University of Denver Digital Commons @ DU

Electronic Theses and Dissertations

Graduate Studies

2020

Xylo-Bot: A Therapeutic Robot-Based Music Platform for Children with Autism

Huanghao Feng

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Child Psychology Commons, Clinical Psychology Commons, and the Robotics Commons

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

A Dissertation

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

In Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

by

Huanghao Feng

August 2020

Advisor: Mohammad H. Mahoor, Ph.D.

©Copyright by Huanghao Feng 2020 All Rights Reserved Author: Huanghao Feng Title: XYLO-BOT: A THERAPEUTIC ROBOT-BASED MUSIC PLATFORM FOR CHILDREN WITH AUTISM Advisor: Mohammad H. Mahoor, Ph.D. Degree Date: August 2020

Abstract

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 without music experience. 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 within the two initial intervention sessions; some even mastered the motor ability throughout the early events. More than half of the subjects could dominate proper turn-taking after few sessions. Music teaching is a good example for accomplishing social skill tasks by taking advantage of customized songs selected by individuals. According to researcher and video annotator, majority of the subjects showed high level of engagement for all music game activities, especially with the free play mode. Based on the conversation and music performance with NAO, subjects showed strong interest in challenging the robot with a friendly way.

Acknowledgements

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. 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 girl-friend, 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.

Table of Contents

 1.1 Autism Spectrum Disorders (ASD) 1.2 Socially Assistive Robotics (SAR) 1.2.1 Socially Assistive Robots for Autism Therapy 1.3 Music Therapy in ASD Treatment 1.4 Contributions 1.5 Organization 2 Related Works in Autism and Robots 2.1 Autism 2.1.1 Motor Control 2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.4 Human Robot Interaction in Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic Robots Designs for Autism 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.4 Lassification and English and the Market System 4 Xylo-Bot: An Interactive Music Teaching System 4.1 Hardware Selection and Design 	1	Intr	oduction	1
 1.2 Socially Assistive Robotics (SAR)		1.1	Autism Spectrum Disorders (ASD)	1
 1.2.1 Socially Assistive Robots for Autism Therapy		1.2	Socially Assistive Robotics (SAR)	5
 1.3 Music Therapy in ASD Treatment 1.4 Contributions 1.5 Organization 2 Related Works in Autism and Robots 2.1 Autism 2.1 Autism 2.1.1 Motor Control 2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.2 Human Robot Interaction in Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic Robots Designs for Autism 2.2.3 Using NAO in Autism 2.3 Music Therapy 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.2 Classification Applications 3.2 Data Acquisition 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.1 Determination of Mother Wavelet 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.4.1 Mathematication Result 3.5 Summary 			1.2.1 Socially Assistive Robots for Autism Therapy	7
1.4 Contributions 1.5 Organization 2 Related Works in Autism and Robots 2.1 Autism 2.1.1 Motor Control 2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.1.4 Interactive and Therapeutic Robots Designs for Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.2 Data Acquisition 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.4.2 Classification Result		1.3	Music Therapy in ASD Treatment	10
1.5 Organization 2 Related Works in Autism and Robots 2.1 Autism 2.1.1 Motor Control 2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 2.3 Music Therapy in robot 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.2 Data Acquisition 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result <t< th=""><th></th><th>1.4</th><th>Contributions</th><th>12</th></t<>		1.4	Contributions	12
2 Related Works in Autism and Robots 2.1 Autism 2.1.1 Motor Control 2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.1.4 Human Robot Interaction in Autism 2.2 Human Robot Interaction in Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 2.4 Summary 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.1.2 Classification Method 3.3 Support Vector Machine 3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.5 Summary		1.5	Organization	13
 2.1 Autism	2	Rela	ited Works in Autism and Robots	15
2.1.1 Motor Control 2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.1.3 Music Therapy 2.2 Human Robot Interaction in Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.2 Data Acquisition 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.5 Summary		2.1	Autism	15
2.1.2 Turn-Taking Skills 2.1.3 Music Therapy 2.1.3 Music Therapy 2.2 Human Robot Interaction in Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 2.4 Summary 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.1.2 Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.4.3 Summary			2.1.1 Motor Control	18
2.1.3 Music Therapy 2.2 Human Robot Interaction in Autism 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.5 Summary 3.5 Summary 3.6 Hardware Selection and Design 4 Xylo-Bot: An Interactive Music Teaching System 4.1 NAO: A Humanoid Robot			2.1.2 Turn-Taking Skills	20
 2.2 Human Robot Interaction in Autism			2.1.3 Music Therapy	23
 2.2.1 Interactive and Therapeutic Robots Designs for Autism 2.2.2 Different Therapeutic approaches for Individuals with ASD		2.2	Human Robot Interaction in Autism	24
 2.2.2 Different Therapeutic approaches for Individuals with ASD 2.3 Using NAO in Autism			2.2.1 Interactive and Therapeutic Robots Designs for Autism	25
 2.2.3 Using NAO in Autism 2.3 Music Therapy in robot 2.4 Summary 3 Pre-Study: A Wavelet-based Approach for Emotion Classification 3.1 Related Work 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4 Experimental Result 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.5 Summary 4 Xylo-Bot: An Interactive Music Teaching System 4.1 Hardware Selection and Design 4.14 NAO: A Humanoid Robot 			2.2.2 Different Therapeutic approaches for Individuals with ASD	28
 2.3 Music Therapy in robot			2.2.3 Using NAO in Autism	32
 2.4 Summary		2.3	Music Therapy in robot	34
 3 Pre-Study: A Wavelet-based Approach for Emotion Classification Related Work Related Work Electrodermal activity (EDA) Classification Applications 3.1 Electrodermal activity (EDA) Classification Applications Data Acquisition Proposed Classification Method Continuous Wavelet Transform Continuous Wavelet Transform Continuous Wavelet Transform Support Vector Machine 3.4 Experimental Result All Determination of Mother Wavelet Classification Result Summary 4 Xylo-Bot: An Interactive Music Teaching System Hardware Selection and Design NAO: A Humanoid Robot 		2.4	Summary	35
 3.1 Related Work	3	Pre-	Study: A Wavelet-based Approach for Emotion Classification	37
 3.1.1 Electrodermal activity (EDA) 3.1.2 Classification Applications 3.2 Data Acquisition 3.3 Proposed Classification Method 3.3.1 Continuous Wavelet Transform 3.3.2 Wavelet-Based Feature Extraction 3.3.3 Support Vector Machine 3.4 Experimental Result 3.4.1 Determination of Mother Wavelet 3.4.2 Classification Result 3.5 Summary 4 Xylo-Bot: An Interactive Music Teaching System 4.1 Hardware Selection and Design 4.1 NAQ: A Humanoid Robot 		3.1	Related Work	38
 3.1.2 Classification Applications			3.1.1 Electrodermal activity (EDA)	38
 3.2 Data Acquisition			3.1.2 Classification Applications	39
 3.3 Proposed Classification Method		3.2	Data Acquisition	41
 3.3.1 Continuous Wavelet Transform		3.3	Proposed Classification Method	43
 3.3.2 Wavelet-Based Feature Extraction			3.3.1 Continuous Wavelet Transform	43
 3.3.3 Support Vector Machine			3.3.2 Wavelet-Based Feature Extraction	46
 3.4 Experimental Result			3.3.3 Support Vector Machine	50
 3.4.1 Determination of Mother Wavelet		3.4	Experimental Result	52
 3.4.2 Classification Result			3.4.1 Determination of Mother Wavelet	53
 3.5 Summary			3.4.2 Classification Result	54
 4 Xylo-Bot: An Interactive Music Teaching System 4.1 Hardware Selection and Design		3.5	Summary	59
4.1 Hardware Selection and Design	4	Xylo	o-Bot: An Interactive Music Teaching System	60
4.1.1 NAO: A Humanoid Robot		4.1	Hardware Selection and Design	61
			4.1.1 NAO: A Humanoid Robot	61
4.1.2 Accessories			4.1.2 Accessories	63

	4.2	Experimental Sessions Designs
		4.2.1 Experiment Room
		4.2.2 Q-Sensor
		4.2.3 Participant Selection
		4.2.4 Session Detail
	4.3	Module-Based Acoustic Music Interactive System Design
		4.3.1 Module 1: Eye-hand Self-Calibration Micro-Adjustment 75
		4.3.2 Module 2: Joint Trajectory Generator
		4.3.3 Module 3: Real-Time Performance Scoring Feedback
	4.4	Summary
5	Res	ults of Social Behavior Intervention and Emotion Classification 84
U	5 1	Social Behavior Result 85
	0.1	5.1.1 Motor Control 85
		5.1.2 Turn-Taking Behavior 87
	52	Music Emotion Classification Result
	53	Summary 94
	5.5	Summary
6	Х-Е	lophone, A New Instrument97
	6.1	Xylophone Modification
		6.1.1 Components Selection
		6.1.2 ChucK: An On-the-fly Audio Programming Language Based on C++106
	6.2	Hardware and Software Design
		6.2.1 Circuit Design
		6.2.2 Software Design
7	Disc	cussion, Conclusion and Future Work 120
	7.1	Each Participant Has A Story
		7.1.1 Subject 101
		7.1.2 Subject 102
		7.1.3 Subject 103
		7.1.4 Subject 104
		7.1.5 Subject 105
		7.1.6 Subject 106
		7.1.7 Subject 107
		7.1.8 Subject 108
		7.1.9 Subject 109
	7.2	Discussion and Conclusion
	7.3	Future Work: New Style Session Proposal
	. –	

Bibliography

List of Tables

3.1	Comparison of different wavelet functions on the feature extraction and emotion classification performance (%) of 2 and 3 classes using SVM clas- sifier with different kernels. The abbreviations "Acc", "Bor", and "Joy"	
	respectively stand for the emotions "Acceptance", "Boredom", and "Joy"	54
3.2	Classification results with SVM and KNN using the presented wavelet-	
	the raw EDA data, and the raw EDA data + statistical features	58
4.1	Session Details: Baseline session, Intervention sessions, and Exit session. Different activities also	
	included in this table.	73
4.2	Continued Table for Session Details.	74
5.1	Emotion change in different events using wavelet-based feature extraction	
	under SVM classifier	91
5.2	Emotion change classification performance in single event with segmenta-	
	tion using both SVM and KNN classifier.	93
5.3	classification rate in kids learn, kids play and robot feedback across warm	
	up (S1) and music practice (S2) sections	94
5.4	TD vs ASD Emotion Changes from Baseline and Exit Sessions	94

List of Figures

2.1	Chronological age-matched, ASD and Mental age-matched brain activities in forward speech.	17
3.1	Two samples of the annotation process, left shows some video frames asso- ciated 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.	43
3.2	The distribution of the emotions across all subjects and events. The ab- breviations "ACC", "BOR", and "JOY" respectively correspond to the emo- tions "Acceptance", "Boredom", and "Joy"	44
3.3	The CWT of a typical EDA signal using the C-Morlet mother wavelet. Dif- ferent scales of the wavelet functions are convolved with the original EDA signal to highlight different features of the raw data. As can be seen in- side 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	
3 /	scaling decreases.	48
5.4	the real and imaginary wavelet coefficients are calculated at different scales. Then the amplitude of these coefficients is calculated to provide the corre-	
3.5	sponding spectrogram. This spectrogram is then used as the feature space Canonical SVM for classifying two linearly separable classes. The decision boundary is shown by OH. Two hyperplanes H_1 and H_2 pass the support	49
3.6	vectors that are circled inside the figure	51 53
4.1	A Humanoid Robot NAO: 25 Degrees of Freedom, 2 HD Cameras and 4 Microphones	62
12	Microphones	64
4.2 4 3	Actual Xylophone and Mallets from $N\Delta\Omega$'s Rottom Camera	65
ч.5 4 4	Mallet Griner	66
4.5	Instrument Stand: (a) Left View (b) Top View (c) Front View	67
4.6	Schematic robot-based therapy session and video capturing setting	68
4.7	Experiment Room	69

4.8 4.9	Block Diagram of Module-Based Acoustic Music Interactive System Color Detection From NAO's Bottom Camera: (a) Single Blue Color De-	75
	tection (b) Full Instrument Color Detection (c) Color Based Edge Detection.	77
4.10	Melody Detection with Short Time Fourier Transform	79
4.11	An Example of Compute Levenshtein Distance for "Sunday" and "Saturday"	81
5.1	Motor Control Accuracy Result	86
5.2 5.3	Main Music Teaching Performance Accuracy	86
5.4	<i>The distribution of the targeted emotions across all subjects and events.</i>	88 90
6.1	Piezo Sensor Schematic	99
6.2	Piezo Sensor VS A Quarter	100
6.3	Schematic of Op-Amp MCP 6002	101
6.4	Schematic of Piezo Sensor Application as A Switch	103
6.5	SparkFun Multiplexer Breakout 8 Channel (74HC4051)	104
6.6	Arduino Uno Microprocessor	105
6.7	The schematic design for the circuit.	108
6.8	Circuit board in real size.	109
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	110
6.10	Arduino code for pin assignment of mutiplexers.	111
6.11	Loop in checking all pins for getting signals from the instrument.	112
6.12	Selecting proper mux by decoding the signals.	113
6.13	Control the input level in order to filter out small viberation noise to make	
	sure getting meaningful input.	114
6.14	A Flow Chart of Using ChucK in Designing Sound Control System to X-	
	Elophone	115
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.	116
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. $\ . \ .$	117
6.17	Chuck Code Part III: Using 1 and 0 to change instrument sound for broad-	110
6.18	casting	118
	'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	119
7.1	Subject 101	123

7.2	Subject 102																		1	23
7.3	Subject 103																		1	24
7.4	Subject 104																		1	25
7.5	Subject 105																		1	27
7.6	Subject 106																		1	27
7.7	Subject 107												•						1	30
7.8	Subject 108																		1	32
7.9	Subject 109				•														1	33

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) Kanner (1944); Wing (1997). 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 Lord et al. (2000). Children with ASD could also experience deficits in inappropriate verbal and nonverbal communication skills, including motor control, emotional facial expressions, and eye gaze attention Dautenhahn et al. (2002). 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 Maenner et al. (2020).

