UNIVERSITY OF DENVER

XYLO-BOT: A SOCIAL MUSIC THERAPY PLATFORM FOR CHILDREN WITH AUTISM

By

Huanghao Feng

A COMPREHENSIVE EXAM

Submitted to the Faculty
of the University of Denver
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Denver, Colorado

August 2019

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Autism Spectrum Disorders (ASD)

Autism is a general term used to describe a spectrum of complex developmental brain disorders causing qualitative impairments in social interaction and results in repetitive and stereotyped behaviors. Currently one in every 88 children in the United States are diagnosed with ASD and government statistics suggest the prevalence rate of ASD is increasing 10-17 percent annually [8]. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, and eye gaze attention [9]. Currently, clinical work such as Applied Behavior Analysis (ABA) [12, 13] has focused on teaching individuals with ASD appropriate social skills in an effort to make them more successful in social situations [20]. With the concern of the growing number of children diagnosed with ASD, there is a high demand for finding alternative solutions such as innovative computer technologies and/or robotics to facilitate autism therapy. Therefore, research into how to design and use modern technology that would result in clinically robust methodologies for autism intervention is vital. In social human interaction, non-verbal facial behaviors (e.g. facial expressions, gaze direction, and head pose orientation, etc.) convey important information between individuals. For instance, during an interactive conversation, the peer may regulate their facial activities and gaze directions actively to indicate the interests or boredom. However, the majority of individuals with ASD show the lack of exploiting and understanding these cues to communicate with others. These limiting factors have made crucial difficulties for individuals with ASD to illustrate their emotions, feelings and also interact with other human beings. Studies have shown that individuals with autism are much interested to interact with machines (e.g. computers, iPad, robots, etc.) than humans [17]. In this regard, in the last decade several studies have been conducted to employ machines in therapy sessions and examine the behavioral responses of people with autism. These studies have assisted researchers to better understand, model and improve the social skills of individuals on the autism spectrum. This thesis presents the methodology and results of a study that aimed to design a humanoid-robot therapy sessions for capturing, modeling and enhancing the social skills of children with Autism. In particular we mainly focus on gaze direction and joint attention modeling and analysis and investigate how the ASD and Typically Developing (TD) children employ their gaze for interacting with the robot. In the following section, we have a brief introduction of the existing assistive robots in the following section and how they have been used in autism applications.

1.2 Socially Assistive Robotics

Socially Assistive Robotics (SAR) can be considered as the intersection of Assistive Robotics (AR) and Socially Interactive Robotics (SIR), which has referred to robots that assist human with physical deficits and also can provide certain terms of social interaction abilities [6]. SAR contains all properties of SIR described in [17], and also a few additional attributes such as: 1) user populations (different groups of users, i.e. elders; individuals with physical impairments; kids diagnosed with ASD; students); 2) social skills (i.e. speech ability; gestures movement); 3) objective tasks (i.e. tutoring; physical therapy; daily life assistance); 4) role of the robot (depends on the task the robot has been assigned for) [6]. Companion robots [10] is one type of SAR that are widely used for elderly people for health care supports. Research shows that this type of social robots can reduce stress and depression of individuals in elderly stage [5]. Service social robots are able to accomplish a variety of tasks for individuals with physical impairments [8]. 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 [9, 12, 13]. Nowadays, there are very few companies that have been designing and producing socially assistive robots. The majority of existing SARs are not commercialized yet and because of being expensive and not well-designed user interfaces, they are mostly used forthe research purposes. Honda, Aldebaran Robotics and Hanson Robokind are the top leading companies that are currently producing humanoid robots. Ideally socially assistive robots can have fully automated systems to detect and express social behaviors while interacting with humans. Some of the existing robot-human interfaces are semi-autonomous and they can recognize some basic biometrics (e.g. visual and audio commands of the user) and behavioral response. Besides, the majority of existing robots are very complicated to work with. Therefore in the last couple of years several companies have started to make these robots more user-friendly and responsive to both the user need and the potential caregiver commands [6]. Intelligent SARs aim to have the capability to recognize visual or audio commands, objects, and specific human gestures. Some of these robots have the ability of detect human face or basic facial expressions. For instance, ASIMO, a robot developed by Honda, it has a sensor for detecting the movements of multiple objects by using visual information captured from two cameras on its head. Plus its “eyes” can measure the distance of the objects from the robot [14]. Another example is from Aldebaran Robotics which designs small size humanoid robots, called NAO. NAO robot has two cameras attached that are used to capture single images and video sequences. This capturing module enables NAO to see the different sides of an object and recognize it for future use. Furthermore, NAO has a remarkable capability of recognizing faces and detecting moving objects. Both of the aforementioned robots have speech recognition system. They can interpret voice commands to accomplish a certain set of tasks which have been pre-programmed in the system. NAO is able to identify words for running specific commands. However ASIMO is able to distinguish between voices and other sounds. This feature empowers ASIMO to perceive the direction of human's speaker or recognize other companion robots by tracking their voice [1]. These robots can also speak in many different languages. For example, NAO can speak in English, French, Chinese, Japanese and other languages up to more than ten languages. This feature gives the robot a great social communication functionality to interact with humans from all over the world.

