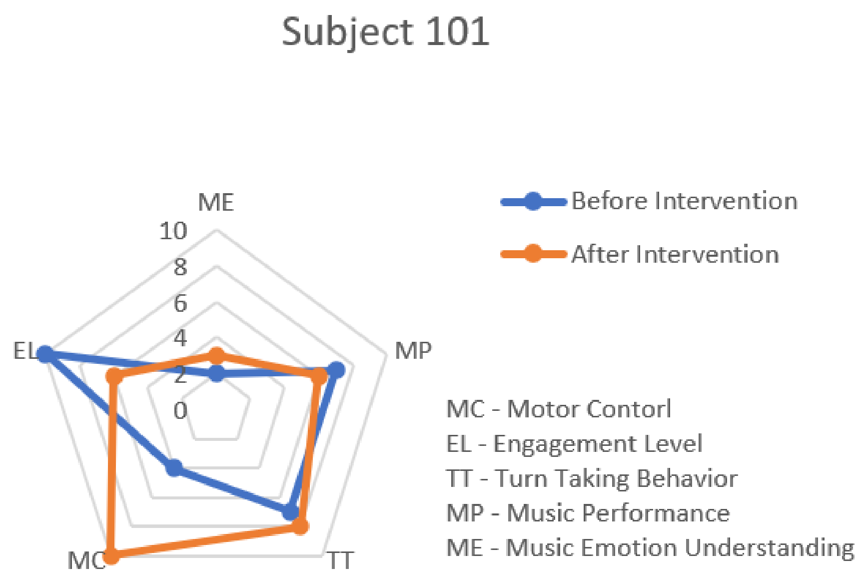


Figure 7.1: *Subject 101*

7.1.2 Subject 102

The most impression from subject 102 is the significant improvement we observed in all the sessions, including all aspects of this study. Playing instruments seemed difficult for him at the beginning. At first this participant had a hard time striking the xylophone accurately. Hitting gesture was somehow challenging to him, and the muffled sound was continuously played even with the warm-up activity. Breaks between activities were often requested across all the sessions, especially during the intervention session with repetitive works. Most of the time he was counting the trials during each activity to leave the experiment room and take a break. However, he never quit any of the sessions or activities, and always completed all the activities as needed.

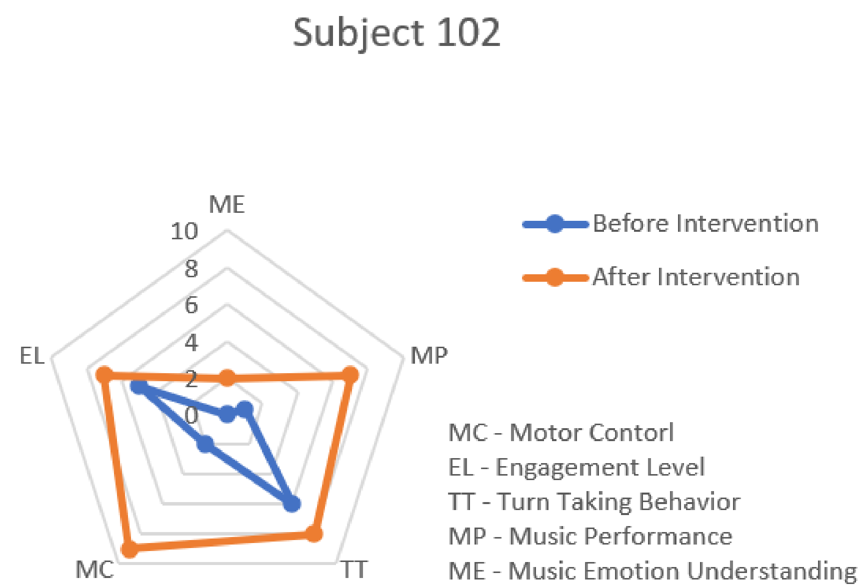


Figure 7.2: *Subject 102*

This participant was willing to learn and improve his skills gradually based on the attitude during the intervention sessions. He started showing no care to the performance at the beginning sessions; to the end, he wanted to play better and appreciated the encouragement from the robot and the researcher. However, it still did not change the fact that he enjoyed taking breaks after each activity and counting the numbers of each trial. The song he used was "Twinkle Twinkle Little Star" due to the fact that there was no favorite song from this subject.

7.1.3 Subject 103

Subject 103 is one of the participants who did not learn much out of this study, which is unfortunate. After all the sessions, he could not learn the hitting technique properly, which means his motor skills were not successfully improved. According to the videos, subject 103 showed acceptable turn-taking behavior across all the sessions. However, he needed some help in concentrating on the tasks most of the time. There was one time that his father had to jump in and interrupt his behaviors while he was not willing to play xylophone but walking and talking of other stuff.

This kid also had some speech issue, which makes the robot had significant trouble with some of the responses from him. For example, the robot could not recognize him saying "yes" and always provided a default behavior. Usually, a "yes" response was requested when participants needed help robot's help. Because the robot could not understand the child's response, it did not deliver the proper content. Hence, whenever the robot asks him whether help is needed, he always responds to "I don't know." To this end, the robot did not have any chance to teach the motor control skills to subject 103.

7.1.4 Subject 104

70% of accuracy for playing a rock song called "I Feel Fantastic" makes his entire session perfect.