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. Council et al. (2001) 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. de Boer-Ott et al. (2004)

Some of the treatments have been proven that has significant and convincing support for ASD children, such as Applied Behavior Analysis (ABA) Cooper et al. (2007), Discrete Trial Teaching (DTT) Sarokoff and Sturmey (2004), and Pivotal Response Training (PRT) Pierce and Schreibman (1995). Currently, ABA Michaud and Caron (2002); Michaud and Théberge-Turmel (2002) has focused on teaching individuals with ASD appropriate social skills in an effort to make them more successful in social situations Wolff and Chess (1964). 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 Mirenda (2001), Joint Action Routines (JARS) Drew et al. (2002), Cognitive Behavioral Modification Epstein and Baucom (1989), Structured Teaching Schopler et al. (1995), and Social Stories Feng et al. (2013); Gray and Garand (1993); Mavadati et al. (2014), 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. de Boer-Ott et al. (2004)

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 Fong et al. (2003). 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. Zane et al. (2008) 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 Cassell and Jenkins (2000); Druin et al. (2000); Laurel (2013); Tapscott (1998). 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 Dautenhahn (1998), 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 Werry and Dautenhahn (1999).

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 Feil-Seifer and Mataric (2005). SAR includes all the characteristics of the SIR mentioned in it Fong et al. (2003), 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) Feil-Seifer and Mataric (2005). Companion robots Wada et al. (2002) 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 Edwards and Beck (2002). Service social robots are able to accomplish a variety of tasks for individuals with physical impairments Huttenrauch and Eklundh (2002). 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 Dautenhahn et al. (2002); Michaud and Caron (2002); Michaud and Théberge-Turmel (2002).

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. Feil-Seifer and Mataric (2005). 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 informa-

tion obtained from two cameras on its head. Besides, its "eyes" will determine the distance between objects and robots. Obringer and Jonathan (2011) 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 Association (2000). 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 Wolff and Chess (1964). 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 Feil-Seifer and Mataric (2005); Fong et al. (2003); Pierno et al. (2008); Villano et al. (2011). 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., Fong et al. (2003)), and to utilize robotic systems to develop novel interventions and enhance existing treatments for children with ASD Association (2000); Obringer and Jonathan (2011); Robins et al. (2006a). Mazzei et al. Dautenhahn and Werry (2004), for example, designed the robot "FACE" to show the details realistically of human facial expressions. Reviewing the literature in SAR Feil-Seifer and Mataric (2005); Fong et al. (2003) 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 Churchill and Bryson (1972), haptic interaction Hutt and Vaizey (1966), and adaptive game interactions based on affective cues inferred from physiological signals Sorosky et al. (1968) 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 Kanner et al. (1943) 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" Reschke-Hernández (2011). 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 Thaut and Clair (2000). 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 Dellatan (2003); Starr and Zenker (1998). 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 Brownell (2002); Katagiri (2009); Kern and Aldridge (2006).

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 Costa et al. (2013); Feng et al. (2013); Robins et al. (2012); Wainer et al. (2010). 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 Gifford et al. (2011). 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 Warwick and Alvin (1991).

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.

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 Lotter (1966). 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)Lotter (1966). 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 communi-

cating 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 Ramachandran and Oberman (2006) 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 Ramachandran and Oberman (2006). There are several studies that focus on the neurological deficits of individuals with autism and studying their brain activities. Figure 2.1 Redcay and Courchesne (2008) demonstrates brain activity difference between groups in forward speech.

Individuals with autism might also have several other unusual social developmental behaviors that may appear in infancy or childhood. For instance, children with autism show less attention to social stimuli (e.g., facial expressions, joint attention), and respond less when calling their names. Compared with typically developing children, older children or adults with autism can read facial expressions less effectively and recognize emotions behind specific facial expressions or the tone of voice with difficulties Popper (2005). 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 Bartak et al. (1975).



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

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 Gowen and Hamilton (2013). Recent studies show that individuals with autism can be observed with abnormal motor skills in the early age stage Brian et al. (2008); Provost et al. (2007b); Teitelbaum et al. (1998). This deficit sticks with them throughout childhood and even in their adulthood as well as Fournier et al. (2010); Ming et al. (2007); Van Waelvelde et al. (2010). It has been reported that the prevalence of motor skill deficits is between 21 and 100 % Ghaziuddin et al. (1994); Manjiviona and Prior (1995); Miyahara et al. (1997), 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 Jasmin et al. (2009), 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 Sutera et al. (2007). 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 Gowen and Hamilton (2013).

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 Wolff and Chess (1964). 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 Pierno et al. (2008); Tang et al. (2011). 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 Villano et al. (2011). Research areas related to motor control include motor synchronization, motor learning, signal processing and the theory of perceptual function.

Hippotherapy (HPOT) Ajzenman et al. (2013) 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 Engel and MacKinnon (2007). Also, active engagement helps to improve in adaptation and increased willingness to participate in daily activities after the therapy sessions Brown and Dunn (2010), and similar effectiveness in using the horse's movements in the HPOT treatment tool Ajzenman et al. (2013). In Ajzenman et al. (2013), 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 Ajzenman

et al. (2013). Gross motor (GM) and fine motor (FM) development in children with ASD has also been studied in Provost et al. (2007a). 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 Provost et al. (2007a).

2.1.2 Turn-Taking Skills

Social communication can be initiated by typically developing kids in their infant stage Neel et al. (1990). 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 Brazelton et al. (1974); Reddy (1991); Trevarthen (1978). However, these skills can be impossible for children with autism, and can be noticed at 7 to 9 months old, according to Lord (1984); Mundy et al. (1986). 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 Kaczmarek (2002). Moreover, these communicative behaviors have been linked to important developmental outcomes in children with ASD McDuffie (2004); Sigman et al. (1999); Stone and Yoder (2001). 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 Aldred et al. (2004); Mundy and Crowson (1997). 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 Edition et al. (2013); Kaczmarek (2002).

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 Wolff and Chess (1964). 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 Pierno et al. (2008); Tang et al. (2011). Conversation turns are a desirable way of engaging in social life in many ways, and thus subject to rivalry Villano et al. (2011). Turntaking 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 Feil-Seifer and Mataric (2005); Fong et al. (2003). 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 Diamond et al. (2008); Guralnick and Neville (1997); Harrist and Waugh (2002); Rieth et al. (2014). 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 Brok and Barakova (2010); Diehl et al. (2012); Rieth et al. (2014); Scassellati et al. (2012).

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 Yoder and Warren (2002). 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 Yoder and Stone (2006). The effectiveness of LEGO© therapy has been examined by a group of researchers recently LeGoff (2004); Legoff and Sherman (2006); Owens et al. (2008). 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 Kim and Clarke (2015). Another research has found in examined the turn-taking behaviors in young children with autism is in Rieth et al. (2014), 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 Boso et al. (2007); Roper (2003). Nowadays, at least 12% of all treatment of individuals with autism consists of music-based therapies Bhat and Srinivasan (2013). 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 Lim and Draper (2011). Recorded music or human played back music are used in single and multiple subjects' intervention sessions from many studies Bhat and Srinivasan (2013); Corbett et al. (2008). Different social skills are targeted and reported (i.e., eye-gaze attention, joint attention and turn-taking activities) in using music-based therapy sessions Kim et al. (2008); Stephens (2008). Noted that improving gross and fine motor skills for ASD through music interventions is a missing part of this field of studies Bhat and Srinivasan (2013).

Early affective activity evolves into relationships in the regular boy, where games dominate Reddy et al. (1997). 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' Jordan and Libby (2011). 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 Wimpory and Nash (1999). 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 Burford (1988). 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 Bruner (1985) 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 Wimpory and Nash (1999).

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 Ricks and Colton (2010). However, only a few groups used humanoid robots for teaching or practicing social communication skills Dautenhahn and Werry (2004); Feil-Seifer and Mataric' (2008); Kozima et al. (2005); Pioggia et al. (2005); Robins et al. (2005, 2006b).

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 Dautenhahn and Werry (2000). 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 Werry et al. (2001). Although a recent case study Ricks and Colton (2010), 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 Ricks and Colton (2010). 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 Robins et al. (2005), while others have created robots with very mechanical forms Dautenhahn and Werry (2004), and others have developed robots with a cartoonish or animal form Kozima et al. (2005). 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 Ricks and Colton (2010), 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 Feil-Seifer and Mataric' (2008). Another non-humanoid robot used by researchers from the University of Hertfordshire called Labo-1 Dautenhahn and Werry (2004), 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 Robins et al. (2005). 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 Robins et al. (2006b). 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 Ricks and Colton (2010). A small soft snowman-shaped robot, called Keepon, was designed to represent as a simple, repeatable, mechanical robot regarding the reason mentioned above Kozima et al. (2005). 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 Pioggia et al. (2005). (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 Feil-Seifer and Mataric' (2008).

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 Dautenhahn and Werry (2004); Feil-Seifer and Mataric' (2008). 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 Dautenhahn and Werry (2004).

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 Dautenhahn and Werry (2004). 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 preprogrammed 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 Robins et al. (2006b). 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 Pioggia et al. (2005).

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, and hear orientations, It has been used for different human-robot interactions. 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 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 Shamsuddin et al. (2012a). 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 Shamsuddin et al. (2012c). 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 Shamsuddin et al. (2012b). 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 Shamsuddin et al. (2013). 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 Feng et al. (2013); Mavadati et al. (2014); Mihalache et al. (2020), 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 Peng et al. (2014); Taheri et al. (2015, 2016). 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 Molnar-Szakacs and Heaton (2012). Although limited research has found in such areas, especially using bio-signals for emotion recognition for ASD and TD kids Feng et al. (2018) 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 Feng et al. (2018), 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 that will be used in the music teaching platform. The purpose of this pre-study is to gain a better understanding of young children's emotional changes and discover an automated method for using electrodermal activity (EDA) signals to classify emotions in children. The dataset used in this paper includes a set of multimodal recordings of social and communicative behavior as well as EDA recordings of 100 children younger than 30 months old. The dataset is annotated by two experts to extract the time sequence corresponding to three main emotions including "Joy", "Boredom", and "Acceptance". 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. Another purpose of this pre-study is to find a method for comparing the differences between groups of non-autistic and autistic children regarding the emotional changes that occur with music social stimuli.

3.1 Related Work

3.1.1 Electrodermal activity (EDA)

Emotions are intense mental experiences often manifested by rapid heartbeat, breathing, sweating, and facial expressions. Using physiological signals to recognize emotions is a challenging problem that needs to be solved because doing so could lead to interesting applications such as developing wearable assistive devices and smart human-computer interfaces. This paper presents an automated method for using electrodermal activity (EDA) signals to classify emotions in children. The time-frequency analysis of the acquired raw EDA signals provides a feature space based on different emotions that can be recognized. Based on this, the complex Morlet (C-Morlet) wavelet function was applied to 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 signal recordings of 100 children younger than 30 months old. The dataset was annotated by two experts to extract the time sequence corresponding to three main emotions including "Joy," "Boredom," and "Acceptance." The annotation process was performed considering the synchronicity between children's facial expressions and EDA signal time sequences. Various experiments were conducted on the annotated EDA signals to classify emotions using a support vector machine (SVM) classifier. The quantitative results show that, compared to other methods, emotion classification performance remarkably improves when the proposed wavelet-based features are used.

For years, EDA signals have been used as an effective and reproducible electrophysiological method for investigating the sympathetic nervous system function Kwon et al. (2016); Shahani et al. (1984); Stagg et al. (2013); Tarvainen et al. (2000). Having said this, sympathetic nervous bursts change skin conductance, a characteristic that can be traced by analyzing EDA signals Kylliäinen and Hietanen (2006); Lidberg and Wallin (1981); Rehg et al. (2013a). The Q-sensor is a convenient, wireless EDA device that does not require cables, boxes, or skin preparation. The Q-sensor can simultaneously track three types of data including EDA, temperature, and acceleration Kappas et al. (2013). To date, no work on emotion classification using EDA signals and this dataset, collected at the Georgia Institute of Technology Rehg et al. (2013a), has been published.

EDA signals are non-stationary and noisy. In the literature, the wavelet-based analysis of EDA signals has been considered Laparra-Hernández and Poveda (2009); Swangnetr and Kaber (2013) as either a pre-processing step or a feature extraction approach for emotion classification. Swangnetr and Kaber (2013) used both a set of wavelet coefficients representing EDA signal features and heart rate signals to increase the percentage of correct classifications of emotional states and provide clearer relationships between physiological response and arousal and valence. Sharma and Kalra (2016) used a feature space based on the discrete wavelet transform (DWT) of the EDA signal to distinguish between subjects suffering from 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 reviewing methods that use various physiological responses, such as facial expressions and other types of biosignals, for recognizing human emotion. A wearable glass device was designed by Kwon et al. (2016) to measure both electrodermal activity (EDA) and photoplethysmogram (PPG) data for emotion recognition purposes. A built-in camera was also used for capturing partial facial expressions from the eye and nose area. This approach remarkably performs facial expression recognition in subject-dependent cases. However, for subject-independent cases, the device results in different accuracies across different types of emotions, which is undesirable.

In the literature, several emotion classification methods using different biosignals have been presented Legiša et al. (2013); Liu et al. (2014); Schmidt and Walach (2000); Xu et al. (2016). Due to the variety of the signals used in these methods, different approaches have been designed to comply with specific method characteristics. Analysis of variance (ANOVA) and linear regression Schmidt and Walach (2000) are the methods commonly used to extract features from biosignals and to recognize different emotional states. These methods assume a linear relationship between the recorded signals and emotional states. A fuzzy-based classification method Liu et al. (2014) has been used to transform EDA signals 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 used for emotion classification tasks based on physiological responses. Lin et al. (2007) developed a multilayer perceptron network (MLP) architecture capable of recognizing five emotions using various features from electrocardiography (ECG) and EDA signals and obtained an accurate classification performance. Nasoz et al. (2004) employed K-nearest neighborhood (KNN) and discriminant function analyses to perform an emotion classification task using different features extracted from EDA signals, body temperature, and heart rate.