1.2.1 Socially Assistive Robots for Autism Therapy

Socially assistive robots are emerging technologies in the field of robotics that aim to utilize social robots to increase engagement of users as communicating with robots, and elicit novel social behaviors through their interaction. One of the goal in SAR is to use social robots either individually or in conjunction with caregivers to improve social skills of individuals who have social behavioral deficits. One of the early applications of SAR is autism rehabilitation. As mentioned before, autism is a spectrum of complex developmental brain disorders causing qualitative impairments in social interaction. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, and gaze regulation. These skill deficits often pose problems in the individual's ability to establish and maintain social relationships and may lead to anxiety surrounding social contexts and behaviors [20]. 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 [15, 18, 19, 6, 17]. These positive behaviors include showing emotional facial expressions (e.g., smiling), gesture imitation, and eye gaze attention. Studies show that these behaviors are rare in children with autism but evidence suggests that robots trigger children to demonstrate such behaviors. These investigations propose that interaction with robots may be a promising approach for rehabilitation of children with ASD. There are several research groups that investigated the response of children with autism to both humanoid robots and non-humanoid toy-like robots in the hope that these systems will be useful for understanding affective, communicative, and social differences seen in individuals with ASD (see Diehl et al., [17]), and to utilize robotic systems to develop novel interventions and enhance existing treatments for children with ASD [14, 1, 2]. Mazzei et al. [4], for example, designed the robot "FACE" to realistically show the details of human facial expressions. A combination of hardware, wearable devices, and software algorithms measured subject's affective states (e.g., eye gaze attention, facial expressions,vital signals, skin temperature and EDA signals), were used for controlling the robot reactions and responses. Reviewing the literature in SAR [6, 17] shows that there are surprisingly very few studies that used an autonomous robot to model, teach or practice the social skills of individuals with autism. Amongst, teaching how to regulate eye-gaze attention, perceiving and expressing emotional facial expressions are the most important ones. Designing robust interactive games and employing a reliable social robot that can sense users' socioemotional behaviors and can respond to emotions through facial expressions or speech is an interesting area of research. In addition, the therapeutic applications of social robots impose conditions on the robot's requirements, feedback model and user interface. In other words, the robot that aims for autism therapy may not be directly used for depression treatment and hence every SAR application requires its own attention, research, and development Only a few adaptive robot-based interaction settings have been designed and employed for communication with children with ASD. Proximity-based closed-loop robotic interaction [3], haptic interaction [7], and adaptive game interactions based on affective cues inferred from physiological signals [16] are some of these studies. Although all of these studies were conducted to analyze the functionality of robots for socially interacting with individuals with ASD, these paradigms were limitedly explored and focused on their core deficits (i.e., Facial expression, eye gaze and joint attention skills). Bekele and colleagues [11] studied the development and application of a humanoid robotic system capable of intelligently administering joint attention prompts and adaptively responding based on within system measurements of gaze and attention. They found out that preschool children with ASD have more frequent eye contact toward the humanoid robot agent, and also more accurate respond in joint attention stimulations. This suggests that robotic systems have the enhancements for successfully improve the coordinated attention in kids with ASD. Considering the existing SAR system and the major social deficits that individuals with autism may have, we have designed and conducted robot-based therapeutic sessions that are focused on different aspects of social skills of children with autism. In this thesis we employed NAO which can be remotely controlled to communicate with the children. We conducted two different protocols to examine the social skills of children with autism and provide feedbacks to improve their behavioral responses. The contribution of our work has been introduced in Section 1.4 and the details of the game setting, experiments, modeling and analysis are provided in Chapter 4.

1.3 Music Therapy for ASD

1.4 Contributions

1.5 Orgizaition

1.6 Dissertation Outline

This thesis is organized as follows: In Chapter two, we present related work for facial features extraction, two dimensional (2-D), three dimensional (3-D), and multi-modal (2-D + 3-D) face recognition. Chapter three explains our algorithm for 2-D facial feature extraction from frontal face images (i.e., Improved ASM) and our algorithm for 3-D facial feature extraction (i.e., the extraction of the three feature points) along with the experimental results. Chapter four presents our approach for 3-D face modeling and recognition based on ridge images. Chapter five describes our multi-modal face modeling and recognition (2-D/3-D) based on attributed relational graphs along with the experiments. In addition, we present two fusion techniques for combining the 2-D and 3-D modalities in this chapter. Finally in Chapter six, we present the conclusion and the future research directions.

Chapter 2

Xylo-Bot: An Interactive Music Teaching System

A novelty Interactive human-robot music teaching system design is presented in this chapter. Two In order to make robot play xylophone properly, several things need to be done before that. First is to find a proper xylophone with correct timber; second, we have to make the xylophone in a proper position in front of the robot that makes it to be seen properly and be reached to play; finally, design the intelligent music system for NAO.

2.1 NAO: A Humanoid Robot

We used a humanoid robot called NAO developed by Aldebaran Robotics in France. NAO is 58 cm (23 inches) tall, with 25 degrees of freedom this robot can conduct most of the human behaviors. It also features an onboard multimedia system including, four microphones for voice recognition, and sound localization, two speakers for text-to-speech synthesis, and two HD cameras with maximum image resolution 1280 x 960 for online observation. As shown in Figure 2.1, these utilities are located in the middle of the forehead and the mouth area. NAO's computer vision module includes facial and shape recognition units. By using the vision feature of the robot, that allows the robot be able to see the instrument from its lower camera and be able to do implement a eye-arm self-calibration system which allows the robot to have real-time micro-adjustment of its arm-joints in case of off positioning during music playing.

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

Four microphones embedded on toy or NAO's head locations see figure 2.2. According the official Aldebaran documentation, these microphones has sensitivity of 20mV/Pa +/-3dB at 1kHz, and the input frequency range of 150Hz - 12kHz, data will be recorded as a 16 bits, 48000Hz, 4 channels wav file which meets the requirements for designing the online feedback audio score system which will be described below.