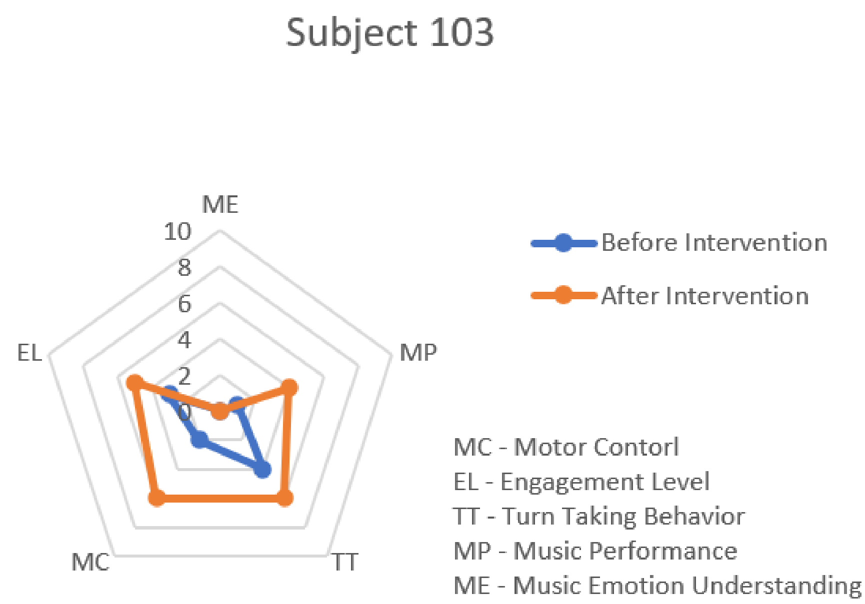


Figure 7.3: *Subject 103*

7.1.5 Subject 105

As mentioned before, there was one activity in this system to ask the feelings of participants regarding randomly generated music by the robot: "How do you feel about what you just heard?". Subject 105 was the only one who shared/expressed his feelings with the robot. This exciting finding showed the potential for children in understanding music emotions. According to the responses from the participants in this study, most of them were thinking about the difficulty in how to playback regarding that random melody, but not how they feel about this piece of music. It is hard to conclude the reason for this result. One possible reason could be the age

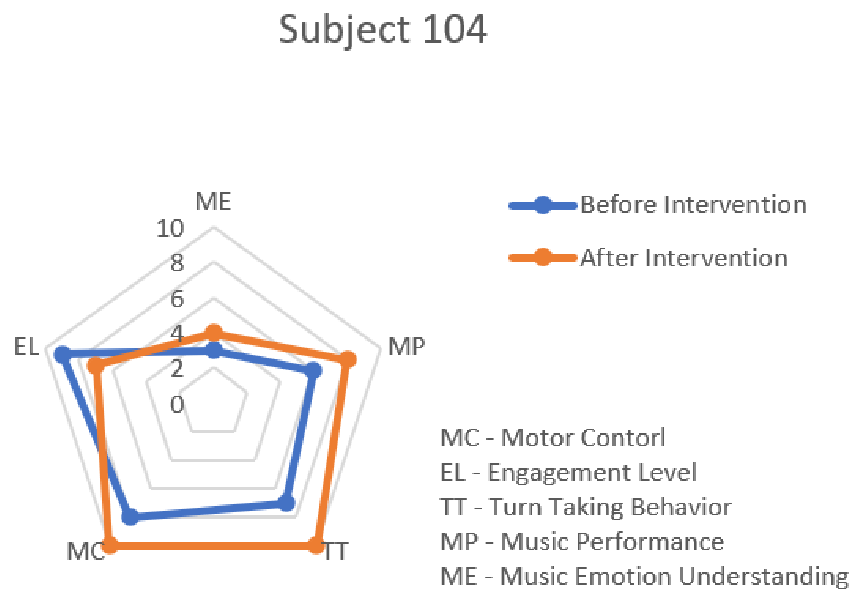


Figure 7.4: *Subject 104*

difference. Older participants may have a better understanding of the questions from the robot as well as the music emotion.

Subject 105 adapted to this music teaching system and performed with a very high accuracy in all the activities. It is also worth mentioning that at the exit session, there was a hidden challenge for all the participants, which added the harmonics to the song they practiced for the past sessions. Subjects would be asked to try this challenge if they wanted. To this end, subject 105 did the best on this task. He almost aced this challenge in playing his requested song, which no other participants could complete this.

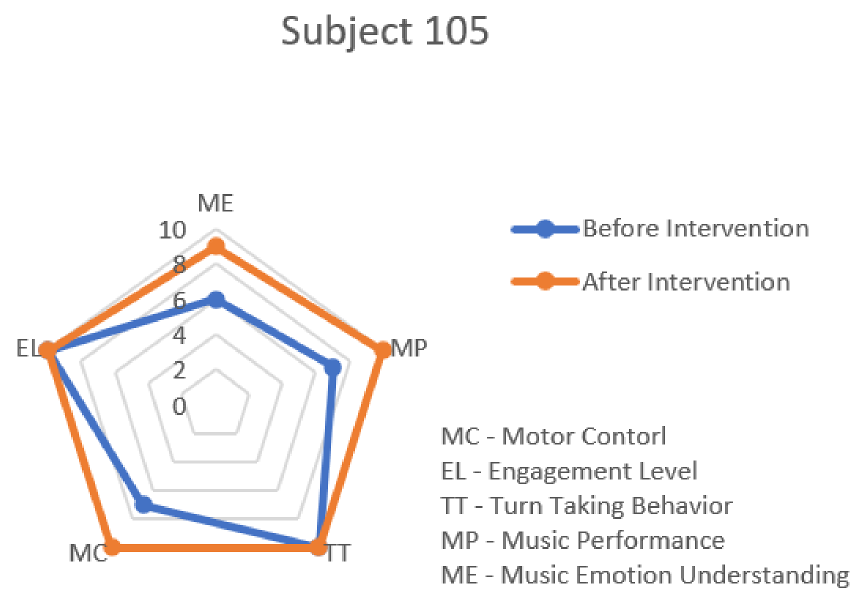


Figure 7.5: *Subject 105*

7.1.6 Subject 106

As the only ASD girl and the only basketball player in this study, one word can be describer her: COMPETITIVE. She was also claimed that she has some experiences in playing the saxophone. Although the instrument was different from a xylophone, the musical experience could have provided some useful knowledge in understanding music concepts such as keys, scales, and melody constructions. Surprisingly, this girl had a difficult time in striking xylophone correctly for the first baseline session, and gradually picked up the technique from the warm-up activity at her first intervention session. Once she got approved by the robot, she began to connect with this research. According to the annotator, subject 106 showed strong engagement in the sessions, focusing on decoding the message one after another. This could also be reflected from the statics result: on average of 80% of the accuracy in the main song practice activity. In this system, a color hint was always be given for all trials; however, the same color sometimes means different pitch on the xylophone. This girl has her way of conquering this challenge. She will play both notes and compare it with the note played by the robot speakers, which makes a perfect play by her. By choosing the song "Three Little Birds" composed by Bob Marley, this performance is impressive. Because she wanted to be better. This could suggest that sports may affect human's mindset and change their behaviors in other activities in daily life.