Machine learning algorithms such as Support Vector Machine (SVM), linear discriminant analysis (LDA), and classification and regression tree (CART) have been employed for emotion classification purposes. Support Vector Machine is a well-known supervised learning algorithm that has extensively been used for pattern classification and regression Cortes and Vapnik (1995). The SVM classifier tends to separate datasets by drawing an optimal hyperplane between classes to create a maximum margin between them. The samples of each class located within the margin are called support vectors and play a primary role in calculating the parameters of hyperplanes between corresponding classes.

In several works, including Jang and Sohn (2014); Sano and Picard (2011), the authors combined various types of biosignals such as ECG, skin temperature (SKT), HR, and PPG for emotion classification purposes. Amershi and McLaren (2006) proposed unsupervised clustering methods for emotion recognition. The authors' 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 few works Legiša et al. (2013); Oberman and Ramachandran (2009) that have studied and compared different automated classification techniques for the emotion recognition of children using EDA signals. This motivated us to conduct this study using an existing dataset that concentrates on the emotion classification of children based on the relationship between their facial expressions and collected EDA signals.

3.2 Data Acquisition

The dataset utilized in this pre-study constitutes a collection of multimodal recordings of the social and communicative behavior of 100 participants younger than 30 months provided by the Georgia Institute of Technology Rehg et al. (2013b). All data were collected within the Child Study Lab (CSL) under a university-approved IRB protocol. The laboratory was a 300-square foot area. The temperature/humidity of the area was kept consistent for all sessions. Based on the dataset description, every session lasted 3–5 minutes, and the EDA signals (a frequency rate of 32 Hz) were collected from two Q-sensors attached to the left and right wrists. The entire experiment was video recorded. A collection of semi-structured play interactions with adults, called Multimodal 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," event 5: "tickling"). These experiments were aimed at analyzing and deciphering children's social-communicative behavior at an early age and are in keeping with the Rapid-ABC play protocol Ousley et al. (2012).

The annotation 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 in terms of the initiation and end of an emotion. Meanwhile, the corresponding EDA signal sequences were recorded to create a dataset for every perceived emotion. During the annotation, two dominant emotions were recognizable. Events 2 (with an average duration of 45 seconds) and 5 (with an average duration of 35 seconds) stimulated the "Joy" emotion and Event 3 (with an average duration of 60 seconds) stimulated the "Boredom" emotion. Concerning Event 1, "greeting," it was difficult to assign an emotion to it; but, the annotators most frequently used the emotion "Acceptance" for this event. We tended to exclude Event 4 from our experiments since the length of the event (on average, nine seconds) was extremely short compared with other events (on average, 50 seconds), and the annotators were not prepared to determine a specific emotion triggered by the event. Fig-



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.

ure 3.2 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 there is a tendency to develop emotion classification methodology that supports the time-frequency analysis of EDA signals, mostly properties of the continuous wavelet transform assumptive C-Morlet wavelet are given here. It followed by the pre-processing steps, as well as the wavelet-based feature extraction stage. Finally, we review the characteristics of the SVM, the classifier used with our approach.

3.3.1 Continuous Wavelet Transform

The EDA signal data recorded using the SC sensors are categorized as non-stationary signals Najafi and Robert (2003); Swangnetr and Kaber (2013). Hence, multiresolution



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".

analysis techniques are suitable for studying the qualitative components of non-stationary biosignals Najafi and Robert (2003). Continuous wavelet transform (CWT) is one of the strongest and most widely used analytical tools for multi-resolution analysis. CWT has received considerable attention regarding processing signals with non-stationary spectra Mallat (1989); Vetterli and Herley (1992); therefore, it is utilized here to perform the time-frequency analysis of the EDA signals. Contrasting many existing methods that utilize wavelet coefficients of raw signals to extract features, our proposed method is based on the spectrogram of the original data in a specific frequency range (0.5, 50)Hz, which provides more information for other post-processing steps (i.e., feature extraction and classification). We applied 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) which can also be used to calculate the spectrogram of the raw signal).

Also, opposed to many related studies that utilize real-valued wavelet functions for feature extraction purposes, we have employed the C-Morlet function since it takes into account both the real and imaginary components of the raw signal, leading to more detailed feature extraction.

The wavelet transform of a 1-D signal provides a decomposition of the time-domain sequence at different scales that are inversely related to their frequency contents Godfrey and Ólaighin (2009); Mallat (1989). This requires the time-domain signal under investigation to be convolved with a time-domain function known as a "mother wavelet." The CWT applies the wavelet function at different scales with a continuous time-shift of the mother wavelet over the input signal. Consequently, it helps represent the EDA signals at different levels of resolution resulting in large coefficients in the transform domain when the wavelet function matches the input signal which provides a multi-scale representation of the EDA signal.

Using a finite energy function $\Psi(t)$ concentrated in the time domain, the CWT of a signal x(t) is given by X(α ,b) as follows Vetterli and Herley (1992):

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

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, $\Psi(\alpha, b)$ is the scaled and translated version of the corresponding mother wavelet. "*" is the conjugation operator. Note that the wavelet coefficients obtained from Eq. (1) essentially evaluate the correlation between the signal x(t) and the wavelet function used at different translations and scales. This implies that the wavelet coefficients calculated over a range of scales and translations can be combined to reconstruct the original signal as follows:

$$x(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} X(a,b) \Psi(\frac{t-b}{a}) dadb$$
(2)

3.3.2 Wavelet-Based Feature Extraction

The time-frequency analysis of varied biosignals has been addressed in a significant amount of literature Golshan et al. (2016, 2017, 2018); Li et al. (2008). Studies have shown that the wavelet-domain feature area will improve the classification performance of various activities when using signals emanating from body responses. Therefore, it primarily enhances the classification performance because of the additional eminence area provided.

In this pre-study, we specialized in the time-frequency analysis of the EDA signal to produce a new feature area that supports the emotion classification task. As opposed to some studies that use raw time-domain signals for classification purposes Greco and Scilingo (2017); Jang and Sohn (2012), we used the amplitude of the CWT of the EDA signals to get options and drive the classifier. Operating within the wavelet-domain is advantageous since the wavelet remodel probes the given signal at completely different scales, extracting an immense amount of information for alternative post-processing steps. Additionally, the localized support of the wavelet function permits CWT-based analysis to match the native variations of the input time sequence Vetterli and Herley (1992). As a result, an elaborate representation of the signal is provided rather than the raw time-domain signal.

Figure 3.3 shows the amplitude of the CWT of a sample EDA signal at different scales using a C-Morlet wavelet function. Different scales of the wavelet function are convolved with the first EDA signal to spotlight completely different data options. 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 information 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 Golshan et al. (2016). 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)









3.3.3 Support Vector Machine

The SVM classifier tends to separate data $D = \{x_i, y_i\}_{i=1}^N, x_i \in \mathbb{R} d, y_i \in \{-1, +1\}$ by drawing an optimal hyperplane w, x + b = 0 between classes such that the margin between them becomes maximum Cortes and Vapnik (1995). With reference to Figure 3.5, The decision boundary is shown by OH. Two hyperplanes H_1 and H_2 pass the support vectors that are circled inside the figure. H_1 and H_2 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 H_1 and H_2 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:

$$\min \frac{1}{2} ||w||^2 s.t. y_i (< w, x_i > +b) \ge 1 \tag{4}$$

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

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

$$\min \frac{1}{2} ||w||^2 + C(\sum_{i=1}^N \xi_i) s.t. y_i (< w, x_i > +b) \ge 1 - \xi_i$$
(5)



Figure 3.5: Canonical SVM for classifying two linearly separable classes. The decision boundary is shown by OH. Two hyperplanes H_1 and H_2 pass the support vectors that are circled inside the figure.

Usually, various kernel functions are used to deal with the non-linearly separable data. As a result, the original data x_i 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 $k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$. 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 $\sum_{n=1}^{\infty} = \lambda_n \varphi_n(x_i) \varphi_n(x_j)$, where $\lambda_n > 0$ are called eigenvalues and functions φ_n are called eigen vectors in the expansion. The fact that all the eigenvalues are non-negative means that the kernel is positive semi-definite Cortes and Vapnik (1995). To maximize the margin, H_1 and H_2 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:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j k(x_i, x_j)$$

$$s.t.0 \le \alpha_i \le C, \sum_{i=1}^{N} \alpha_i y_i = 0, i = 1, ..., N$$
(6)

where, α_i s 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 $w = \sum_{i=1}^{N} \alpha_i y_i x_i$. Therefore, the label of the test sample (y_z) is given by:

$$y_z = sgn(\sum_{i=1}^N \alpha_i y_i k(x_i, z) + b) \tag{7}$$

3.4 Experimental Result

This work employed the EDA signals of 64 subjects. The signals were annotated based on participant facial expressions to evaluate the accuracy of the proposed wavelet-based feature extraction method on emotion classification performance. The EDA dataset is classified based on the different emotions perceived in the annotation step: Joy, Boredom, and Acceptance. The SVM classifier was 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) $k(x, y) = exp(\gamma \parallel x - y \parallel^2)$, where x and y are two feature vectors, and γ , c, and d are constant values.



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.

Before proceeding with the quantitative evaluation of the performance of the emotion classification method, the impact of various families of wavelet functions on the feature extraction stage, as well as emotion classification ability, had to be tested.

3.4.1 Determination of Mother Wavelet

Table 3.1 shows the classification results given by different wavelet functions. For the sake of brevity, only the results of the "db1," "coif1," "sym2," and "C-Morlet" wavelets and three kernels with SVM classifiers are shown. As typically seen, the time-frequency features calculated by the C-Morlet end up in the following classification performance, possibly due to the distinctive feature space provided by this classification performance. Figure 3.6 shows the difference between the wavelet functions. Note that the C-Morlet wavelet has successfully been applied to different types of biosignals (e.g., EEG, LFP brain signals) and has led to promising results, specifically for feature extraction functions. One of the foremost characteristics of this wavelet function is its sophisticated nature of primarily extracting various features from input time sequences.

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		61	56	61	75
ACC-JOY	Lincor	50	46	50	69
BOR-JOY	Lilleal	51	69	57	90
BOR-JOY-ACC		51	35	39	66
ACC-BOR		51	56	58	64
ACC-JOY	Polynomial	54	54	57	81
BOR-JOY	Forynonnai	55	64	69	86
BOR-JOY-ACC		43	46	50	61
ACC-BOR		59	60	58	74
ACC-JOY	DDE	55	44	57	84
BOR-JOY	KDI	68	51	69	89
BOR-JOY-ACC		45	35	50	69

3.4.2 Classification Result

The pre-processing stages performed on the raw EDA dataset are explained at the beginning of this section. Afterward, the recognition results with various modalities under SVM and KNN classifiers are presented. Classification performances of the suggested wavelet-based feature extraction method used on the raw EDA signals are compared here. Plus, statistical feature extraction methods Liu et al. (2016); Mera and Ichimura (2004) used for EDA signal performance are compared to the proposed feature extraction method. The extracted features for the proposed method are primarily based on statistical moments of acquired EDA time sequences 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 was applied to the segments of the EDA signals obtained from the annotation step in order to smooth the signal, eliminating some existing impulsive noise that may happen due to the sudden movement of subjects during the experiment. Next, the amplitude of the wavelet coefficients was calculated for the frequency range of (0.5, 50)Hz. The reason for using such a wide frequency range was to ensure that all detailed components of the EDA signal were considered (See Figure 3.3).

Principal component analysis (PCA) Abdi and Williams (2010) was then applied to 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 that is extensively used for data analysis before classification. PCA can decrease the chance of overfitting which can happen because of the enormous size of the feature vectors. In our experiments, 95% of the eigenvalues corresponding to maximum variance directions were kept. Since the spectrogram of the raw EDA data (see Figure 3.4) was calculated for 100 scales (e.g., frequency range (0.5, 50)Hz with a resolution of 0.5Hz), for a fair comparison, we first down-sampled 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 was applied. As a result, on average, the length of the wavelet-based feature vector, before and after PCA, was 1000 and 35 samples respectively, while, for the raw data, the lengths were 1000 and 15 samples respectively.