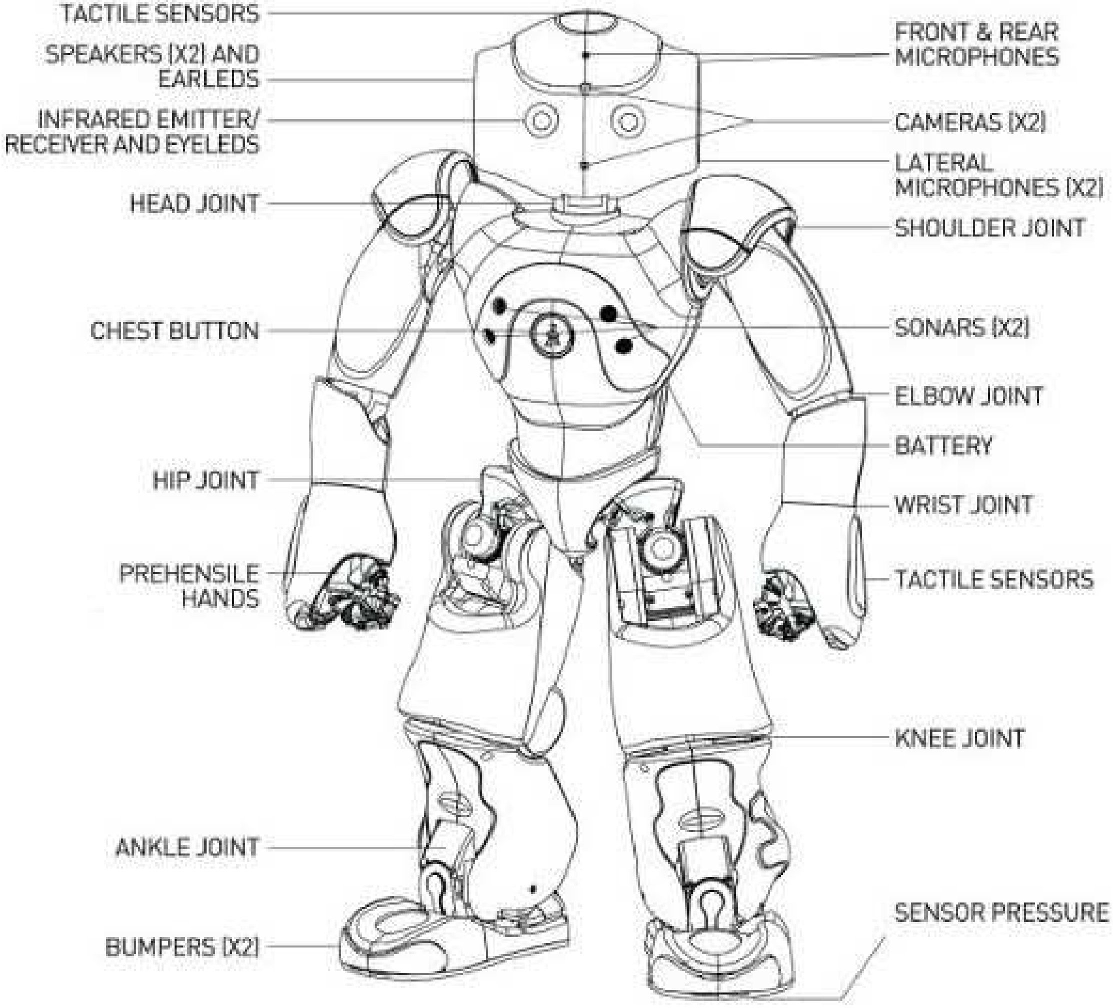


Figure 2.1: *A Humanoid Robot NAO: 25 Degrees of Freedom, 2 HD Cameras and 4 Microphones*

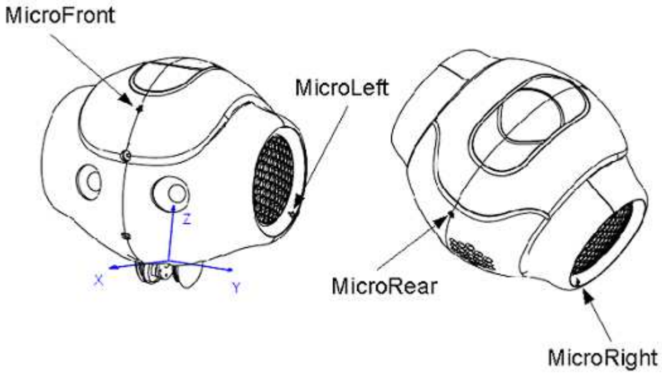


Figure 2.2: *Microphone locations on NAO's head*

2.2 Accessories

The purpose of this study is to have a toy size humanoid robot to play music, some necessary accessories need to be purchased and made before teach the robot to play music. All accessories will be discussed in the following sections.

2.2.1 Xylophone: A Toy for Music Beginner

In this system, according NAO's open arms' length, we choose a Sonor Toy Sound SM soprano-xylophone with 11 sound bars of 2 cm in width. The instrument has a size of 31 cm x 9.5 cm x 4 cm, including the resonateing 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. The beaters/mallets, we use the pair which come with the xylophone with a modified 3D printed grips (details in next subsection) to allow the robot's hands to hold them properly. The mallets are approximately 21 cm in length include a head of 0.8 cm radius. See Figure 2.3.

11 bars represent 11 different notes (11 frequencies) which covers approximate one