The most joyful moment for most of the participants would be the free playtime, in this activity it allows kids to challenge NAO to play whatever they just played for 5 seconds each time and with no limitation of melody structure. Particularly with subject 106, she spent most of the time in this section every single session. Also worth to be mentioned is, after each time the robot playback to the participants, NAO will ask for a grade from them. Most of the participants did not take this seriously and sometimes will provide a ridiculously high or low score regardless of the actual performance of the robot. Other than these subjects, participant 106 would carefully rate the robot's performance base on what she felt like, and most of the time is pretty reasonable, according to the researcher in the room. This activity usually took over 15 minutes every time she visits NAO and, most of the time, have to end the session manually from the computer side.

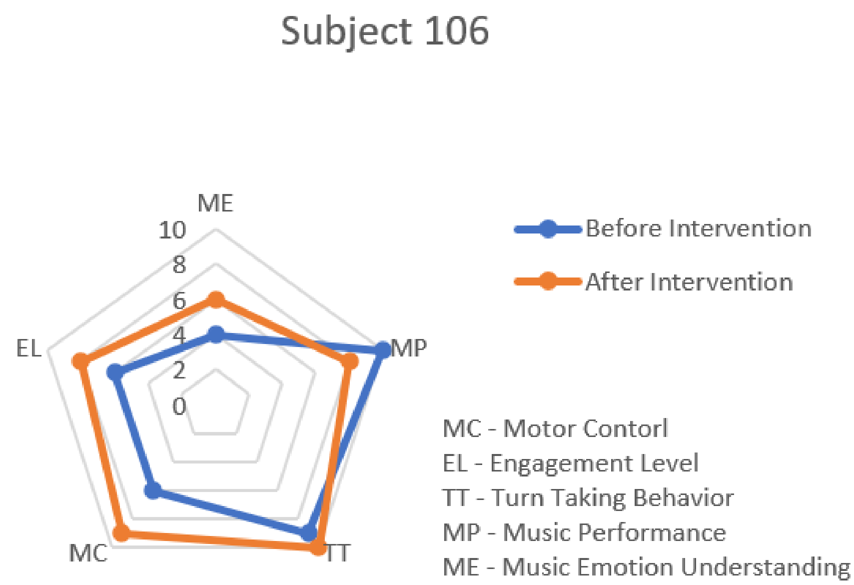


Figure 7.6: *Subject 106*

7.1.7 Subject 107

At the beginning of the session, this subject needed significant help from his caregiver for the entire time. It is also apparent to notice that he may request extra help in future visits. According to his mom, music therapy treatment had been given to subject 107 before, and it is also clear that he did enjoy this method of treatment. The baseline session was not good; this boy did not listen to the robot and could not provide a meaningful turn-taking behavior then. Since he had music therapy before, it was unclear whether the therapist used xylophone before or how frequently it has been used. Regardless of this uncleared question, subject 107 did a good job of playing the xylophone, and a nice and clear note could be played by him for most of the time.

Subject 107 showed signs in constant repetitive hand gestures during sessions, which somehow made it difficult to hold the mallet properly in the first few sessions. This fact created some delay in responding to the robot during the intervention session. From Figure 5.2, it is easy to see that this subject could not follow the turn-taking rule properly; most of the results were not recorded in the right way. However, according to the report of the annotators, "...this participant was able to understand the color hint from the robot and may provide correct input to the robot but off the time limit, which will not be recorded in the system and could be determined as incorrect answers from the computer side...".

As session ongoing, this particular participant started to show more engagement behavior for the last intervention session. The first verbal response to the question from the robot happened in session 5. After all the intervention, finally, subject 107 started to respond to the music agent for the very first time, and multiple times in this session as well positively. One significant difference reflected in the music playing activity. The participant started to repeat the color names while trying to strike the bars accordingly. This phenomenon may suggest that for some particular users, they may need more time to adjust themselves in getting used to the new system or platform. Besides, proper turn-taking behavior can be improved by using this platform after intervention sessions.