To generate the training and test sets for the classification step, the leave-one-out crossvalidation (LOOCV) approach was adopted. To achieve the best accuracy of classification on the validation set, the parameters of the hyperplane were fixed in terms of the SVM classifier (LibSVM library Chang and Lin (2011)). 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 classifiers with different kernel functions applied and the dataset acquired from 64 annotated subjects are shown in Table 3.2.. In terms of the binary classification cases (i.e., "Acceptance vs Boredom," "Acceptance vs Joy," and "Boredom vs Joy"), besides the classification accuracy, the precision of the quantitative measures (true positive/(true positive + false positive)), recall (true positive/(true positive + false negative)), and AUC (area under the receiver operating characteristic curve), are also given in the table. To calculate precision and recall, first, a positive class is chosen from one of the emotions and then the precision and recall values are calculated. Next, the order is changed and another emotion is used as the positive class and the precision and recall values are calculated. The final step is to calculate the average of all the precision and recall values. The result of this calculation is presented in Table 3.23.2. As shown in the 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.Compared to other feature

extraction methods, the proposed wavelet-based features lead to a higher classification performance among almost all cases (both SVM and KNN classifiers). For example, using the SVM classifier with the linear kernel SVM and raw EDA signal, the classification rate of a 3-class case is about 38%. However, the feature space introduced in this paper reaches an accuracy of 68%. From Table 3.2, note the competitive classification performance for the 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, three different values (K = 1, 3, 5) were used in this study. 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 the K = 1 and "Acceptance vs Boredom" task, it obtains an accuracy of 73% compared to the proposed method which reaches an accuracy of 70% for the same task. However, the proposed method shows better classification performance in most of the other cases. For 3-class classification, the proposed method achieves, on average, 64% accuracy for all K values, while the other two feature extraction approaches result in 57% and 44%, respectively. These results indicate the superiority of the proposed method for 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, making it unsuitable to use the raw time sequence for practical signal processing approaches. Prior studies have represented stochastic physiological signals using statistical features to classify emotional states Mera and Ichimura (2004). Unfortunately, information can be lost with such features since 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 that are not yet identified Swangnetr and Kaber (2012).

the raw EDA d	ata, and the	e raw EDA	data	+ statistic	cal feati	ures.							
	SVM	Wavelet-bé	nsed			Statistics-b	ased fea	tture + Raw	data	Raw data			
	Kernels	Accuracy	AUC	Precision	Recall	Accuracy	AUC	Precision	Recall	Accuracy	AUC	Precision	Recall
ACC-BOR		75	75	81	71	56	55	57	56	56	55	58	43
ACC-JOY		69	84	75	62	47	56	47	47	49	57	49	39
BOR-JOY	Linear	06	87	82	88	55	55	54	55	53	54	52	67
ACC-JOY-BOR		66				35				36			
ACC-BOR		64	68	70	64	74	82	83	74	70	78	79	70
ACC-JOY	Delmanial	81	83	78	81	LL	90	84	LL	09	68	72	60
BOR-JOY	Folynomial	86	87	85	86	09	<u>66</u>	59	60	57	62	57	57
ACC-JOY-BOR		61				55				46			
ACC-BOR		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	KBF	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		70	50	70	70	73	53	72	73	68	52	68	68
ACC-JOY		76	58	76	76	76	63	LL	76	62	60	62	62
BOR-JOY		80	64	76	80	57	99	57	57	56	68	56	56
ACC-JOY-BOR		60				56				43			
ACC-BOR		70	63	71	70	81	LL	79	81	71	68	75	71
ACC-JOY	с — 4	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		64	64	67	64	75	78	74	75	71	67	75	71
ACC-JOY	2 - V	LL LL	84	73	LL	LL	82	LL	LL	57	50	58	57
BOR-JOY		85	86	LL	85	61	58	63	61	46	57	46	46
ACC-JOY-BOR		99				58				43			

Table 3.2: Classification results with SVM and KNN using the presented wavelet-based feature extraction,
3.5 Summary

Three basic emotions were recognized within the annotation step: Acceptance, Joy, and Boredom. Numerous experiments were conducted on the dataset mistreatment either the raw segmented EDA signal or its corresponding time-frequency illustration as options. The quantitative results show that the emotion classification performance is remarkably improved when the planned wavelet-based options are used with the SVM classifier. Apart from the "C-Morlet," we also 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 utilizing information in real-time would be a remarkable extension of the current work. This would permit us to observe emotions and give participants feedback throughout experimental sessions. Moreover, due to the limitation of obtainable datasets, making a more comprehensive dataset is necessary for longer-term analysis. The quantitative results show that, compared to other methods, the emotion classification performance remarkably improves when the proposed wavelet-based features are used. This pre-study also provided a possibility of using a C-Morlet wavelet function as a 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 above, playing music can be an effective method for expressing emotion and for non-verbally communicating. Individuals with ASD are interested in interacting with a social robot. Therefore, implementing music therapy such as intervention sessions using a humanoid social robot is possible. A novel, interactive human-robot music teaching system design is presented in this chapter. The hardware and software design including experiment room setup, robot selection, instrument accessories, and platform design will be discussed in the following sections.

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, a proper xylophone with perfect timber had to be found. Second, we had to arrange the xylophone in a proper position in front of the robot to make it visible and reachable to play. Next, a set of challenge-based experimental sessions had to be constructed including a baseline session, intervention sessions, and an exit session with various levels of activities. Finally, the module-based in-

teractive music teaching system had to be designed and programmed so that it could be implemented into the experimental sessions.

4.1 Hardware Selection and Design

4.1.1 NAO: A Humanoid Robot

NAO, a humanoid robot, was selected for the current research which murchandised by SoftBank Group Corporation. NAO is 58 cm (23 inches) tall and has 25 degrees of freedom. Most human body movements can be performed by NAO. NAO also features an on-board multimedia system including four microphones for voice recognition and sound localization, two speakers for text-to-speech synthesis, and two HD cameras with a maximum image resolution of 1280 x 960 for high-quality online observation. As shown in Figure 4.1, these utilities are in the middle of the forehead and mouth areas of NAO. NAO's computer vision module includes facial and shapes recognition units. By using the vision feature of the robot, NAO can see an instrument with its lower camera and implement an eye-hand self-calibration system that allows the robot to micro-adjustment its arm joints in real-time in case of off positioning before music play.

NOA's arms have a length of approximately 31 cm. Position feedback sensors are equipped in each of the robot's joints to obtain real-time localization information from them. This helps protect people and allows robot operators to provide robot safety. Each robot arm has five degrees of freedom and is equipped with sensors to measure the position of joint movement. To determine the position of the xylophone and the mallets' heads, the robot analyzes images from the lower monocular camera located in its head, which has a



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

diagonal field of view of 73 degrees. By using these dimensions, properly sized instruments can be selected and more accessories can be built. Information regarding this will be presented in the following sections.

Four microphones embedded on NAO's head can be seen in Figure 4.2. According to the official Aldebaran documentation, these microphones have a sensitivity of 20mV/Pa +/-3dB at 1kHz and an input frequency range of 150Hz- 12kHz. Data is recorded as a 16 bit, 48000Hz, 4 channel wave file which meets the requirements for designing the online feedback audio score system that will be described in a later section.

4.1.2 Accessories

The purpose of this study is to have a toy-size humanoid robot play and teach music to children with autism. Some necessary accessories needed to be purchased and made before the robot was capable of completing 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 Glockenspiel with 11 sound bars 2 cm in width was selected and purchased. The instrument is 31 cm x 9.5 cm x 4 cm, including the resonating body. The smallest sound bar is playable in an area of 2.8 cm x 2 cm, the largest in an area of 4.8 cm x 2 cm. The instrument is diatonically tuned in C major/A minor (see Figure 4.3).The 11 bars of the xylophone represent 11 different notes, or frequencies, that cover one and a half octave scales, from C6 to F7.



Figure 4.2: Microphone locations on NAO's head

To provide a music teaching environment system for children with autism, the xylophone is one of the best choices for such a study. The xylophone, also known as the marimba or the glockenspiel, is categorized as a percussion instrument that consists of a set of metal/wooden bars that are struck with mallets to produce delicate musical tones. Along with the keyboard or drum, to play the xylophone properly, a unique technique needs to be applied. A proper strike movement is required to produce a beautiful note. The action required is perfect for practicing motor control, and the melody played by the user can support music emotion learning.

Mallet Gripper Design

For the mallets, we used the pair that came with the xylophone but added a modified 3D-printed gripper that allowed the robot hands to hold them properly. The mallets are approximately 21 cm in length and include a head with a 0.8 cm radius. Compared to other



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

designs, the mallet gripper we added encourages a natural holding position for the robot. In turn, the robot properly models how participants should hold the mallet stick (see Figure 4.4).

Instrument Stand Design

Using carefully measured dimensions, a wooden base was designed and laser cut to hold the xylophone at a proper height for the robot to play in a crouching position. In this position, the robot could easily be fixed in a location and have the same height as the participants, making it more natural for the robot to teach activities (see Figure 4.5).



Figure 4.4: Mallet Griper

4.2 Experimental Sessions Designs

4.2.1 Experiment Room

All the experiment sessions were held in an 11ft x 9.5ft x 10ft room located in the Ritchie School of Engineering Room 248, University of Denver. Six HD surveillance cameras were installed in the corners and on the sidewalls and ceiling of the experimental room (see Figure 4.6). One miniature hidden microphone was attached to the ceiling camera to sending real-time audio to the observation room so that the caregiver could listen in. During experiment sessions, an external, hand-held audio recorder was set in front of participants to collect high-quality audio to use for future processes.

As shown in Figure 4.7, the observation room was located behind a one-way mirror that was behind the participants so that they were not distracted by it during sessions. Real-time video and audio were broadcasted to the observation room during sessions, allowing





(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

researchers to observe and record throughout the sessions. Parents in the observation room could also call off sessions in the 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.



Figure 4.7: Experiment Room

4.2.3 Participant Selection

Nine ASD participants (average age: 11.73, std: 3.11) and seven TD participants (average age: 10.22, std: 2.06) were recruited for this study. All participants were selected from the potential subject pool with help from the psychology department. For each participant in the ASD group, six sessions were delivered, including a baseline session, four intervention sessions, and an exit session. As for the TD control group, only baseline and exit sessions were required for each participant. Every session lasted for a total of 30-60 minutes depending on the difficulty of each session and the performance of individuals. Typically, each participant had baseline and exit session lengths that were comparable. For intervention sessions, the duration gradually increased due to the challenge level increasing.

4.2.4 Session Detail

Baseline and exit sessions contained two activities: 1) music practice and 2) music gameplay. The goal of the participants was to complete a full song play. They started by playing a single note strike with a color hint. One-by-one, participants were asked to play multiple notes, half song play, and full song play after they perfected the previous task. After the mentioned music practice was completed, a freshly designed music game, that contained three novel entertaining game modes in it, was presented to participants. Participants could communicate with the robot regarding which mode to play with.

Mode 1): The robot randomly picks a song from its song bank and plays it for participants. After each play, participants were asked to identify a music feeling to find out whether music emotion can be recognized from an early age.

Mode 2): A sequence of melodies was randomly generated by the robot with a conso-

nance (happy or comfortable feeling) or dissonance (sad or uncomfortable feeling) style. Participants were asked to express their emotions verbally and a playback were required afterwards.

Mode 3): Participants have five seconds of free play then challenge the robot to imitate what was just played. After the robot was done playing, the participant rated the performance, providing a teaching experience for all human subjects.

There was no limit on how many trials or modes each individual could play for each session, but each mode had to be played at least once in a single session. The only difference between the baseline and exit sessions was the song that was used in them. In the baseline session, "Twinkle, Twinkle, Little Star" was used as an entry-level song for every participant. For the exit session, participants chose a customized song so that they would be more motivated to learn the music. In turn, this made the exit session more difficult than the baseline session. By using the Module-Based Acoustic Music Interactive System, inputting multiple songs became possible and less time-consuming. More than 10 songs were collected in the song bank including "Can Can" by Offenbach, "Shake It Off" by Taylor Swift, "Spongebob SquarePants" from the animated show Spongebob SquarePants, and "You Are My Sunshine" by Johnny Cash. Music styles covered kid's songs, classic, pop, ACG (Anime, Comic and Games), folk, proving such a platform to be versatile. Because of the various music styles that NAO can play, learning motivation was successfully delivered across all sessions.

Each intervention session was divided into three parts: S1) warm-up; S2) single activity practice (with a color hint); and S3) music gameplay. A user-customized song was used in every intervention session to have participants engaged in multiple repetitive activities. The purpose of having a warm-up section was to have participants practice motor control skills

while also helping them have a refreshed memory, allowing them to implement motor skills in the following activities. The single activity practice was based on music practice from the baseline/exit session. In each intervention session, the single activity practice only had one type of music practice each session. For instance, single-note play was delivered in the first intervention session. For the next intervention session, the single activity practice was multiple notes play. The level of difficulty for music play was gradually increased across sessions. This was done to make a challenge-based engagement activity for ASD groups that improved motivation and had better emotion stimuli. As for music, the gameplay remained the same as the baseline/exit session (see Table 4.1).

4.3 Module-Based Acoustic Music Interactive System Design

In this section, a novel, module-based robot-music teaching system will be presented. To be successful, several tasks need to be accomplished: a) make the robot play a sequence of notes or melody fluently; b) make the robot play notes accurately; c) make the robot adapt to multiple songs easily; d) make the robot be able to have social communication with participants; e) make the robot be able to deliver learning and teaching experiences to participants; f) make the robot have fast responses and accurate decision-making. To accomplish these tasks, a module-based acoustic music interactive system was designed. Three modules were built in this intelligent system: Module 1: eye-hand self-calibration micro-adjustment; Module 2: joint trajectory generator; and Module 3: real-time performance scoring feedback (see Figure 4.8).

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

#	Session Type	Outlines	Session Details	Purpose
-	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 Wusic Game: Music Game: Mode 1) robot play song(s) based on user's request Mode 2) 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.
7	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.

			Continued table	
			Warm-Up:	
			Same as previous session	Review previous music
		First half customized	Half Song Practice:	and social behaviors
		song practice.	Randomly select a chunk of	and skills. Implement all
			melody from first half of the	skills to a harder challenge.
Intervent	ion		song (5-7 notes), ask user to	
Sessio	ц		memorise the music and	
			playback, with color hits.	
			Music Game:	
			Same as previous session	
			Warm-Up:	
			Same as previous session	Review previous music and
		Second half customized	Half Song Practice:	social behaviors and skills.
		song practice.	Randomly select a chunk of	Implement all skills to harder
			melody from second half of the	challenge.
			song (5-7 notes), ask user to	
			memorise the music and	
			playback, with color hits.	
			Music Game:	
			Same as previous session	
			Music Practice:	
			Single note strike	
			Multiple notes strike	
ц ;		Include all music activities	Half song practice	Compare music and social
Connio	ş	in one session, play	Whole song practice	performance with baseline
0199210	=	customized song.	Music Game:	session.
			Same as previous session	
			Music Game:	
			Same as previous session	

Table 4.2: Continued Table for Session Details.



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

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

Knowledge about the parameters of the robot's kinematic model was essential for programming tasks requiring high precision such as playing the xylophone. While the kinematic structure was known because of the construction plan, errors could occur because of factors such as imperfect manufacturing. After multiple rounds of testing, it was determined that the targeted angle chain of arms never equals the returned chain. Therefore, we used a calibration method to eliminate this error.

Color-Based Object Tracking

To play the xylophone, the robot had to be able to adjust its motions according to the estimated relative position of the xylophone and the heads of the mallets it was holding. To estimate the poses adopted in this paper, we used a color-based technique.