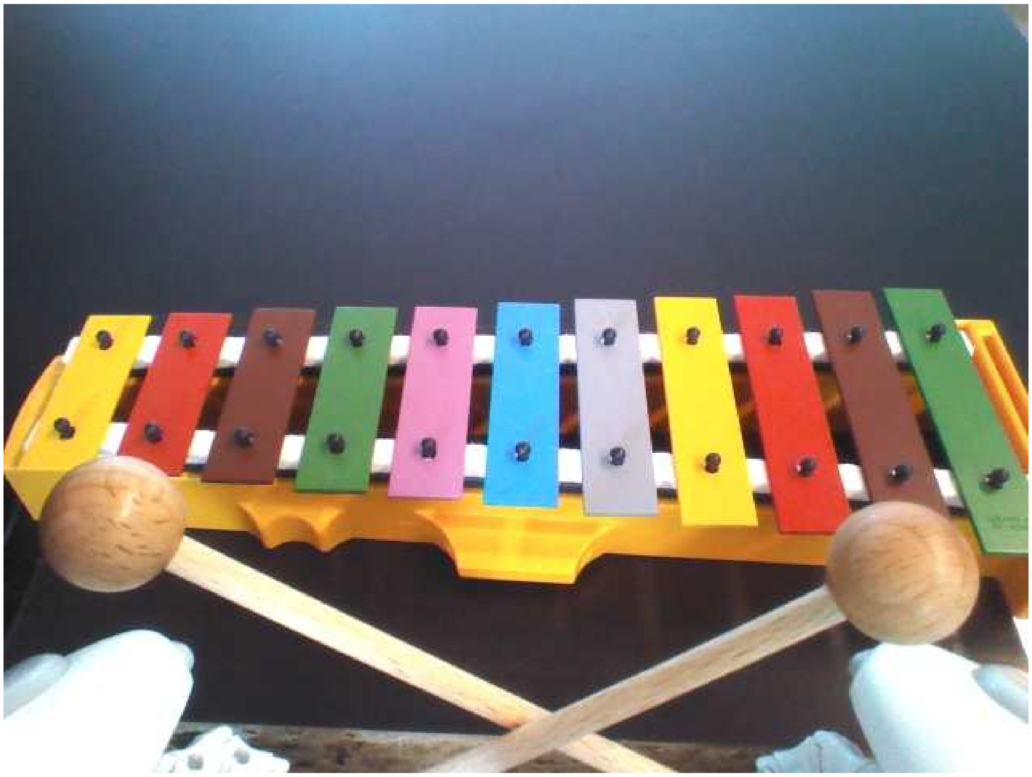


Figure 2.3: *Actual Xylophone and Mallets from NAO's Bottom Camera*

and half octave scale starting from C6 to F7.

2.2.2 Mallet Gripper Design

According to NAO's hands size, we designed and 3D printed a pair of gripers to have the robot be able to hold the mallets properly. All dimensions can be found in figure somewhere.

2.2.3 Instrument Stand Design

A wooden base has designed and laser cut to hold the instrument in a proper place in order to have the robot be able to play music. All dimensions can be found in figure somewhere below.

2.3 Module-Based Acoustic Music Interactive System Design

In this section, a novelty module-based robot-music teaching system will be presented. Three modules have built in this intelligent system including module 1: eye-hand self-calibration micro-adjustment; module 2: joint trajectory generator; and module 3: real time performance scoring feedback. See Figure 2.4

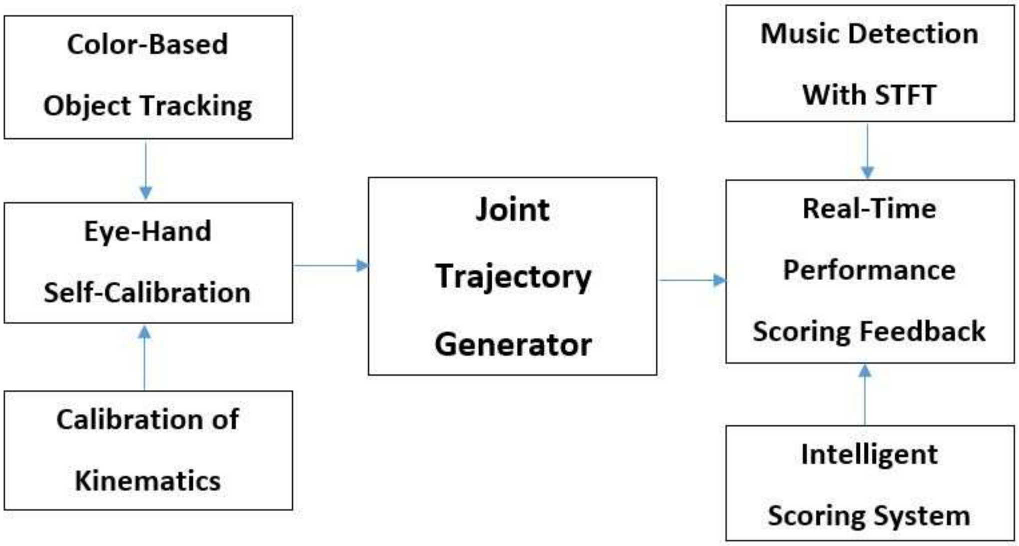

Color-Based Object Tracking
Eye-Hand Self-Calibration
Calibration of Kinematics
Joint Trajectory Generator
Music Detection With STFT
Real-Time Performance Scoring Feedback
Intelligent Scoring System

Figure 2.4: *Block Diagram of Module-Based Acustic Music Interactive System*

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

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

A. Color-Based Object Tracking

To play the xylophone, the robot has to be able to adjust its motions according to the estimated relative poses of the instrument and the heads of the beaters it is holding. The approach to estimating these poses which adopted in this thesis, we uses a color-based technique.