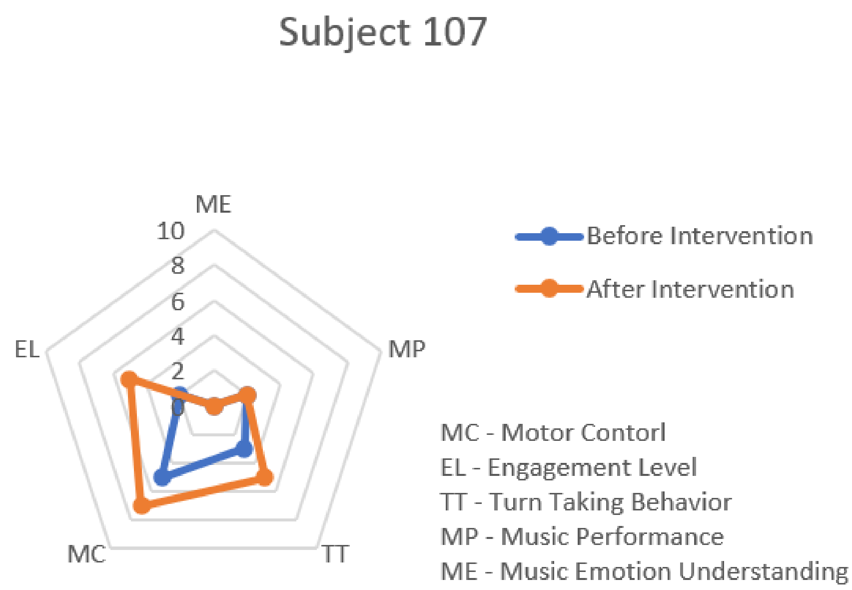


Figure 7.7: *Subject 107*

7.1.8 Subject 108

One talented musician among all subjects. Subject 108, a violin player, who suppose to have the most precise ear playing ability, surprisingly did not perform very well from the static result. 68.75% is the accuracy for the customize song play across all intervention sessions. After dig into his files, the mystery starts to appear. The biggest challenge is the song he chooses: "Can Can" by Offenbach. Such a famous and difficult classical music in string instrument. And he accepted this challenge willingly. From the performance result, subject 108 shows decreasing in the play accuracy rate after sessions, which makes perfect sense in such a high-level challenge. Based on the setting of this practice, the notes will significantly increase at the last two intervention sessions in order to have the whole song or main melody covered to be taught to the

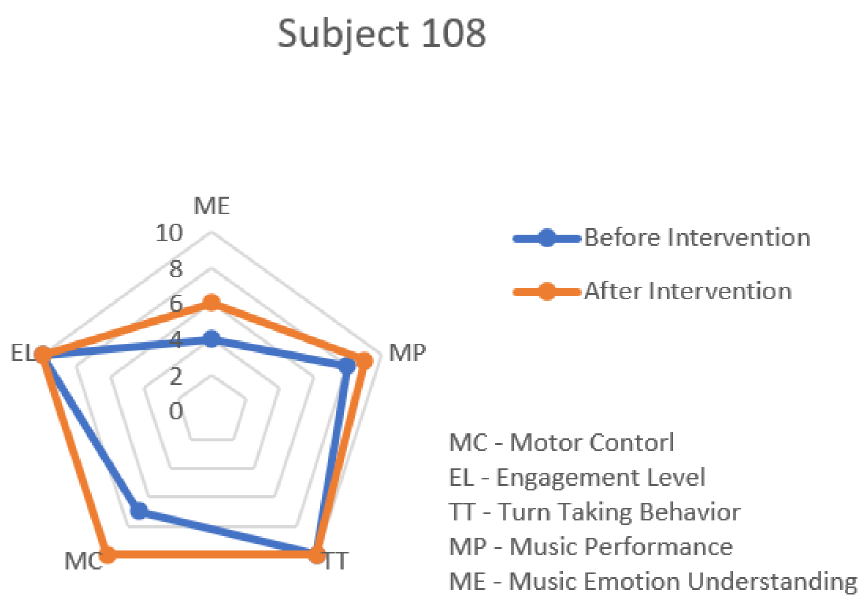


Figure 7.8: *Subject 108*

kids. "Can Can" which has already included a massive amount of notes in it, and this could suggest the lousy performance at the last intervention session of subject 108. However, almost 70% of accuracy in such a problematic song; he did an ace job.

Quiet, focus, pinpoint, can be the best definition of subject 108. This participant was able to complete the tasks smoothly with a high level of engagement. At the very beginning of the session, he also confused about the technique of how to play percussion properly since it is different from strings. However, it did not take long for him to figure out. Starting from the first intervention session, 100% of the accuracy from the warm-up task explains everything.

7.1.9 Subject 109

Unstoppable can be suitable to describe this participant. According to his caregiver, subject 109 has a hearing disability from one side of his ear. However, after all sessions, this fact did not seem to be an influence on his performance in playing music. Ever since he started this study, he showed hyperactive in the experiment room with the robot. Among all kids, he was the one who touched the robot most, and there was one time almost accidentally pushed the robot to lean backward. This makes the researcher had to pay extra attention to protecting both participants and robots not to get injured. Impatience was also an issue for subject 109; the researcher had to spend time to help this participant focusing on the tasks most of the time.

Although the session did not go smoothly as usual, the music performance from subject 109 was acceptable. Without much help, he quickly picked up the striking technique from the robot and be able to hit the notes accurately. One thing that needs to be mentioned is his hearing ability, he is one of the few participants who have the sensitivity in distinguishing different pitches with the same color bars by only listen to the sound. Somehow even better then the ones who claimed has the musical experience, for example, subject 108. Subject 108 is gifted. And after few sessions, he began to accept this platform, and showed more extended concentration period in the last few sessions, which provided the potential uses for this assistive music teaching platform a chance in daily life.

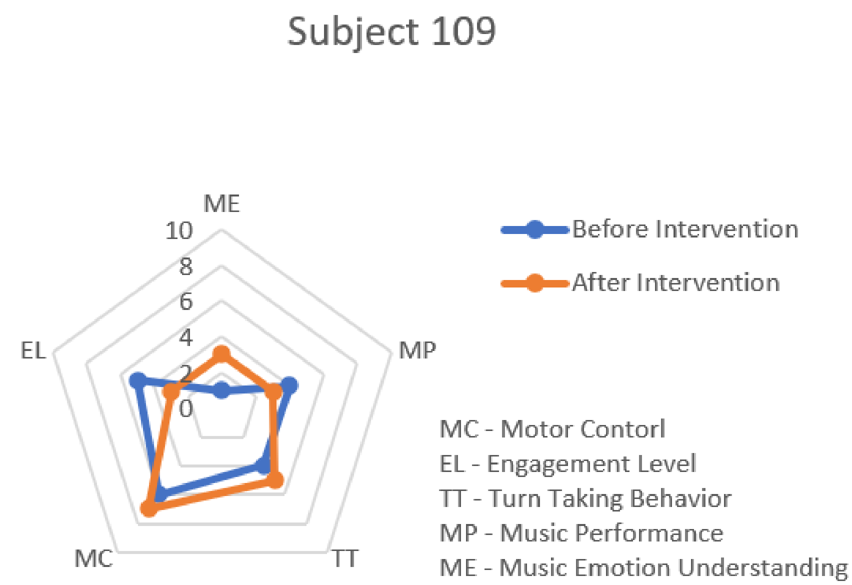


Figure 7.9: *Subject 109*

7.2 Discussion and Conclusion

The results of our study indicate that the conferred music education platform will be thought-about as a decent tool to facilitate improving fine motor control, turn-taking skills, and social activities engagement in children with autism. The automated music detection system created a self-adjusted surrounding for participants in early sessions. Most of the ASD youngsters began to develop the strike movement whiting the initial 2 intervention sessions; some even mastered the motor ability throughout the very first warm-up event. Though the robot might provide verbal directions and demonstrations by voice command input from participants whenever they need it. However, the majority of the participants did not request such a service while playing with NAO. This finding suggests that fine motor control ability will be learned from specific well-designed activities by the young ASD population.