The main idea in object tracking is that, based on the RGB color of the center blue bar, given a hypothesis about the xylophone's position, one can project the contour of the xy-lophone's model into the camera image and compare them to an observed contour. In this way, it is possible to estimate the likelihood of the position hypothesis. Using this method, the robot can track the xylophone with extremely low cost in real-time (see Figure 4.9)

4.3.2 Module 2: Joint Trajectory Generator

Our system parsed a list of hexadecimal numbers (from 1 to b) to obtain the sequence of notes to play. The system converted the notes into a joint trajectory using the striking configurations obtained from inverse kinematics as control points. The timestamps for the control points are defined by the user to meet the experiment requirement. The trajectory was then computed by the manufacturer-provided API, using Bezier curve interpolation in the joint space, and then sent to the robot controller for execution. This process allowed the robot to play in-time with songs.

4.3.3 Module 3: Real-Time Performance Scoring Feedback

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



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.

Music Detection

Music, from a science and technology perspective, is a combination of time and frequency. To make the robot detect a sequence of frequencies, we adopted the Short-time Fourier transform (STFT) for its audio feedback system. Doing so allowed the robot to be able to understand the music played by users and provide proper feedback as a music instructor.

The 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 separately compute the Fourier transform for each shorter segment. Doing so reveals the Fourier spectrum on each shorter segment. The changing spectra can then be plotted as a function of time. In the case of discrete-time, data to be transformed can be broken up into chunks of frames that usually overlap each other to reduce artifacts at boundaries. Each chunk is Fourier transformed. The complex results are then added to a matrix that records magnitude and phase for each point in time and frequency (see Figure 4.10). This can be expressed as:

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

likewise, with signal x[n] and window w[n]. In this case, m is discrete and ω is continuous, but in most 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:



Figure 4.10: Melody Detection with Short Time Fourier Transform

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

After the robot detects the notes from user input, a list of hexadecimal numbers are returned. This list is used for two purposes: 1) to compare with the target list for scoring and sending feedback to the user and 2) a new input for having robot playback in the game session.

Intelligent Scoring-Feedback System

To compare the detected notes and the target notes, we used Levenshtein distance, an algorithm that is typically used in information theory linguistics. 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 $lev_{a,b}(|a|, |b|)$ where

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

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 $lev_{a,b}(i, j)$ is the distance between the first *i* characters of *a* and the first *j* characters of *b*.

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



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

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

where, if the likelihood is more than 66% (not including single note practice since, in that case, it is only correct or incorrect), the system will consider it to be a good result. This result is then 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 about how this relates to the experiment design will be discussed in the next chapters.

4.4 Summary

In this chapter, we have discussed both the hardware and software designs for the humanoid social assistive robot, NAO, that we used in experimental music teaching and playing sessions. To create a system, we decided to use a music teaching instructor that could both teach children pure music and deliver social content simultaneously. We first chose the proper agent, a robot named NAO, that is kid-friendly and has complex social abilities. Second, based on the size of the robot, some necessary accessories were purchased and handcrafted. We determined that a toy-sized color-coded xylophone was the best musical instrument option. Based on the size and position, a wooden xylophone stand was customized and assembled. Due to the limitation of NAO's hand size, a pair of mallet grippers was 3-D printed and customized. Last, an intelligent, module-based acoustic music interactive system was fully designed from scratch to accommodate the well-designed experiment sessions. Three modules were created to allow the robot to freely play, listen, and teach music. All of these preparations allow NAO to be a great music learning and social communication companion for children.

There were three types of experiment sessions: a baseline session, intervention sessions, and an exit session. A set of music teaching activities gradually increased in difficulty in order to deliver better music content to the participants. Both the ASD and TD groups experienced a game-like music challenge that led them from playing a single note to an entire song. The music game was designed to keep the session less tedious than the sessions that included learning, and it also provided an opportunity for us to learn more about the difference in learning and to teach the robot music created by the participants.

Module-based acoustic music interactive system design was challenging. Several problems needed to be solved. To improve the robot's xylophone strike accuracy, Module 1 provided an autonomous self-awareness positioning system for the robot to localize the instrument and make micro-adjustments to its arm joints, helping NAO play note bars correctly. NAO needed to be able to play multiple songs and programming specific arm movements for each song was unrealistic.

An easy music score inputting method needed to be complete before the experiment started. Module 2 allowed the robot to be able to play any customized song the user requested. This meant that for any songs that could be translated to either a C-major or A-minor key, a well-trained person could type in the playable hexadecimal score and the robot could play it within seconds.

Music teaching requires real-time feedback. Module 3 was designed to provide reallife music teaching experience to system users. Two key features were designed for this module: music detection and smart scoring feedback. Short-time Fourier transform and Levenshtein distance tools were adapted to fulfill the requirement, allowing the robot to understand music and provide users with a proper dosage of practice and oral feedback.

Chapter 5

Results of Social Behavior Intervention and Emotion Classification

This study examines several concepts including turn-taking, motor control, social engagement, and emotion fluctuation. To examine these concepts, the following questions were asked: **Turn-taking:** How well do ASD participants behave during music activities? **Motor control:** How well do ASD participants play the xylophone, in terms of volume, pitch, and accuracy, after intervention sessions (e.g., excellent multiple strikes should be recognized by STFT as a sequence of frequencies)? **Social Engagement:** How do participants engage with different difficult levels of music teaching events? **Emotion Fluctuation:** How do emotions change from one activity to another? How do emotions change during a single event? What are the different target and control groups? Answers to these questions are examined in the following sections.

5.1 Social Behavior Result

5.1.1 Motor Control

Nine ASD and seven TD participants completed the study over the course of eight months. All ASD subjects completed all six sessions (baseline, intervention, and exit) while all the TD subjects completed the required baseline and exit sessions. Conducting a Wizard of Oz experiment, a well-trained researcher was involved in the baseline and exit sessions in order for there to be high-quality observations and quality of performance evaluations. With well-designed, fully automated intervention sessions, NAO was able to initiate music teaching activities with participants.

Since the music detection method was sensitive to the audio input, a clear and longlasting sound from the xylophone was required. As seen in Figure 5.1, , it is evident that most subjects were able to strike the xylophone properly after one or two sessions. Notice that subject 101 and 102 had significant improvement during intervention sessions. Some of the subjects started at a higher accuracy rate and kept the rate above 80%. Even with ups and downs, subjects with this type of accuracy rate are considered to have consistent motor control performance. Two subjects (103 & 107) had a difficult time playing the xylophone and following turn-taking cues given by the robot. For both subjects, this affected performance in the following activities.

Figure 5.2 shows the accuracy of the first music teaching activity that was part of intervention sessions across all participants. Learning how to play one's favorite song can be a motivation that helps ASD participants understand and learn turn-taking skills. As described in the previous section, the difficulty level of this activity was designed to increase across sessions. Because of this, the accuracy of participant performance was expected to



Figure 5.1: Motor Control Accuracy Result



Figure 5.2: Main Music Teaching Performance Accuracy

decrease or remain consistent. This activity required participants to be able to concentrate and using joint attention skills during the robot teaching stage and respond properly afterward. 5 to 10 seconds was given after the robot said: "Now, you shall play right after my eye flashes." Participants also received an eye color change cue from the robot to complete a desired music-based social interaction. Different from the warmup section, notes played in the correct sequence were considered to be a good-count strike. As seen in Figure 5.2, most of the participants were able to complete single/multiple notes practice with an average of a 77.36%/69.38% accuracy rate although, even with color hints, the pitch difference of notes was still a primary challenge for them. Due to the difficulty of sessions 4 and 5, a worse performance, when compared to the previous two sessions, was acceptable. However, more than half of the participants showed a consistently high accuracy performance or an even better performance than previous sessions. Combining the report from video annotators, 6 out of 9 subjects showed stable, engaging behavior when playing music, especially after the first few sessions. Better learn-and-play turn-taking rotation was performed over time. Three subjects showed a significant increase in performance, revealing that 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



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

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 developed our emotion classification method based on the time-frequency analysis of EDA signals, the main properties of the CWT, assuming that it is a C-Morlet wavelet, is presented here. Then, the pre-processing steps, as well as the wavelet-based feature extraction scheme, are discussed. Finally, we briefly review the characteristics of the SVM, the classifier used in our approach.

EDA signals were collected in this study. By using the annotation and analysis method from our pre-study Feng et al. (2018), we were able to produce a music-event-based emotion classification result that is presented below. To determine the emotions of the ASD group, multiple comparisons were made after annotating the videos. We noticed that different activities may cause a change in emotional arousal. As mentioned above, the warmup section and single activity practice section use the same activity with different intensities levels. The gameplay section has the lowest difficulty and is purposely designed to be more relaxing.

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



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

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, 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 5.4 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.

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

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

In the first part of the analysis, EDA signals were segmented into small event-based pieces according to the number of "conversations" in each section. One "conversation" was defined by three movements: 1) robot/participant demonstrates the note(s) to play; 2) robot/participant repeats the note(s); and 3) robot/participant presents the result. Each segmentation lasts about 45 seconds. The CWT of the data, assuming the use of the C-Morlet wavelet function, was used inside a frequency range of (0.5, 50)Hz. An SVM classifier was then employed to classify "conversation" segmentation among three sections using the wavelet-based features. Table 5.1shows the classification accuracy for the SVM classifier with different kernel functions. As can be seen, emotion arousal change between warmup (S1) and music practice (S2) and S2 and music game (S3) sections can be classified using a wavelet-based feature extraction SVM classifier with an average accuracy of 76% and 70%, respectively. With the highest percentage of accuracy for S1 and S3 being 64%, fewer emotion changes between the S1 and S3 sections may be indicated.

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 prepared 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 kids learn, 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

	Seg	mentation C	omparisor	n in Single 7	lask			
		Warm up S	ection			Song Practice	e Section	
	Kernels	Accuracy	K value	Accuracy	Kernels	Accuracy	K value	Accuracy
learn vs play		52.62		54		53.79		52.41
learn vs feedback		53.38	V = 1	50.13		53.1	V = 1	51.72
play vs feedback	LIIIAI	47.5		50.38	LIIIAI	54.31		50.86
learn vs play vs feedback		35.08		36.25		35.52		36.55
learn vs play		49		50.25		53.79		50.69
learn vs feedback	Dolymomial	50.75	$\mathcal{L} = \mathcal{I}$	50.13	Dolumoniol	50.86	$\mathcal{L} = \mathcal{I}$	50.34
play vs feedback	roiyiioiiiiai	49.87		49.5	roiynoiniai	49.14		52.07
learn vs play vs feedback		33.92		35.83		34.71		35.29
learn vs play		54.38		48.37		50.86		50.17
learn vs feedback	DDE	55.75	2 - N	52.75	DDE	53.97	2 - N	50.17
play vs feedback	NDL	51.12		50	NDF	53.79		52.93
learn vs play vs feedback		36.83		34.17		34.83		33.1

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

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

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

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	50 50	81 19
	12 88	25 75	25 75

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 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 (see Table 5.4), 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. Some of the questions that we presented can be answered by them. According to the report from video annotators, most of
the participants (in both the ASD and TD groups) showed good turn-taking communication behavior in all the sessions. However, differences can also be found when comparing the groups. All TD participants were able to initiate activities from the beginning sessions, while some of the ASD participants needed some help. After several intervention sessions, most of the ASD group could perform turn-taking skills as well as the TD group. In terms of motor control skills, as can be seen in Figure 5.1, most of the ASD group mastered the skill after a few visits. For the participants who initially struggled to play the xylophone properly, improvement can be seen in Figure 5.1. Based on recorded videos and Figure 5.2, more than half of the ASD participants were well engaged during the intervention sessions. A few of them needed help from the researcher to complete tasks during the first one or two sessions. Allowing every participant to choose the music could have provided a certain level of motivation that led to them being more engaged, even with repetitive activities. From Chapter 3, we learned that emotion classification using the EDA signal could be possible. A wavelet-based feature extraction method was developed and applied in previous research with younger aged children. In this chapter, we adopted the proposed method to decode the emotional fluctuation of children with autism during music social activities. Multiple experiments were established in this chapter. Emotion change was compared across different events, within one activity, and between the target and control groups. Detailed information was shown in tables. 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 warmup 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. Further discussion of this data and a conclusion can be found in the last chapter.

Chapter 6

X-Elophone, A New Instrument

After the first phase of the experiment was done, some limitations of the current design were noticed. An acoustic instrument limits music representation, especially in regard to the rich content of some songs. Some participants may request rock or electric songs for practice, making an acoustic xylophone barely able to meet sound quality requirements. In this chapter, the design and prototyping of a novel instrument based on the xylophone are described. The purpose of this design was to increase the possibilities of playing different timbers and major/minor keys. The increased amount of possible notes would allow the system to play more customized songs for participants.

6.1 Xylophone Modification

6.1.1 Components Selection

A. Piezo Vibration Sensor: The LDT0-028K is a flexible component comprising a 28 μ m 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.

B. Op-Amp: An operational amplifier (often called op-amp or opamp) is a DCcoupled 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 tolerance in the op-amp itself.



Figure 6.1: Piezo Sensor Schematic



Figure 6.2: Piezo Sensor VS A Quarter



Figure 6.3: Schematic of Op-Amp MCP 6002

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 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 de-



SENSOR CAPACITANCE = 22.2nF CUT-OFF FREQUENCY = 0.7Hz



Figure 6.4: Schematic of Piezo Sensor Application as A Switch



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

sired 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 (I/O) pins that may be in-



Figure 6.6: Arduino Uno Microprocessor

terfaced 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. 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 timbers, and potential for Fantastical automation make it a compelling platform for expression and experimentation, but only to the extent that we can express to the computer what to do and how to do it. To this end, programming languages have 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 different ways of thinking to their users Wang et al. (2003).

6.2 Hardware and Software Design

In order to 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 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.



Figure 6.7: The schematic design for the circuit.

6.2.2 Software Design

As mentioned above, the well designed circuit has been tested and implemented. However, to make xylophone sound different, software control and design also plays important 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 reasons for selecting ChucK as our music design tool was due to its real-time sound synthesis and music creation. Figure 6.14 shows the flowchart for music software design. This design allows users to switch keys between different scales and majors/minors to create emotional music. From Figure 6.15 to Figure 6.18 shows the Chuck code in detail.