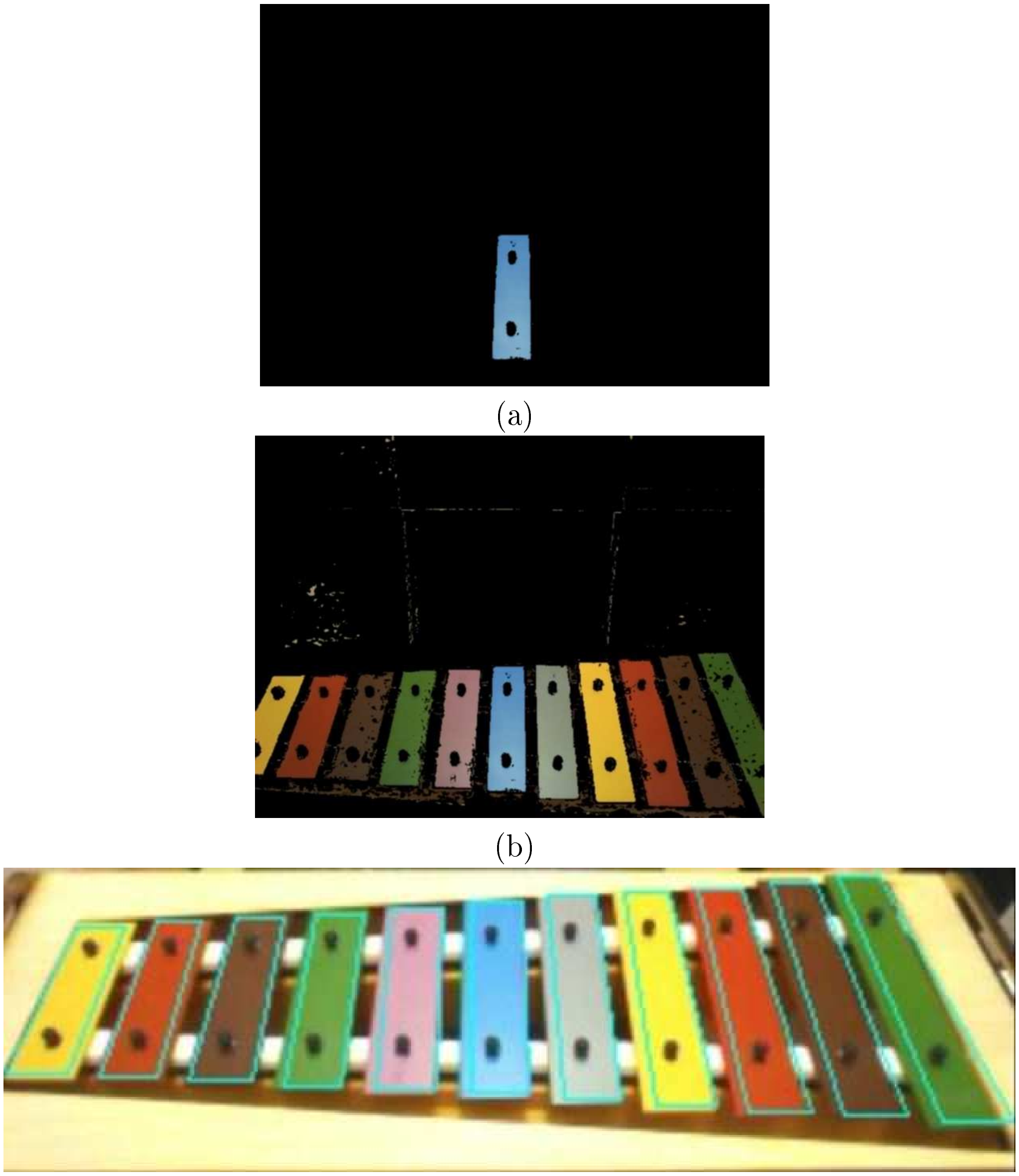
The main idea is, based on the RGB color of the center blue bar, given a hypothesis about the instrument's pose, one can project the contour of the object's model into the camera image and compare them to actually observed contour. In this way, it is possible to estimate the likelihood of the pose hypothesis. By using this method, it allows the robot to track the instrument with very low cost in real-time. See Figure 2.5

B. Calibration of Kinematic Parameters

(In progress, will not present in this version. The idea is to use both positions of the instrument and beaters' heads to computes for each sound bar a suitable beating configuration for arm kinematics chain. Suitable means that the beater's head can be placed on the surface of the sound bar at the desired angle. From this configuration, the control points of a predefined beating motion are updated.)

2.3.2 Module 2: Joint Trajectory Generator

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



(a)

(b)

(c)

Figure 2.5: *Color Detection From NAO's Bottom Camera: (a) Single Blue Color Detection (b) Full Instrument Color Detection (c) Color Based Edge Detection.*

2.3.3 Module 3: Real-Time Performance Scoring Feedback

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

A. Music Detection

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

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

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

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

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

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

B. Intelligent Scoring-Feedback System

In order to compare the detected notes and the target notes, we used an algorithm which normally used in information theory, linguistics called Levenshtein Distance. This algorithm is a string metric for measuring the difference between two sequences.

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

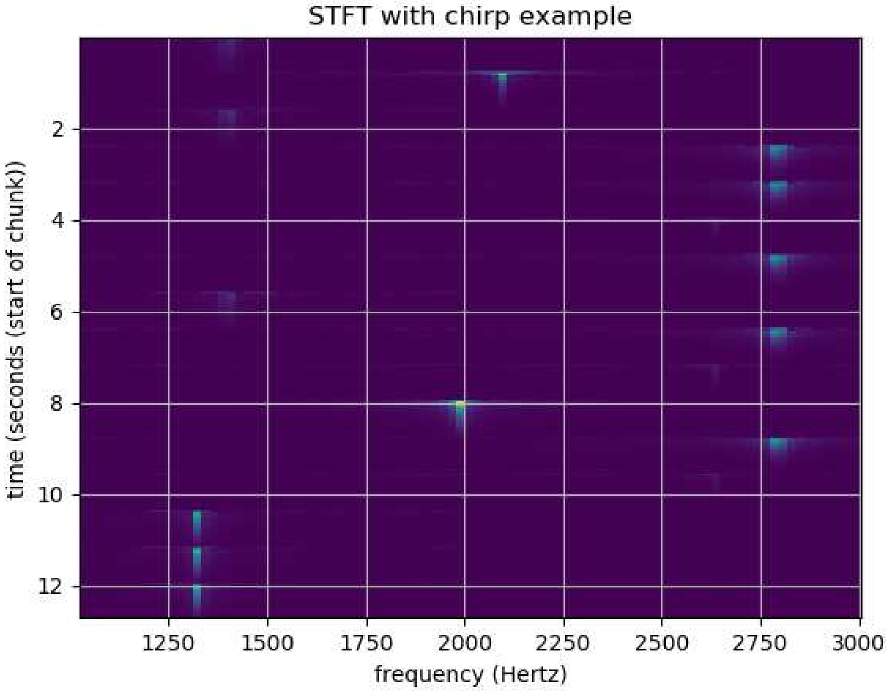
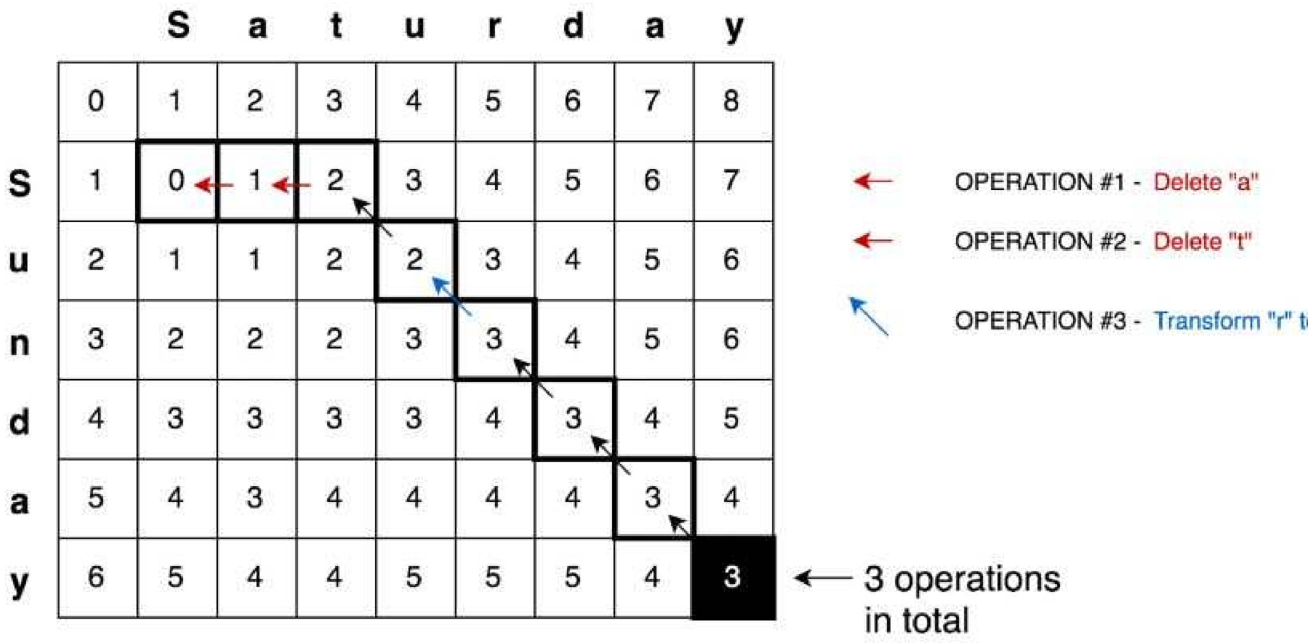