The purpose of using a music teaching scenario as the main activity in the current research is to give the kids a chance for practicing fine and natural turn-taking behaviors during social interactions. By observing all experimental sessions, 6 out of 9 subjects could dominate proper turn-taking skill after one or two sessions. Note that subject 107 had significant improvement in the last few sessions compared to the baseline session. Subject 109 had trouble listening to the robot for most of the time. However, with the researcher's help, this kid could perform better music activities with social interaction skills for a short time period. For practicing turn-taking skills, a fun, motivating activity should be designed for children with autism. Music teaching could be a good example for accomplishing this task by taking the advantage of customized songs selected by each individual.

Starting the latter half of the sessions, participants could start to recognize their favorite songs, where over half of the participants were getting more into the activities, although the difficulty level for playing proper notes was much higher. It was 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 the more complex the music was, the more challenging it was and more concentration was needed by the participants. Thus, older individuals might be willingly to accept the challenges and enjoy the sense of accomplishment afterward based on their verbal feedback to the research at the end of each session. Having some knowledge of music could also be one of the reasons that caused this result, since older participants might have had more chance to learn music at school. High engagement level can be found in the music game play section among all sessions. The reason behind it is not only because of the music game was fun and more relax, but also provided an opportunity in switching the role from student to instructor for participants. Some of the participants enjoyed this role play activity especially for subject 106, who spent a significant amount of time in free play game mode, this was more notable. According to the session executioner and video annotators, this particular subject (106) showed a high level of engagement in all the activities including the free play. Based on the conversation and music performance with the robot, the subject showed a strong interest in challenging the robot in a friendly way.

Measuring emotion in children with autism is often difficult and challenging. Biosignals provide a possible way of doing that. The vvent-based emotion classification method presented in this research suggests that the same activity with different intensities can cause emotion changes in the arousal dimension, although it is difficult to label the emotions based on facial expression changes in the video annotation phase for the ASD group. Less emotion expressed in particular activities 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 that helped participants to focus on learning music with proper communication behaviors. This may explain the improvement of music play performance in the song practice (S2) through the intervention sessions illustrated in Figure 5.2. Comparing emotion patterns from baseline and exit sessions between TD and ASD groups in Table 5.3, Some differences could be found. This may suggest a potential way to assist autism diagnoses using bio-signals an early age. According to annotators and observers, TD kids showed a strong passion. Excitement, stressful, and disappointment were easy to be recognized and labeled from the recorded videos. On the other hand, limited facial expression changes could be detected in the ASD group. That makes it challenging to learn whether they have different feelings or have the same feelings but different bio-signal activities compared to the TD group. Further research need to be done to better understand these issues. Furthermore, due to the limitation of the sample size, future research can focus on this problem with different classification methods and with a larger sample size.

7.3 Future Work: New Style Session Proposal

Our newly designed electric xylophone has two significant differences compared to the original acoustic one. The most obvious improvement is the sound. Various timbers, keys, and scales can be programmed on the board and switched in real-time, which provides the infinite possibilities with limited note bars. Gentle touch has also been embedded in the play style. Previously, one with only proper motor control could strike a sweet melody. However, by using this new design, one soft touch of the fine-tuned bar will also provide a friendly and clean note out of the speaker. These two designs provide unlimited possibilities of music play based on the proposed music teaching platform.

Bibliography

[1] Leo Kanner. Early infantile autism. *The Journal of Pediatrics*, 1944.

[2] Lorna Wing. The autistic spectrum. *The lancet*, 350(9093):1761–1766, 1997.

[3] Catherine Lord, Edwin H Cook, Bennett L Leventhal, and David G Amaral. Autism spectrum disorders. *Neuron*, 28(2):355–363, 2000.

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

[5] Matthew J Maenner, Kelly A Shaw, Jon Baio, et al. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2016. *MMWR Surveillance Summaries*, 69(4):1, 2020.

[6] National Research Council et al. *Educating children with autism*. National Academies Press, 2001.

[7] Sonja de Boer-Ott, Deborah E Griswold, Brenda Smith Myles, Sara E Byrd, Jennifer B Ganz, Katherine T Cook, Kaye Otten, Josefa Ben-Arieh, Sue Ann Kline, Lisa Garriott Adams, et al. *Autism spectrum disorders: Interventions and treatments for children and youth*. Corwin press, 2004.

[8] John O Cooper, Timothy E Heron, William L Heward, et al. Applied behavior analysis. 2007.

[9] Randi A Sarokoff and Peter Sturmey. The effects of behavioral skills training on staff implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis*, 37(4):535–538, 2004.

[10] Karen Pierce and Laura Schreibman. Increasing complex social behaviors in children with autism: Effects of peer-implemented pivotal response training. *Journal of applied behavior analysis*, 28(3):285–295, 1995.

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

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

[13] Sulammith Wolff and Stella Chess. A behavioural study of schizophrenic children. *Acta Psychiatrica Scandinavica*, 1964.

[14] Pat Mirenda. Autism, augmentative communication, and assistive technology: What do we really know? *Focus on autism and other developmental disabilities*, 16(3):141–151, 2001.