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++)</pre>
  {
   pinMode(selectPinsl[i], OUTPUT);
   pinMode(selectPins2[i], OUTPUT);
   digitalWrite(selectPinsl[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
ł
```

Figure 6.10: Arduino code for pin assignment of mutiplexers.

```
void loop() {
  // loop through all ll pins
  for (byte pin=0; pin<=7; pin++)</pre>
  {
    selectMuxPinl(pin); // Select one at a time from muxl
    int inputValuel = analogRead(Al); // read zl
    int hitBarl = checkSensor(inputValuel, pin);
    seq[pin] = hitBarl;
  }
  for (byte pin=0; pin<=7; pin++)</pre>
  {
    selectMuxPin2(pin); // Select one at a time from mux2
    int inputValue2 = analogRead(A2); // read z2
    int hitBar2 = checkSensor(inputValue2, pin);
    seq[pin+7] = hitBar2;
  1
  hasInput = checkInput(seq);
  if (hasInput == 1) {
    for (int j=0; j<16; j++)</pre>
    {
      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 selectMuxPinl(byte pin)
{
  for (int i=0; i<3; i++)
  {
    if (pin & (l<<i))
     digitalWrite(selectPinsl[i], HIGH);
    else
      digitalWrite(selectPinsl[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[])
Ł
  for(int i=0; i<16; i++)</pre>
  Ł
    if(seq[i] != 0) {
      return 1;
    ł
    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

```
1 // connect to the serial port
 2 SerialIO.list() @=> string list[];
 3 for(int i; i < list.cap(); i++)</pre>
 4 {
 5
       chout <= i <= ": " <= list[i] <= IO.newline();</pre>
 6 }
 7 // parse first argument as device number
 8 0 => int device;
 9 if(me.args()) {
     me.arg(0) => Std.atoi => device;
10
11 }
12 if(device >= list.cap())
13 {
14
      cherr <= "serial device #" <= device <= " not available\n";</pre>
15
      me.exit();
16 }
17 SerialIO cereal;
18 if(!cereal.open(device, SerialIO.B9600, SerialIO.ASCII))
19 {
      chout <= "unable to open serial device '" <= list[device] <= "'\n";</pre>
20
21
     me.exit();
22 }
```

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.

```
23 // Set up the music generators
24 // synchronize to period
25.5::second => dur T;
26 T - (now % T) => now;
27 // setup 2 instrument and differnt sound effect to them
28 BeeThree s => NRev n => Echo e => dac;
29 Mandolin ss => NRev ee => dac;
30
31.1 => n.mix;
32.2 => e.mix;
33.2 => ee.mix;
34
35 0 => int change;
360 => int cnt;
37 0 => int select;
38
39 <<<"Starting with minor lower key">>>;
40 // assign different frequencies to two xylophone arrays
41 [60, 62, 64, 65, 67, 69, 71, 72, 74, 76, 77] @=> int xylol[];
42 [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.

```
44 while (true)
45 {
46
      // wait for event
47
      cereal.onLine() => now;
48
      cereal.getLine() => string line;
      // instrument selection
49
      if (select == 0)
50
51
       Ł
52
           1 => select;
53
       }
54
      else if(select == 1)
55
       Ł
56
           0 => select;
57
           continue;
58
       }
```

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