Figure 2.6: *Melody Detection with Short Time Fourier Transform*

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

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 $\operatorname{lev}_{a,b}(i,j)$ 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 2.7 demonstrates how to apply this principle in finding the Levenshtein distance of two words "Sunday" and "Saturday".



		S	a	t	u	r	d	a	y
	0	1	2	3	4	5	6	7	8
S	1	0	1	2	3	4	5	6	7
u	2	1	1	2	2	3	4	5	6
n	3	2	2	2	3	3	4	5	6
d	4	3	3	3	3	4	3	4	5
a	5	4	3	4	4	4	4	3	4
y	6	5	4	4	5	5	5	4	3

OPERATION #1 - Delete "a"
OPERATION #2 - Delete "t"
OPERATION #3 - Transform "r" t
3 operations in total

Figure 2.7: *An Example of Compute Levenshtein Distance for "Sunday" and "Saturday"*

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

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

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

2.4 Summary

In this chapter, based on the purpose of this research, we have discussed both hardware and software design for getting ready to design the experiment sessions.

From Chapter One, we determined to have NAO as a music teaching instructor be able to teach children simple music and be able to deliver social content in the mean time. In order to have the system ready, first of all, we have to choose the proper agent which should be a kids friendly with social ability robot NAO. Second, based on the size of the robot, some necessary accessories have to purchased and handcrafted. A toy size color coded xylophone became the best option and based on the size and position, a wooden based xylophone stand was customized and assembled. Due to the limitation of NAO's hand size, a pair of mallet gripers have been 3-D printed and customized. Last, a intelligent module-based acoustic music interactive system has designed from scratch. Three modules has designed to have the robot to play, listen and teach the music freely and can become a great companion for children in both music learning and social life.

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

Chapter 3

Protocol 1: Acoustic Music Teaching Experiment Design Data Acquisition and Result

Few questions will be answered in terms of social skills in this chapter: 1) Turn Taking: How well kids with autism behave during the teaching and learning process compare to TD group, e.g., kid listen to the instruction or demonstration from robot before plays instrument; 2) Joint Attention and Eye-Gaze Attention: How well kids with ASD follow instructions and adapt hints compare to TD kids, e.g., kid follows hitting positions demonstrated by robot or follows the change of eye colors; 3) Motor Control: How well ASD kids play the xylophone in terms of volume, pitch and accuracy compare to TD group, e.g., a good multiple strikes should be recognized by STFT as a sequence of frequencies that designed from previous chapter; 4) Engagement and Event Based Emotion: How kids engage to different music teaching events and what are the emotions during certain events, e.g., using EDA signal to find out the emotions regarding different situations for example having hard time memorizing

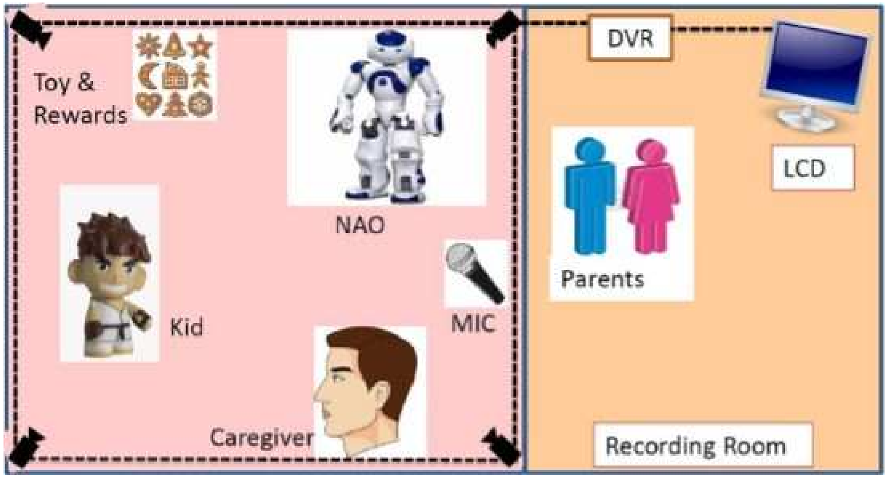


Figure 3.1: *Schematic robot based therapy session and video capturing setting*

a sequence of notes; 5) Facial Expression and Emotions Correlation: (will not be presented in this version); 6) Music Emotions and Feelings: How kids perceive emotions in small pieces of music within different melodies, e.g., music in different keys.