[15] Auriol Drew, Gillian Baird, Simon Baron-Cohen, Antony Cox, Vicky Slonims, Sally Wheelwright, John Swettenham, Bryony Berry, and Tony Charman. A pilot randomised control trial of a parent training intervention for pre-school children with autism. *European child & adolescent psychiatry*, 11(6):266–272, 2002.

[16] Norman Epstein and Donald H Baucom. Cognitive-behavioral marital therapy. In *Comprehensive handbook of cognitive therapy*, pages 491–513. Springer, 1989.

[17] Eric Schopler, Gary B Mesibov, and Kathy Hearsey. Structured teaching in the teacch system. In *Learning and cognition in autism*, pages 243–268. Springer, 1995.

[18] Carol A Gray and Joy D Garand. Social stories: Improving responses of students with autism with accurate social information. *Focus on autistic behavior*, 8(1):1–10, 1993.

[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] 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.

[21] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166, 2003.

[22] Thomas Zane, Cheryl Davis, and Mary Rosswurm. The cost of fad treatments in autism. *Journal of Early and Intensive Behavior Intervention*, 5(2):44, 2008.

[23] Brenda Laurel. *Computers as theatre*. Addison-Wesley, 2013.

[24] Don Tapscott. *Growing up digital*, volume 302. McGraw-Hill Companies San Francisco, 1998.

[25] Justine Cassell and Henry Jenkins. *From Barbie to Mortal Kombat: gender and computer games*. MIT press, 2000.

[26] Allison Druin, James A Hendler, and James Hendler. *Robots for kids: exploring new technologies for learning*. Morgan Kaufmann, 2000.

[27] Kerstin Dautenhahn. The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied artificial intelligence*, 12(7-8):573–617, 1998.

[28] Iain Werry and Kerstin Dautenhahn. Applying mobile robot technology to the rehabilitation of autistic children. In *In: Procs SIRS99, 7th Symp on Intelligent Robotic Systems*, 1999.

[29] David Feil-Seifer and Maja J Mataric. Defining socially assistive robotics. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, pages 465–468. IEEE, 2005.

[30] 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.

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

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

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

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

[35] Andrea C Pierno, Morena Mari, Dean Lusher, and Umberto Castiello. Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46(2):448–454, 2008.

[36] Michael Villano, Charles R Crowell, Kristin Wier, Karen Tang, Brynn Thomas, Nicole Shea, Lauren M Schmitt, and Joshua J Diehl. Domer: A wizard of oz interface for using interactive robots to scaffold social skills for children with autism spectrum disorders. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 279–280. IEEE, 2011.

[37] 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.

[38] Kerstin Dautenhahn and Iain Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35, 2004.

[39] 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.

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

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

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

[43] Alaine E Reschke-Hernández. History of music therapy treatment interventions for children with autism. *Journal of Music Therapy*, 48(2):169–207, 2011.

[44] Michael H Thaut and Alicia Ann Clair. *A scientific model of music in therapy and medicine*. IMR Press, 2000.

[45] Elizabeth Starr and Krista Zenker. Understanding autism in the context of music therapy: Bridging theory and practice. *Canadian Journal of Music Therapy*, 1998.

[46] Alana K Dellatan. The use of music with chronic food refusal: A case study. *Music Therapy Perspectives*, 21(2):105–109, 2003.

[47] Mike D Brownell. Musically adapted social stories to modify behaviors in students with autism: Four case studies. *Journal of music therapy*, 39(2):117–144, 2002.

[48] Petra Kern and David Aldridge. Using embedded music therapy interventions to support outdoor play of young children with autism in an inclusive community-based child care program. *Journal of music therapy*, 43(4):270–294, 2006.

[49] June Katagiri. The effect of background music and song texts on the emotional understanding of children with autism. *Journal of Music Therapy*, 46(1):15–31, 2009.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

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

[55] Victor Lotter. Epidemiology of autistic conditions in young children. *Social psychiatry*, 1(3):124–135, 1966.

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

[57] Elizabeth Redcay and Eric Courchesne. Deviant functional magnetic resonance imaging patterns of brain activity to speech in 2–3-year-old children with autism spectrum disorder. *Biological psychiatry*, 64(7):589–598, 2008.

[58] Karl Popper. *The logic of scientific discovery.* Routledge, 2005.

[59] Lawrence Bartak, Michael Rutter, and Antony Cox. A comparative study of infantile autism and specific developmental receptive language disorder: I. the children. *The British Journal of Psychiatry*, 126(2):127–145, 1975.

[60] Emma Gowen and Antonia Hamilton. Motor abilities in autism: a review using a computational context. *Journal of autism and developmental disorders*, 43(2):323–344, 2013.

[61] Philip Teitelbaum, Osnat Teitelbaum, Jennifer Nye, Joshua Fryman, and Ralph G Maurer. Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences*, 95(23):13982–13987, 1998.

[62] Beth Provost, Brian R Lopez, and Sandra Heimerl. A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. *Journal of autism and developmental disorders*, 37(2):321–328, 2007.

[63] Jessica Brian, SE Bryson, Nancy Garon, Wendy Roberts, Isabel M Smith, Peter Szatmari, and Lonnie Zwaigenbaum. Clinical assessment of autism in high-risk 18-month-olds. *Autism*, 12(5):433–456, 2008.

[64] Kimberly A Fournier, Chris J Hass, Sagar K Naik, Neha Lodha, and James H Cauraugh. Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of autism and developmental disorders*, 40(10):1227–1240, 2010.

[65] Xue Ming, Michael Brimacombe, and George C Wagner. Prevalence of motor impairment in autism spectrum disorders. *Brain and Development*, 29(9):565–570, 2007.

[66] Hilde Van Waelvelde, Ann Oostra, Griet Dewitte, Christine Van den Broeck, and Marian J Jongmans. Stability of motor problems in young children with or at risk of autism spectrum disorders, adhd, and or developmental coordination disorder. *Developmental Medicine & Child Neurology*, 52(8):e174–e178, 2010.