```
59
       if(change == 1)
60
       Ł
61
           if (line$Object != null)
62
           Ł
63
               chout <= "read line: " <= line <= IO.newline();</pre>
64
               StringTokenizer tok;
               tok.set(line);
65
               Std.atoi(line) => int pos;
66
67
               chout <= "pos: " <= pos <= IO.newline();</pre>
68
               //playChord(xylo[pos], "major", T)
69
               Std.mtof(xylol[pos]) => float f;
70
               chout <= "Freq: " <= f <= IO.newline();</pre>
71
               s.freq(f); // Change sin wave frequency
72
               .5 => s.noteOn;
73
               // advance time
74
               // note: Math.randomf() returns value between 0 and 1
75
               if( Math.randomf() > .25 ) .25::T => now;
76
               else .5::T => now;
               if (pos == 10 && cnt == 1)
77
78
               £
79
                   0 => change;
80
                   0 => cnt;
81
                    <<< "Play lower key minor next round" >>>;
82
               }
83
               else
84
                Ł
85
                    1 => change;
86
                    1 => cnt;
87
               }
88
           }
89
               .02 => s.noteOn;
90
                .0 => s.noteOff;
91
               0.0 :: second => now;
92
               chout <= "play done" <= IO.newline();</pre>
93
       }
```

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

It is well known that no two autistic children experience the same challenges. This made us curious to see how different autistic children behaved in this study. In the next few pages, a brief report about each participant is presented. A 5-dimention performance eveluation system were used to compare social and music activities before and after intervention sessions. Overall performance were eveluated in following aspects: Motor Control, Engagement Level, Turn Taking Behavior, Music Performance, Music Emotion Understanding. By comparing the area covered in the pentagon, difference between baseline and exit session can be visualized easily.

7.1.1 Subject 101

As the first ASD participant to join this study, subject 101 provided excellent insight into how to improve future intervention sessions for himself and other participants. One song, "Baby Shark," fascinated him. The song was created by the South Korean education company Pinkfong and became popular in 2016. Due to the song's popularity all over the world, it was a perfect song to use in music intervention sessions for children. According to the researcher, "Baby Shark" was were the words most frequently said by subject 101. Every time the robot asked what the boy wanted to play, "Baby Shark" was the only song he mentioned. In the game session, during free play, "Baby Shark" was the only song played. To make the song more challenging for the subject, three different versions of "Baby Shark" were pre-programmed into the system. This was possible due to the simplicity of the song, which allows for three different keys to be rearranged because of the available bars on the current xylophone. Because of familiarity with the song, subject 101 provided a constant level of performance even when changing the three keys of the song.

Subject 101 is a music lover and showed a strong passion for playing the xylophone. Based on the recorded videos, it appears that motor control was appropriately taught during the first few sessions and a nice, clean note was delivered to the robot. However, sometimes subject 101 would hit the bar a bit too hard, causing damage to the instrument or the base stand. There was one time one of the base's handles broke. Fortunately, this issue did not affect the rest of the sessions. Such a high level of engagement supports, with confidence, that this platform is good to use for practice in the future. 5-dimention performance eveluation can be seen in Figure 7.1

7.1.2 Subject 102

Subject 102 made the biggest impression because he had the most significant improvement among all the sessions and in all aspects of the study. In the beginning, playing an instrument seemed difficult for him. Based on our experience with Subject 101, it was easy to determine that this participant was having a hard time striking the xylophone accurately. The hitting gesture was somehow challenging to him, and a muffled sound was continuously played even after taking part in the warmup activity. Across all sessions, breaks between activities were often requested, especially during the intervention session that had repetitive work. Most of the time, subject 102 was counting the trials during each activity to determine when he could leave the experiment room and take a break. Despite this, he never quit any of the sessions or activities, always completing the activities as needed. It was clear by subject 102's attitude during intervention sessions that he was willing to learn and improve the skill gradually. In the beginning, he seemed to not care about his performance during sessions. Towards the end, he wanted to play better and appreciated the encouragement received from the robot and the researcher. This desire to improve and appreciation of encouragement did not change the fact that he enjoyed taking breaks after each activity and counting the number of each trial. The song he used was "Twinkle, Twinkle, Little Star" because subject 102 had no favorite song. 5-dimention performance eveluation can be seen in Figure 7.2

7.1.3 Subject 103

Based on data obtained, it appears that Subject 103 did not learn much from the study. After all the sessions, he had not learned the hitting technique properly, meaning the motor skills were not successfully delivered to him. According to the videos, subject 103 showed acceptable turn-taking behavior across all sessions. However, most of the time, he needed help concentrating on tasks. There was one time that his father had to interrupt his behaviors because he was not willing to play the xylophone and instead was walking and talking about stuff.



Figure 7.1: Subject 101





Figure 7.2: Subject 102



Figure 7.3: Subject 103

Subject 103 had a lip deformity, making it very hard for the robot to understand some of his responses. For example, the robot could not recognize his "yes" response and always provided a default behavior. In most cases, a participant was required to say "yes" whenever they needed help from the robot. Since the robot could understand subject 103's response, it did not deliver the content it was supposed to. Whenever the robot asked him whether help was needed, since it could not understand the response, he always defaulted to responding to "I don't know." Because of this, the robot did not have any opportunities to teach subject 103 motor control skills. 5-dimention performance eveluation can be seen in Figure 7.3



Figure 7.4: Subject 104

7.1.4 Subject 104

He was quite, but focused at the tasks for most of the time. According to the motor control result and music performance, 70% of accuracy for playing a rock song called "I Feel Fantastic" makes his entire session perfect. At the beginning of the session, subject 104 did not had perfect strike technique for playing xylophone. However, after first intervention session, that skill was aced by him. It can be also noticed that subject 104 did provide a good music performance of all time. Emotion change was hard to detected from the videos among all session to this subject. 5-dimention performance eveluation can be seen in Figure 7.4

7.1.5 Subject 105

As mentioned before, in this system there is one activity that asks participants about their feelings regarding randomly generated music from robots. Subject 105 was the only one who told the robot his feelings. This exciting finding showed the potential for children to understand music emotions. Based on the general feedback from participants, it appears that most of them were thinking about the difficulty of doing the playback for the robot rather than about how the music made them feel. It is unclear why this occurred. One possible reason could be the difference in participants ages. Older participants may have a better understanding of the questions the robot asked as well as music emotion.

Subject 105 was smart enough to adapt to the music teaching system. He had very high-performance accuracy for all the activities. It is also worth mentioning that, at the exit session, there was a hidden challenge for all participants. Harmonics were added to the song they had practiced in the previous sessions. Subjects had the option of trying this challenge if they wanted. Subject 105 decided to participate in this challenge and was the best at this task. He played his requested song perfectly when no other participants were able to complete it. 5-dimention performance eveluation can be seen in Figure 7.5

7.1.6 Subject 106

As the only ASD girl and basketball player in this study, one word can be used to describe her: competitive. She also claimed to have musical experience playing the sax-ophone. Although the saxophone is 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, subject 106 had a difficult time striking the



Figure 7.5: Subject 105





Figure 7.6: Subject 106

xylophone correctly for the first baseline session but gradually picked up the technique from the warmup activity at her first intervention session. Once she got approved by the robot, she began to connect with the research. According to the annotators, subject 106 showed strong engagement in the sessions, focusing on decoding melodies one after another. This engagement is reflected in the results. She had an average of 80% accuracy in the main song practice activity. In this system, a color hint was given for all trials; however, the same color sometimes meant a different pitch on the xylophone. Subject 106 conquered this challenge. She would play both notes and compare them to the notes played by the robot speakers, allowing her to play perfectly. This was extremely impressive since she chose the song "Three Little Birds" composed by Bob Marley. It was clear that she was motivated by wanting to play better. This level of motivation might suggest that playing sports can affect a human's mindset, leading to a change in their behaviors in other daily life activities.

Most participants found the free play time to be the most fun. This activity allowed participants to challenge NAO to play whatever they just played for 5 seconds with no limitation on melody structure. Subject 106 spent most of the time in this section every session. This activity usually took her over 15 minutes and, most of the time, the session had to be manually ended from the computer side.

After each time the robot played music back to the participants, NAO would ask them to grade its accuracy. Most of the participants did not take this seriously and oftentimes would provide a ridiculously high or low score regardless of the robot's actual performance. In contrast, subject 106 carefully rated the robot's performance base on how she felt. According to the researcher in the room, most of the time her ratings were reasonable. 5-dimention performance eveluation can be seen in Figure 7.6

7.1.7 Subject 107

At the beginning of the session, subject 107 needed significant help from his caregiver the entire time. It was also apparent that he would need extra help in future visits. According to his mom, music therapy treatment had been given to subject 107 before, and it is was clear that he enjoyed this method of treatment.

The baseline session did not go well. Subject 107 did not listen to the robot and was not able to exhibit meaningful turn-taking behavior. With subject 107 having taken part in music therapy before, it was unclear whether he had used a xylophone before and, if so, how frequently it had been used. Despite this question, subject 107 did a good job playing the xylophone, and a nice and clear note was played most of the time.

In the first few sessions, subject 107 exhibited constant, repetitive hand gestures that made it difficult for him to hold the mallet properly. This created a delay in him responding to the robot during intervention sessions. From Figure 5.2, it is easy to see that this subject could not follow the turn-taking rule properly; most of his performance were not recorded the right way by the robot. However, according to the annotators' report, "...this participant was able to understand the color hint from the robot and sometimes provided correct input to the robot but outside of the time limit. Input outside of the time limit is not recorded in the system and could be determined to be incorrect answers by the computer...".

As sessions went on, subject 107 started to show more engagement behavior, showing the most in the last intervention session. The first verbal response to the robot's question



Figure 7.7: Subject 107

happened in session 5. Not only did the subject respond for the first time but also positively responded multiple times. One significant difference in verbal response was reflected during the music playing activity. Subject 107 started to repeat the color names while trying to strike the bars accordingly. This phenomenon may suggest that some users may need more time to adjust and get used to a new system or platform. This platform is a tool that can also be used after intervention sessions to teach and improve turn-taking behavior. 5-dimention performance eveluation can be seen in Figure 7.7

7.1.8 Subject 108

Of all the subjects, subject 108 was the most talented musician. With subject 108 being a violin player, it was assumed that they would have a precise ear playing ability.
Surprisingly, subject 108 did not perform very well. Across all intervention sessions, for customized song play, subject 108 had 68.75% accuracy. After delving into his files, it became clearer why he was obtaining less than desirable results. The biggest challenge was the song he chose: "Can Can" by Offenbach. This is an extremely difficult classical music piece to play, even on a stringed instrument, but he accepted the challenge willingly. Subject 108's performance results show a decreased rate in play accuracy after sessions. With such a high-level challenge, these results make sense.

Based on the setting of this practice, the number of notes to be played significantly increases during the last two intervention sessions for the whole song or melody to be taught to the participants. "Can Can" has a massive number of notes in it. This could be why subject 108's performance was not optimal. Having said this, having an almost 70% accuracy rate with such a problematic song is amazing.

Quiet, focused, and able to pinpoint are the best words and phrased to describe subject 108. Subject 108 was able to smoothly complete the tasks with a high level of engagement. At the very beginning of the session, he was confused about the technique used to play percussion properly since it is different from a stringed instrument. However, it did not take long for him to figure it out. Starting from the first intervention session, he had 100% accuracy on warmup tasks. 5-dimention performance eveluation can be seen in Figure 7.8

7.1.9 Subject 109

Unstoppable is a suitable word for describing subject 109. According to his caregiver, subject 109 had a hearing disability in one ear. However, this did not seem to have an

Subject 108



Figure 7.8: Subject 108

impact on his music playing performance in any of the sessions. From the beginning of the study, he exhibited hyperactive behavior with the robot in the experiment room. Of all the participants, he was the one who touched the robot the most. There was one time he almost accidentally pushed the robot backward. This meant the researcher had to pay extra attention to protect both the subject and the robot from getting injured. Impatience was also an issue for subject 109; the researcher had to spend time helping him focus on the tasks most of the time.

Although the session did not go smoothly, subject 109's music performance was acceptable. Without much help, he quickly picked the striking technique up from the robot and was able to hit notes accurately. Subject 109's hearing ability was extraordinary. He was one of the few participants who had sensitive enough hearing to be able to distinguish

Subject 109



Figure 7.9: Subject 109

different pitches created from the same-colored bars by only listen to the sound. Somehow, subject 109 was able to do this even better than subjects who claimed to have musical experience, such as subject 108. After a few sessions, subject 109 began to accept the platform and showed extended periods of concentration in the last few sessions. This was more proof that this assistive music teaching platform has potential uses for improving daily life. 5-dimention performance eveluation can be seen in Figure 7.9

7.2 Discussion and Conclusion

The results of this study indicate that the music education platform in this study can be thought of as a decent tool to facilitate the improvement of fine motor control, turn-taking skills, and social activities engagement. The automated music detection system created a self-adjusting environment for participants in early sessions. Most of the ASD participants began to develop the strike movement in the initial two intervention sessions; some even mastered the motor ability throughout the very first warmup event. The robot was able to provide verbal directions and demonstrations by participants providing voice command input whenever they need to. However, the majority of the participants did not request this feedback, instead just playing with NAO. This finding suggests that the young ASD population can learn fine motor control ability from specific, well-designed activities.

The purpose of using a music teaching scenario as the main activity in the current research was to create a natural turn-taking behavior opportunity during social interaction. Observing all experimental sessions, six out of nine subjects exhibited proper turn-taking behavior after one or two sessions. Specifically, subject 107 significantly improved in the last few sessions when comparing their results with their baseline session. Subject 109 had trouble focusing on listening to the robot most of the time. However, with the researcher prompting, he performed better at the music turn-taking activity for a short period of time. For practicing turn-taking skills, fun, motivating activities should be designed for children with autism. Music teaching that takes advantage of individuals selecting customized songs is a good example.

During the latter half of the sessions, participants started to recognize their favorite songs. Even though the difficulty for playing proper notes was much higher, over half of the participants became more engaged in the activities. Upon observations, it became clear that older participants spent more time engaging with the activities during the song practice session compared to younger participants, especially during the half/whole song play sessions. This could be for several reasons. First, the more complex the music, the more challenging it is, and the more concentration participants need to be successful. Older individuals may be more willing to accept the challenge and are better able to enjoy the sense of accomplishment they receive from their verbal feedback at the end of each session. Music knowledge base could also be another reason for this result, since older participants may have had more opportunities to learn music at school.

The game section of each session provided the highest engagement level, not only because it was relaxing and fun play, but also because it was an opportunity for the participants to challenge the robot to mirror their free play. This exciting phenomenon could be a game of "revenge." Subject 106 exhibited this behavior by spending a significant amount of time in free play game mode. According to the session executioner and video annotators, this subject 106 showed a high level of engagement for all the activities, including free play. Based on the conversation and music performance with the robot, subject 106 showed a strong interest in challenging the robot in a friendly way.

Conducting emotion studies with children with autism is difficult. Biosignals provide a possible way to do so. The event-based emotion classification method presented in this research suggests that the same activity with different intensities can cause emotion change in the arousal dimension, although, for the ASD group, it is difficult to label emotions based on facial expression changes in the video annotation phase. Fewer emotion fluctuations in a particular activity, as seen in Table 5.2, suggests that a mild, friendly, game-like teaching system may encourage better social content learning for children with autism, even when there are repetitive movements. These well-designed activities could provide a relaxed learning environment that helps participants focus on learning music content with proper communication behaviors. This may explain the improvement in music play performance during the song practice (S2) section of intervention sessions, as seen in Figure 5.2. When comparing emotion patterns from baseline and exit sessions between TD and ASD groups in Table 5.3, differences can be found. This may suggest that we have discovered a potential way of using biosignals to help diagnose autism at an early age. According to annotators and observers, TD participants showed a strong passion for this research. Excitement, stress, and disappointment were easy to recognize and label when watching the recorded videos. On the other hand, limited facial expression changes were detected in the ASD group. This makes it challenging to determine whether the ASD participants had different feelings or had the same feelings but different biosignal activity compared to the TD group. This concept should be explored in future research. Furthermore, due to the limited sample size, future research with different classification methods and a broader population should be conducted.

7.3 Future Work: New Style Session Proposal

The newly designed, X-Elophone, has two significant differences compared to the original acoustic xylophone. The most obvious improvement is the sound. Various timbers, keys, and scales can be programmed on the micro-processor board and switched in realtime, providing infinite song possibilities with a limited number of note bars. Gentle touch was also embedded in the play style. Previously, only proper motor control could create a melody. By using this new design, a fine-tuned bar softly touched also produces a clean note. These two design qualities provide unlimited possibilities for the music play feature of the music teaching platform we proposed.

Bibliography

- H. Abdi and L. J. Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010. 55
- H. F. Ajzenman, J. W. Standeven, and T. 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. 19