3.1 Room Setup and Participants

Figure 3.1 shows the experiment

3.2 Experiment Design

In order to collect all the data for answering the questions above, a set of intervention sessions has designed using music therapy concept. 6 - 7 sessions has settled for ASD group and 2 sessions for TD group. For ASD group, entire sessions have been divided into 3 parts: baseline session, intervention sessions and exit session. Similar to ASD group, TD group only includes baseline session and exit session for comparison purpose. In addition, a social music game play has also included in each session for system testing and entertainment purpose.

Baseline Session: Participants has been asked to follow all the instructions contains all the practices from the following intervention sessions. Including single bar strike, multiple bars strike, half song play and the whole song play. We choose a very popular kid's song "Twinkle Twinkle Little Star" for this specific session due to the well known of this song in almost everyone's childhood.

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

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

Due to the difficulty of user selected songs and difference performance score of participants, the total session numbers can be various, 6 visits will be the minimum requirement for ASD group and total visits cannot go beyond 8 times. More detailed experiment design is shown in the table below:

3.3 Methodology and Experiment Result

3.3.1 Social Aspects Annotation and Coding Methods

This part is to describe how to do video coding in these social features. Basically just post process by watching the front face videos and by listen to the audio to code how the social behaviors been performed. Have to copy some of the stuff from previous studies.

Turn-Taking Coding

Joint Attention Coding

3.3.2 Motor Control Scoring System

This is related to hitting practice and play practice in real time during the session. Need to have connections with the module 2 real time performance scoring feedback system. Describe the how the flow goes during the session and how to have small add on dosage if the accuracy not good. Also have to make up a story when it goes to the half/whole song, how to manage that if the kid cannot play well, just cut the whole chunk into small continuous pieces and use the same methods and scoring system to deliver the idea.

3.3.3 Emotion Classification

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

A. Continuous Wavelet Transform The EDA data recorded using the SC sensors are categorized as non-stationery signals (Najafi et al., 2003; Swangnetr et al., 2013). Hence, multiresolution analysis techniques are essentially suitable to study the qualitative components of these kinds of bio-signals (Najafi et al., 2003). Note that continuous wavelet transform (CWT) is one of the strongest and most widely used analytical tools for multiresolution analysis. CWT has received considerable attention in processing signals with non-stationary spectra (Vetterli, & Herley, 1992; Mallat, 1989); therefore, it is utilized here to perform the time-frequency analysis of the EDA signals. In contrast to many existing methods that utilize the wavelet coefficients of the raw signal to extract features, our proposed method is essentially based on the spectrogram of the original data in a specific range of frequency (0.5, 50)Hz, which provides more information for other post-processing (i.e., feature extraction and classification) steps. We apply the wavelet transform at various scales corresponding to the aforementioned frequency range to calculate the spectrogram of the raw signal (i.e., Short Time Fourier Transform (STFT) can also be used to calculate the spectrogram of the raw signal). In addition, as opposed to many related studies that utilize real-valued wavelet functions for feature extraction purposes, we have employed the complex Morlet (C-Morlet) function with the proposed approach, as it takes into account both the real and imaginary components of the raw signal, leading to a more detailed feature extraction. The wavelet transform of a 1-D signal provides a decomposition of the time-domain sequence at different scales, which are inversely related to their frequency contents (Mallat, 1989; Godfrey et al., 2009). This requires the time-domain signal under investigation to be convolved with a time-domain function known as "mother wavelet". The CWT applies the wavelet function at different scales with continuous time-shift of the mother wavelet over the input signal. As a consequence, it helps represent the EDA signals at different levels of resolution. For instance, it results in large coefficients in the transform domain when the wavelet function matches the input signal, providing a multiscale representation of the EDA signal. Using a finite energy function Ψ(t) concentrated in the time domain, the CWT of a signal x(t) is given by X(α,b) as follows (Vetterli, & Herley, 1992):

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

where, α, is the scale factor and represents dilation or contraction of the wavelet function and b is the translation parameter that slides this function on the time-domain sequence under analysis. Therefore, Ψ(α,b) is the scaled and translated version of the corresponding mother wavelet. "*" is the conjugation operator. Note that the wavelet coefficients obtained from Eq. (1) essentially evaluate the correlation between the signal x(t) and the wavelet function used at different translations and scales. This implies that the wavelet coefficients calculated over a range of scales and translations can be combined to reconstruct the original signal as follows:

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

B. Wavelet-Based Feature Extraction The time-frequency analysis of various bio-signals has been addressed in many related literature (Golshan et al., 2016;

Golshan et al., 2017; Zhang et al., 2009). It has been shown that the wavelet-domain feature space can improve the recognition performance of different human activities using the signals emanated from the body responses. Therefore, it essentially enhances the classification performance due to the more distinctive feature space provided. In this paper, we focus on the time-frequency analysis of the EDA signal to provide a new feature space based on which emotion classification task can be done. As opposed to some related studies that employ the raw time-domain signals for classification purposes (Greco et al., 2017; Jang et al., 2012), we use the amplitude of the CWT of the EDA signals to generate the features and drive the classifier. Working in the wavelet-domain is essentially advantageous since the wavelet transform probes the given signal at different scales, extracting more information for other post-processing steps. In addition, the localized support of the wavelet functions enables CWT-based analysis to match to the local variations of the input time sequence (Vetterli, & Herley, 1992). As a result, a more detailed representation of the signal is provided in comparison with the raw time-domain signal.