[67] Mohammad Ghaziuddin, E Butler, L t Tsai, and N Ghaziuddin. Is clumsiness a marker for asperger syndrome? *Journal of intellectual disability research*, 38(5):519–527, 1994.

[68] Janine Manjiviona and Margot Prior. Comparison of asperger syndrome and high-functioning autistic children on a test of motor impairment. *Journal of autism and developmental disorders*, 25(1):23–39, 1995.

[69] Motohide Miyahara, Masatsugu Tsujii, Miwako Hori, Kazunori Nakanishi, Hidenori Kageyama, and Toshiro Sugiyama. Brief report: motor incoordination in children with asperger syndrome and learning disabilities. *Journal of autism and developmental disorders*, 27(5):595–603, 1997.

[70] Emmanuelle Jasmin, Mélanie Couture, Patricia McKinley, Greg Reid, Eric Fombonne, and Erika Gisel. Sensori-motor and daily living skills of preschool children with autism spectrum disorders. *Journal of autism and developmental disorders*, 39(2):231–241, 2009.

[71] Saasha Sutera, Juhi Pandey, Emma L Esser, Michael A Rosenthal, Leandra B Wilson, Marianne Barton, James Green, Sarah Hodgson, Diana L Robins, Thyde Dumont-Mathieu, et al. Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of autism and developmental disorders*, 37(1):98–107, 2007.

[72] K Tang, JJ Diehl, M Villano, K Wier, B Thomas, N Shea, Z DuBois, MA Millea, KA Uhland, and CR Crowell. Enhancing empirically-supported treatments for autism spectrum disorders: A case study using an interactive robot. In *Poster session presented at the International Meeting for Autism Research, San Diego, CA*, 2011.

[73] Heather F Ajzenman, John W Standeven, and Tim L Shurtleff. Effect of hippotherapy on motor control, adaptive behaviors, and participation in children with autism spectrum disorder: a pilot study. *American Journal of Occupational Therapy*, 67(6):653–663, 2013.

[74] Barbara Teichmann Engel and Joyce R MacKinnon. *Enhancing human occupation through hippotherapy: A guide for occupational therapy.* AOTA Press, 2007.

[75] Natalie Bennett Brown and Winnie Dunn. Relationship between context and sensory processing in children with autism. *American Journal of Occupational Therapy*, 64(3):474–483, 2010.

[76] Beth Provost, Sandra Heimerl, and Brian R Lopez. Levels of gross and fine motor development in young children with autism spectrum disorder. *Physical & Occupational Therapy in Pediatrics*, 27(3):21–36, 2007.

[77] Richard S Neel, Zoe Neibrugge Jenkins, and Nancy Meadows. Social problem-solving behaviors and aggression in young children: A descriptive observational study. *Behavioral Disorders*, 16(1):39–51, 1990.

[78] T Berry Brazelton, Barbara Koslowski, and Mary Main. The origins of reciprocity: The early mother-infant interaction. 1974.

[79] Colwyn Trevarthen. Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. *Action, gesture, and symbol: The emergence of language*, 1978.

[80] Vasudevi Reddy. Playing with others' expectations: Teasing and mucking about in the first year. 1991.

[81] Catherine Lord. The development of peer relations in children with autism. *Advances in applied developmental psychology*, pages 165–229, 1984.

[82] Peter Mundy, Marian Sigman, Judy Ungerer, and Tracy Sherman. Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of child psychology and psychiatry*, 27(5):657–669, 1986.

[83] Louise A Kaczmarek. *Assessment of social-communicative competence: An interdisciplinary model.* Paul H Brookes Publishing Co., 2002.

[84] Andrea Sweetman McDuffie. *The relationship of four prelinguistic skills and rapid word-object mapping to vocabulary outcomes in young children with autism.* PhD thesis, Vanderbilt University, 2004.

[85] Marian Sigman, Ellen Ruskin, Shoshana Arbelle, Rosalie Corona, Cheryl Dissanayake, Michael Espinosa, Norman Kim, Alma López, Cynthia Zierhut, Carolyn B Mervis, et al. Continuity and change in the social competence of children with autism, down syndrome, and developmental delays. *Monographs of the society for research in child development*, pages i–139, 1999.

[86] Wendy L Stone and Paul J Yoder. Predicting spoken language level in children with autism spectrum disorders. *Autism*, 5(4):341–361, 2001.

[87] Catherine Aldred, Jonathan Green, and Catherine Adams. A new social communication intervention for children with autism: pilot randomised controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry*, 45(8):1420–1430, 2004.

[88] Peter Mundy and Mary Crowson. Joint attention and early social communication: Implications for research on intervention with autism. *Journal of Autism and Developmental disorders*, 27(6):653–676, 1997.

[89] Fifth Edition et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, 2013.

[90] KE Diamond, S Hong, and AE Baroody. Promoting young children's social competence in early childhood programs. *Social competence of young children: Risk, disability, and intervention*, pages 165–184, 2008.

[91] Michael J Guralnick and Brian Neville. Designing early intervention programs to promote children's social competence. *The effectiveness of early intervention*, 579:610, 1997.

[92] Amanda W Harrist and Ralph M Waugh. Dyadic synchrony: Its structure and function in children's development. *Developmental review*, 22(4):555–592, 2002.

[93] Sarah R Rieth, Aubyn C Stahmer, Jessica Suhrheinrich, Laura Schreibman, Joanna Kennedy, and Benjamin Ross. Identifying critical elements of treatment: Examining the use of turn taking in autism intervention. *Focus on Autism and Other Developmental Disabilities*, 29(3):168–179, 2014.

[94] Jeroen CJ Brok and Emilia I Barakova. Engaging autistic children in imitation and turn-taking games with multiagent system of interactive lighting blocks. In *International conference on entertainment computing*, pages 115–126. Springer, 2010.

[95] Joshua J Diehl, Lauren M Schmitt, Michael Villano, and Charles R Crowell. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in autism spectrum disorders*, 6(1):249–262, 2012.