- C. Aldred, J. Green, and C. 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. 20
- C. 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. 41
- A. P. Association. Diagnostic and statistical manual of mental disorders: Dsm-iv. 2000. 7, 8
- L. Bartak, M. Rutter, and A. 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. 16
- A. N. Bhat and S. Srinivasan. A review of "music and movement" therapies for children with autism: embodied interventions for multisystem development. *Frontiers in integrative neuroscience*, 7:22, 2013. 23
- M. Boso, E. Emanuele, V. Minazzi, M. Abbamonte, and P. 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. 23
- T. B. Brazelton, B. Koslowski, and M. Main. The origins of reciprocity: The early motherinfant interaction. 1974. 20

- J. Brian, S. Bryson, N. Garon, W. Roberts, I. M. Smith, P. Szatmari, and L. Zwaigenbaum. Clinical assessment of autism in high-risk 18-month-olds. *Autism*, 12(5):433–456, 2008. 18
- J. C. Brok and E. 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. 21
- N. B. Brown and W. Dunn. Relationship between context and sensory processing in children with autism. *American Journal of Occupational Therapy*, 64(3):474–483, 2010.
 19
- M. 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. 11
- J. Bruner. Child's talk: Learning to use language. *Child Language Teaching and Therapy*, 1(1):111–114, 1985. 24
- B. 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. 24
- J. Cassell and H. Jenkins. *From Barbie to Mortal Kombat: gender and computer games*. MIT press, 2000. 4
- C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011. 55
- D. W. Churchill and C. Q. Bryson. Looking and approach behavior of psychotic and normal children as a function of adult attention of preoccupation. *Comprehensive Psychiatry*, 13(2):171–177, 1972. 9
- J. O. Cooper, T. E. Heron, W. L. Heward, et al. Applied behavior analysis. 2007. 2
- B. A. Corbett, K. Shickman, and E. 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. 23
- C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. 41, 50, 51

- S. Costa, H. Lehmann, B. Robins, K. Dautenhahn, and F. Soares. "where is your nose?": developing body awareness skills among children with autism using a humanoid robot. 2013. 11
- N. R. Council et al. Educating children with autism. National Academies Press, 2001. 2
- K. 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. 4
- K. Dautenhahn and I. Werry. Issues of robot-human interaction dynamics in the rehabilitation of children with autism. *Proc. From animals to animats*, 6:519–528, 2000. 25
- K. Dautenhahn and I. Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35, 2004. 8, 25, 26, 29, 30
- K. Dautenhahn, I. Werry, J. Rae, P. Dickerson, P. Stribling, and B. Ogden. Robotic playmates. In *Socially intelligent agents*, pages 117–124. Springer, 2002. 1, 6
- S. de Boer-Ott, D. E. Griswold, B. S. Myles, S. E. Byrd, J. B. Ganz, K. T. Cook, K. Otten, J. Ben-Arieh, S. A. Kline, L. G. Adams, et al. *Autism spectrum disorders: Interventions* and treatments for children and youth. Corwin press, 2004. 2, 3
- A. K. Dellatan. The use of music with chronic food refusal: A case study. *Music Therapy Perspectives*, 21(2):105–109, 2003. 11
- K. Diamond, S. Hong, and A. Baroody. Promoting young children's social competence in early childhood programs. *Social competence of young children: Risk, disability, and intervention*, pages 165–184, 2008. 21
- J. J. Diehl, L. M. Schmitt, M. Villano, and C. 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. 21
- A. Drew, G. Baird, S. Baron-Cohen, A. Cox, V. Slonims, S. Wheelwright, J. Swettenham,
 B. Berry, and T. 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. 3
- A. Druin, J. A. Hendler, and J. Hendler. *Robots for kids: exploring new technologies for learning*. Morgan Kaufmann, 2000. 4

- F. Edition et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, 2013. 21
- N. E. Edwards and A. M. Beck. Animal-assisted therapy and nutrition in alzheimer's disease. *Western journal of nursing research*, 24(6):697–712, 2002. 5
- B. T. Engel and J. R. MacKinnon. *Enhancing human occupation through hippotherapy: A guide for occupational therapy*. AOTA Press, 2007. 19
- N. Epstein and D. H. Baucom. Cognitive-behavioral marital therapy. In *Comprehensive* handbook of cognitive therapy, pages 491–513. Springer, 1989. 3
- D. Feil-Seifer and M. 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. 25, 26, 29
- D. Feil-Seifer and M. J. Mataric. Defining socially assistive robotics. In 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005., pages 465–468. IEEE, 2005. 5, 6, 8, 9, 21
- H. Feng, A. Gutierrez, J. Zhang, and M. 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. 3, 11, 34
- H. Feng, H. M. Golshan, and M. H. Mahoor. A wavelet-based approach to emotion classification using eda signals. *Expert Systems with Applications*, 112:77–86, 2018. 35, 89
- T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166, 2003. 3, 5, 8, 9, 21
- K. A. Fournier, C. J. Hass, S. K. Naik, N. Lodha, and J. H. Cauraugh. Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of autism and developmental disorders*, 40(10):1227–1240, 2010. 18
- M. 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. 18

- T. Gifford, S. Srinivasan, M. Kaur, D. Dotov, C. Wanamaker, G. Dressler, K. MARSH, and A. BHAT. Using robots to teach musical rhythms to typically developing children and children with autism. *University of Connecticut*, 2011. 11
- C. R. L. M. M. 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. 45
- H. M. Golshan, A. O. Hebb, S. J. Hanrahan, J. Nedrud, and M. 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. 46, 47
- H. M. Golshan, A. O. Hebb, S. J. Hanrahan, J. Nedrud, and M. 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. 46
- H. M. Golshan, A. O. Hebb, S. J. Hanrahan, J. Nedrud, and M. H. Mahoor. A hierarchical structure for human behavior classification using stn local field potentials. *Journal of neuroscience methods*, 293:254–263, 2018. 46
- E. Gowen and A. Hamilton. Motor abilities in autism: a review using a computational context. *Journal of autism and developmental disorders*, 43(2):323–344, 2013. 18
- C. A. Gray and J. D. Garand. Social stories: Improving responses of students with autism with accurate social information. *Focus on autistic behavior*, 8(1):1–10, 1993. 3
- V. G. C. L. Greco, A. and E. P. Scilingo. Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*, 17(3), 2017. 46
- M. J. Guralnick and B. Neville. Designing early intervention programs to promote children's social competence. *The effectiveness of early intervention*, 579:610, 1997. 21
- A. W. Harrist and R. M. Waugh. Dyadic synchrony: Its structure and function in children's development. *Developmental review*, 22(4):555–592, 2002. 21
- C. Hutt and M. J. Vaizey. Differential effects of group density on social behaviour. *Nature*, 209(5030):1371–1372, 1966. 9

- H. Huttenrauch and K. S. Eklundh. Fetch-and-carry with cero: observations from a longterm user study with a service robot. In *Proceedings*. *11th IEEE International Workshop on Robot and Human Interactive Communication*, pages 158–163. IEEE, 2002. 6
- P. B. J. K. 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. 46
- P. B. J. K. S. H. C. M. A. P. 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. 41
- E. Jasmin, M. Couture, P. McKinley, G. Reid, E. Fombonne, and E. 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. 18
- R. Jordan and S. Libby. Developing and using play in the curriculum. In *Autism and Learning (Classic Edition)*, pages 35–49. Routledge, 2011. 24
- L. A. Kaczmarek. Assessment of social-communicative competence: An interdisciplinary model. Paul H Brookes Publishing Co., 2002. 20, 21
- L. Kanner. Early infantile autism. The Journal of Pediatrics, 1944. 1
- L. Kanner et al. Autistic disturbances of affective contact. *Nervous child*, 2(3):217–250, 1943. 9
- A. Kappas, D. Küster, C. Basedow, and P. Dente. A validation study of the affectiva qsensor in different social laboratory situations. In 53rd Annual Meeting of the Society for Psychophysiological Research, Florence, Italy, 2013. 39
- J. 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. 11
- P. Kern and D. 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. 11

- J. Kim, T. Wigram, and C. 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. 23
- S. Kim and E. Clarke. Case study: An ipad-based intervention on turn-taking behaviors in preschoolers with autism. *Behavioral Development Bulletin*, 20(2):253, 2015. 22
- H. Kozima, C. Nakagawa, and Y. 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. 25, 27
- J. Kwon, D.-H. Kim, W. Park, and L. Kim. A wearable device for emotional recognition using facial expression and physiological response. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 5765–5768. IEEE, 2016. 38, 39
- A. Kylliäinen and J. K. Hietanen. Skin conductance responses to another person's gaze in children with autism. *Journal of autism and developmental disorders*, 36(4):517–525, 2006. 38
- B.-L. J. M. M. E. C. 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. 39
- B. Laurel. Computers as theatre. Addison-Wesley, 2013. 4
- J. Legiša, D. S. Messinger, E. Kermol, and L. Marlier. Emotional responses to odors in children with high-functioning autism: autonomic arousal, facial behavior and selfreport. *Journal of autism and developmental disorders*, 43(4):869–879, 2013. 40, 41
- D. B. LeGoff. Use of lego© as a therapeutic medium for improving social competence. *Journal of autism and developmental disorders*, 34(5):557–571, 2004. 22
- D. B. Legoff and M. Sherman. Long-term outcome of social skills intervention based on interactive lego© play. *Autism*, 10(4):317–329, 2006. 22
- J. Li, L. Zhang, D. Tao, H. Sun, and Q. 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. 46

- L. Lidberg and B. G. Wallin. Sympathetic skin nerve discharges in relation to amplitude of skin resistance responses. *Psychophysiology*, 18(3):268–270, 1981. 38
- H. A. Lim and E. 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. 23
- Y.-P. Lin, C.-H. Wang, T.-L. Wu, S.-K. Jeng, and J.-H. Chen. Multilayer perceptron for eeg signal classification during listening to emotional music. In *TENCON 2007-2007 IEEE region 10 conference*, pages 1–3. IEEE, 2007. 40
- F. Liu, G. Liu, and X. Lai. Emotional intensity evaluation method based on galvanic skin response signal. In 2014 Seventh International Symposium on Computational Intelligence and Design, volume 1, pages 257–261. IEEE, 2014. 40
- M. Liu, D. Fan, X. Zhang, and X. 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. 54
- C. Lord. The development of peer relations in children with autism. *Advances in applied developmental psychology*, pages 165–229, 1984. 20
- C. Lord, E. H. Cook, B. L. Leventhal, and D. G. Amaral. Autism spectrum disorders. *Neuron*, 28(2):355–363, 2000. 1
- V. Lotter. Epidemiology of autistic conditions in young children. *Social psychiatry*, 1(3): 124–135, 1966. 15
- M. J. Maenner, K. A. Shaw, J. 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. 2
- S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on pattern analysis and machine intelligence*, 11(7), 1989. 44, 45
- J. Manjiviona and M. 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. 18

- S. M. Mavadati, H. Feng, A. Gutierrez, and M. 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. 3, 34
- A. S. 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. 20
- K. Mera and T. Ichimura. Emotion analyzing method using physiological state. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 195–201. Springer, 2004. 54, 57
- F. Michaud and S. Caron. Roball, the rolling robot. *Autonomous robots*, 12(2):211–222, 2002. 2, 6
- F. Michaud and C. Théberge-Turmel. Mobile robotic toys and autism. pages 125–132, 2002. 2, 6
- D. Mihalache, H. Feng, F. Askari, P. Sokol-Hessner, E. J. Moody, M. H. Mahoor, and T. D. Sweeny. Perceiving gaze from head and eye rotations: An integrative challenge for children and adults. *Developmental Science*, 23(2):e12886, 2020. 34
- X. Ming, M. Brimacombe, and G. C. Wagner. Prevalence of motor impairment in autism spectrum disorders. *Brain and Development*, 29(9):565–570, 2007. 18
- P. Mirenda. Autism, augmentative communication, and assistive technology: What do we really know? *Focus on autism and other developmental disabilities*, 16(3):141–151, 2001. 3
- M. Miyahara, M. Tsujii, M. Hori, K. Nakanishi, H. Kageyama, and T. Sugiyama. Brief report: motor incoordination in children with asperger syndrome and learning disabilities. *Journal of autism and developmental disorders*, 27(5):595–603, 1997. 18
- I. Molnar-Szakacs and P. Heaton. Music: a unique window into the world of autism. *Annals* of the New York Academy of Sciences, 1252(1):318–324, 2012. 35
- P. Mundy and M. 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. 20

- P. Mundy, M. Sigman, J. Ungerer, and T. 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. 20
- A. K.-P.-I. A. L. F. B. 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. 43, 44
- F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology* & Work, 6(1):4–14, 2004. 40
- R. S. Neel, Z. N. Jenkins, and N. Meadows. Social problem-solving behaviors and aggression in young children: A descriptive observational study. *Behavioral Disorders*, 16(1): 39–51, 1990. 20
- W. 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. 41
- L. A. Obringer and S. Jonathan. Honda asimo robot. July 2011. 7, 8
- O. Ousley, R. Arriaga, G. Abowd, and M. Morrier. Rapid assessment of socialcommunicative abilities in infants at risk for autism. *Technical Report, Georgia Tech*, 2012. 42
- G. Owens, Y. Granader, A. Humphrey, and S. 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. 22
- Y.-H. Peng, C.-W. Lin, N. M. Mayer, and M.-L. 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. 34

- K. Pierce and L. 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. 2
- A. C. Pierno, M. Mari, D. Lusher, and U. Castiello. Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46(2):448–454, 2008. 8, 19, 21
- G. Pioggia, R. Igliozzi, M. Ferro, A. Ahluwalia, F. Muratori, and D. 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. 25, 28, 32
- K. Popper. The logic of scientific discovery. Routledge, 2005. 16
- B. Provost, S. Heimerl, and B. 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, 2007a. 20
- B. Provost, B. R. Lopez, and S. 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, 2007b. 18
- V. S. Ramachandran and L. M. Oberman. Broken mirrors: A theory of autism. *Scientific American*, 295(5), 2006. 16
- E. Redcay and E. 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. 16
- V. Reddy. Playing with others' expectations: Teasing and mucking about in the first year. 1991. 20
- V. Reddy, D. Hay, L. Murray, and C. Trevarthen. Communication in infancy: Mutual regulation of affect. *Infant development: Recent advances*, page 247, 1997. 23
- J. Rehg, G. Abowd, A. Rozga, M. Romero, M. Clements, S. Sclaroff, I. Essa, O. Ousley, Y. Li, C. Kim, et al. Decoding children's social behavior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3414–3421, 2013a. 38, 39

- J. Rehg, G. Abowd, A. Rozga, M. Romero, M. Clements, S. Sclaroff, I. Essa, O. Ousley, Y. Li, C. Kim, et al. Decoding children's social behavior. In *Proceedings of the IEEE* conference on computer vision and pattern recognition, pages 3414–3421, 2013b. 42
- A. E. Reschke-Hernández. History of music therapy treatment interventions for children with autism. *Journal of Music Therapy*, 48(2):169–207, 2011. 10
- D. J. Ricks and M. B. Colton. Trends and considerations in robot-assisted autism therapy. In 2010 IEEE international conference on robotics and automation, pages 4354–4359. IEEE, 2010. 24, 25, 26, 27
- S. R. Rieth, A. C. Stahmer, J. Suhrheinrich, L. Schreibman, J. Kennedy, and B. 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. 21, 22
- B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. 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. 25, 27
- B. Robins, K. Dautenhahn, and J. Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction studies*, 7(3):479–512, 2006a.
- B. Robins, K. Dautenhahn, and J. Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction studies*, 7(3):479–512, 2006b. 25, 27, 32
- B. Robins, K. Dautenhahn, and P. 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. 11
- N. Roper. Melodic intonation therapy with young children with apraxia. *Bridges*, 1(8):1–7, 2003. 23
- 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. 41

- R. A. Sarokoff and P. Sturmey. The effects of behavioral skills training on staff implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis*, 37(4):535–538, 2004. 2
- B. Scassellati, H. Admoni, and M. Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14, 2012. 21
- S. Schmidt and H. Walach. Electrodermal activity (eda)–state-of-the-art measurements and techniques for parapsychological purposes. *Journal of Parapsychology*, 64(2), 2000. 40
- E. Schopler, G. B. Mesibov, and K. Hearsey. Structured teaching in the teach system. In *Learning and cognition in autism*, pages 243–268. Springer, 1995. 3
- B. T. Shahani, J. Halperin, P. Boulu, and J. Cohen. Sympathetic skin response–a method of assessing unmyelinated axon dysfunction in peripheral neuropathies. *Journal of Neurology, Neurosurgery & Psychiatry*, 47(5):536–542, 1984. 38
- S. Shamsuddin, H. Yussof, L. Ismail, F. A. Hanapiah, S. Mohamed, H. A. Piah, and N. I. 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, 2012a. 33
- S. Shamsuddin, H. Yussof, L. I. Ismail, S. Mohamed, F. A. Hanapiah, and N. I. Zahari. Humanoid robot nao interacting with autistic children of moderately impaired intelligence to augment communication skills. *Procedia Engineering*, 41:1533–1538, 2012b. 33
- S. Shamsuddin, H. Yussof, L. I. Ismail, S. Mohamed, F. A. Hanapiah, and N. I. 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, 2012c. 33
- S. Shamsuddin, H. Yussof, M. Miskam, A. C. Hamid, N. A. Malik, and H. 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. 33
- P. 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. 39

- M. Sigman, E. Ruskin, S. Arbelle, R. Corona, C. Dissanayake, M. Espinosa, N. Kim, A. López, C. Zierhut, C. 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. 20
- A. D. Sorosky, E. M. Ornitz, M. B. Brown, and E. R. Ritvo. Systematic observations of autistic behavior. *Archives of General Psychiatry*, 18(4):439–449, 1968. 9
- S. D. Stagg, R. Davis, 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):2303–2311, 2013. 38
- E. Starr and K. Zenker. Understanding autism in the context of music therapy: Bridging theory and practice. *Canadian Journal of Music Therapy*, 1998. 11
- C. E. Stephens. Spontaneous imitation by children with autism during a repetitive musical play routine. *Autism*, 12(6):645–671, 2008. 23
- W. L. Stone and P. J. Yoder. Predicting spoken language level in children with autism spectrum disorders. *Autism*, 5(4):341–361, 2001. 20
- S. Sutera, J. Pandey, E. L. Esser, M. A. Rosenthal, L. B. Wilson, M. Barton, J. Green, S. Hodgson, D. L. Robins, T. 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. 18
- 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):63–75, 2012. 57
- 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. 39, 43
- A. Taheri, M. Alemi, A. Meghdari, H. Pouretemad, N. M. Basiri, and P. 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. 34

- A. Taheri, A. Meghdari, M. Alemi, H. Pouretemad, P. Poorgoldooz, and M. Roohbakhsh. Social robots and teaching music to autistic children: Myth or reality? In *International Conference on Social Robotics*, pages 541–550. Springer, 2016. 34
- K. Tang, J. Diehl, M. Villano, K. Wier, B. Thomas, N. Shea, Z. DuBois, M. Millea, K. Uhland, and C. 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. 19, 21
- D. Tapscott. *Growing up digital*, volume 302. McGraw-Hill Companies San Francisco, 1998. 4
- M. Tarvainen, P. Karjalainen, A. Koistinen, and M. Valkonen-Korhonen. Principal component analysis of galvanic skin responses. In *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.* 00CH37143), volume 4, pages 3011–3014. IEEE, 2000. 38
- P. Teitelbaum, O. Teitelbaum, J. Nye, J. Fryman, and R. 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. 18
- M. H. Thaut and A. A. Clair. *A scientific model of music in therapy and medicine*. IMR Press, 2000. 10
- C. Trevarthen. Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. *Action, gesture, and symbol: The emergence of language*, 1978. 20
- H. Van Waelvelde, A. Oostra, G. Dewitte, C. Van den Broeck, and M. 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. 18
- M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *IEEE transactions* on signal processing, 40(9), 1992. 44, 45, 46
- M. Villano, C. R. Crowell, K. Wier, K. Tang, B. Thomas, N. Shea, L. M. Schmitt, and J. 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. 8, 19, 21

- K. Wada, T. Shibata, T. Saito, and K. Tanie. Analysis of factors that bring mental effects to elderly people in robot assisted activity. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 1152–1157. Ieee, 2002. 5
- J. Wainer, K. Dautenhahn, B. Robins, and F. 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. 11
- G. Wang, P. R. Cook, et al. Chuck: A concurrent, on-the-fly, audio programming language. In *ICMC*, 2003. 106
- A. Warwick and J. Alvin. *Music therapy for the autistic child*. Oxford University Press, 1991. 12
- I. Werry and K. Dautenhahn. Applying mobile robot technology to the rehabilitation of autistic children. In *In: Procs SIRS99, 7th Symp on Intelligent Robotic Systems*, 1999. 4
- I. Werry, K. Dautenhahn, and W. Harwin. Investigating a robot as a therapy partner for children with autism. *Procs AAATE 2001*, 2001. 25
- D. C. Wimpory and S. Nash. Musical interaction therapy-therapeutic play for children with autism. *Child Language Teaching and Therapy*, 15(1):17–28, 1999. 24
- L. Wing. The autistic spectrum. The lancet, 350(9093):1761-1766, 1997. 1
- S. Wolff and S. Chess. A behavioural study of schizophrenic children. *Acta Psychiatrica Scandinavica*, 1964. 2, 8, 18, 21
- B. Xu, Y. Fu, Y.-G. Jiang, B. Li, and L. Sigal. Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization. *IEEE Transactions on Affective Computing*, 9(2):255–270, 2016. 40
- P. Yoder and W. L. Stone. Randomized comparison of two communication interventions for preschoolers with autism spectrum disorders. *Journal of consulting and clinical psychology*, 74(3):426, 2006. 22
- P. J. Yoder and S. 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. 22

T. Zane, C. Davis, and M. Rosswurm. The cost of fad treatments in autism. *Journal of Early and Intensive Behavior Intervention*, 5(2):44, 2008. 4

Appendices

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

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

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

VITA

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!