[96] Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14, 2012.

[97] Paul J Yoder and Steven F Warren. Effects of prelinguistic milieu teaching and parent responsivity education on dyads involving children with intellectual disabilities. *Journal of Speech, Language, and Hearing Research*, 2002.

[98] Paul Yoder and Wendy L Stone. Randomized comparison of two communication interventions for preschoolers with autism spectrum disorders. *Journal of consulting and clinical psychology*, 74(3):426, 2006.

[99] Daniel B LeGoff. Use of lego© as a therapeutic medium for improving social competence. *Journal of autism and developmental disorders*, 34(5):557–571, 2004.

[100] Daniel B Legoff and Michael Sherman. Long-term outcome of social skills intervention based on interactive lego© play. *Autism*, 10(4):317–329, 2006.

[101] Gina Owens, Yael Granader, Ayla Humphrey, and Simon Baron-Cohen. Lego® therapy and the social use of language programme: An evaluation of two social skills interventions for children with high functioning autism and asperger syndrome. *Journal of autism and developmental disorders*, 38(10):1944, 2008.

[102] Sojung Kim and Elizabeth Clarke. Case study: An ipad-based intervention on turn-taking behaviors in preschoolers with autism. *Behavioral Development Bulletin*, 20(2):253, 2015.

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

[104] 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.

[105] 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.

[106] 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.

[107] 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.

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

[109] 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.

[110] Vasudevi Reddy, Dale Hay, Lynne Murray, and Colwyn Trevarthen. Communication in infancy: Mutual regulation of affect. *Infant development: Recent advances*, page 247, 1997.

[111] Rita Jordan and Sarah Libby. Developing and using play in the curriculum. In *Autism and Learning (Classic Edition)*, pages 35–49. Routledge, 2011.

[112] Dawn C Wimpory and Susan Nash. Musical interaction therapy-therapeutic play for children with autism. *Child Language Teaching and Therapy*, 15(1):17–28, 1999.

[113] Bronwen Burford. Action cycles: Rhythmic actions for engagement with children and young adults with profound mental handicap. *European Journal of Special Needs Education*, 3(4):189–206, 1988.

[114] Jerome Bruner. Child's talk: Learning to use language. *Child Language Teaching and Therapy*, 1(1):111–114, 1985.

[115] Daniel J Ricks and Mark B Colton. Trends and considerations in robot-assisted autism therapy. In *2010 IEEE international conference on robotics and automation*, pages 4354–4359. IEEE, 2010.

[116] David Feil-Seifer and Maja Mataric'. Robot-assisted therapy for children with autism spectrum disorders. In *Proceedings of the 7th international conference on Interaction design and children*, pages 49–52, 2008.

[117] Ben Robins, Kerstin Dautenhahn, Rene Te Boekhorst, and Aude Billard. Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Universal access in the information society*, 4(2):105–120, 2005.

[118] Ben Robins, Kerstin Dautenhahn, and Janek Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction studies*, 7(3):479–512, 2006.

[119] Hideki Kozima, Cocoro Nakagawa, and Yuriko Yasuda. Interactive robots for communication-care: A case-study in autism therapy. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 341–346. IEEE, 2005.

[120] Giovanni Pioggia, Roberta Igliozzi, Marcello Ferro, Arti Ahluwalia, Filippo Muratori, and Danilo De Rossi. An android for enhancing social skills and emotion recognition in people with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(4):507–515, 2005.

[121] Kerstin Dautenhahn and Iain Werry. Issues of robot-human interaction dynamics in the rehabilitation of children with autism. *Proc. From animals to animats*, 6:519–528, 2000.

[122] Iain Werry, Kerstin Dautenhahn, and William Harwin. Investigating a robot as a therapy partner for children with autism. *Procs AAATE 2001,*, 2001.

[123] Syamimi Shamsuddin, Hanafiah Yussof, Luthffi Ismail, Fazah Akhtar Hanapiah, Salina Mohamed, Hanizah Ali Piah, and Nur Ismarrubie Zahari. Initial response of autistic children in human-robot interaction therapy with humanoid robot nao. In *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, pages 188–193. IEEE, 2012.

[124] Syamimi Shamsuddin, Hanafiah Yussof, Luthffi Idzhar Ismail, Salina Mohamed, Fazah Akhtar Hanapiah, and Nur Ismarrubie Zahari. Initial response in hri-a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot nao. *Procedia Engineering*, 41:1448–1455, 2012.

[125] Syamimi Shamsuddin, Hanafiah Yussof, Luthffi Idzhar Ismail, Salina Mohamed, Fazah Akhtar Hanapiah, and Nur Ismarrubie Zahari. Humanoid robot nao interacting with autistic children of moderately impaired intelligence to augment communication skills. *Procedia Engineering*, 41:1533–1538, 2012.

[126] Syamimi Shamsuddin, H Yussof, MA Miskam, A Che Hamid, Norjasween Abdul Malik, and Hafizan Hashim. Humanoid robot nao as hri mediator to teach emotions using game-centered approach for children with autism. In *HRI 2013 Workshop on Applications for Emotional Robots*, 2013.

[127] 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.

[128] 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.

[129] 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.

[130] 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.

[131] 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.

[132] 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.

[133] 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.

[134] 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.

[135] 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.

[136] 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.

[137] 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.

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

[139] 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.

[140] 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.

[141] 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.

[142] 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.

[143] 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.

[144] 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.

[145] 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.

[146] 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.

[147] 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.

[148] 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.

[149] 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.

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

[151] 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.

[152] 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.

[153] 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.

[154] 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.

[155] 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.

[156] 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.

[157] 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.

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

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

[160] 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.

[161] 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.

[162] 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.

[163] 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.

[164] 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.

[165] 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.

[166] 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.

[167] 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.

[168] 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.

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

[170] 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.

[171] 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.

[172] Ge Wang, Perry R Cook, et al. Chuck: A concurrent, on-the-fly, audio programming language. In *ICMC*, 2003.

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 July 2020.

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