Figure 3.2 shows the amplitude of the CWT of a sample EDA signal at different scales using complex Morlet (C-Morlet) wavelet function. As can be seen, due to the localization property of the CWT, different structures of the input signal are extracted at each level of decomposition, providing useful information for analyzing the recorded EDA signals. In this work, we have 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 3.3 shows the wavelet-based feature extraction. Note that the impact of different families of the wavelet functions (e.g., Symlets, Daubechies, Coiflets) on the emotion classification will be evaluated in 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:

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

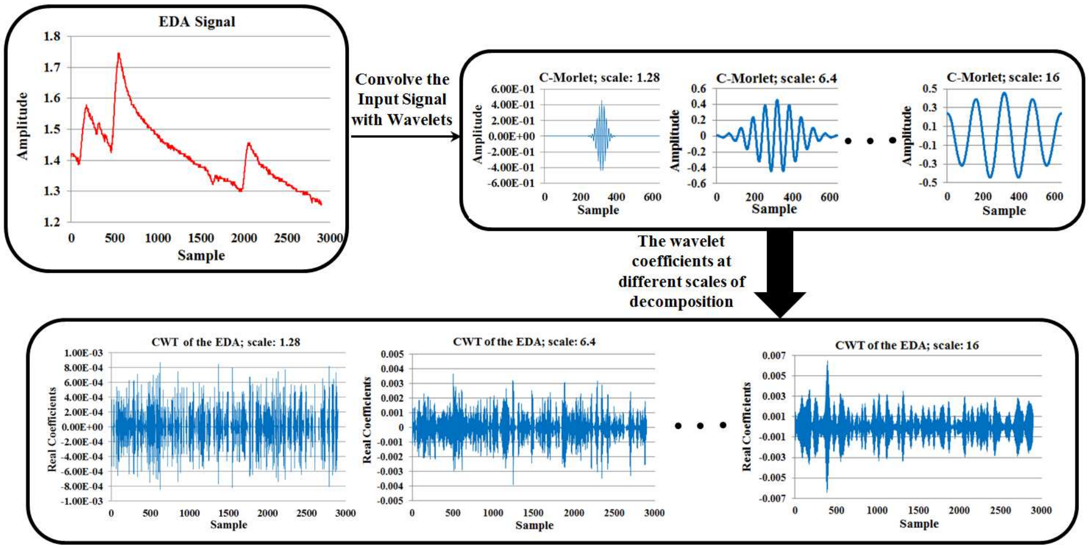


Figure 3.2: *The CWT of a typical EDA signal using the C-Morlet mother wavelet. Different scales of the wavelet functions are convolved with the original EDA signal to highlight different features of the raw data. As can be seen inside the bottom box, when the scaling parameter of the wavelet function increases, the larger features of the input signal are augmented. On the other hand, the detailed structures of the signal are better extracted when the scaling factor decreases.*

C. Support Vector Machine The SVM classifier tends to separate data $D = \{x_i, y_i\}_{i=1}^{N}, x_i \in^{d}, y_i \in \{-1, +1\}$ by drawing an optimal hyperplane <w,x>+b=0 between classes such that the margin between them becomes maximum (Cortes, & Vapnik, 1995). With reference to Figure 3.4, H1 and H2 are the supporting planes and the optimal hyperplane (OH) splits this margin such that it stands at the same distance from each supporting hyperplane. This implies that the margin between H1 and H2 is equal to 2 / ‖w‖. In terms of linearly separable classes, the classifier is obtained by maximizing the margin 2 / ‖w‖, which is equivalent to minimizing ‖w‖ / 2 with a constraint in convex quadratic programming (QP) as follows:

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

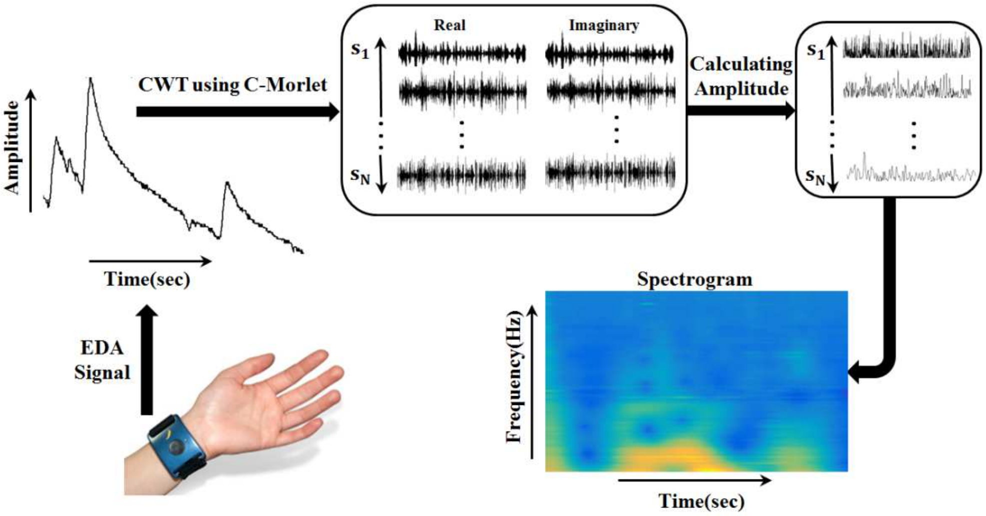


Figure 3.3: *The wavelet-based feature extraction. Using the C-Morlet mother wavelet, the real and imaginary wavelet coefficients are calculated at different scales. Then the amplitude of these coefficients is calculated to provide the corresponding spectrogram. This spectrogram is then used as the feature space.*

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

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

Usually, various kernel functions are used to deal with the nonlinearly separable data. As a result, the original data xi is mapped onto another feature space through a projection function $\varphi(\cdot)$.It is not necessary to exactly know the equation of the projection $\varphi(\cdot)$, but one can use a kernel function $k(x_i, x_j) =< \varphi(x_i), \varphi(x_j) >$. This function is symmetric and satisfies the Mercer's conditions. The Mercer's conditions determine if a candidate kernel is actually an inner-product kernel. Let $k(x_i, x_j)$ be

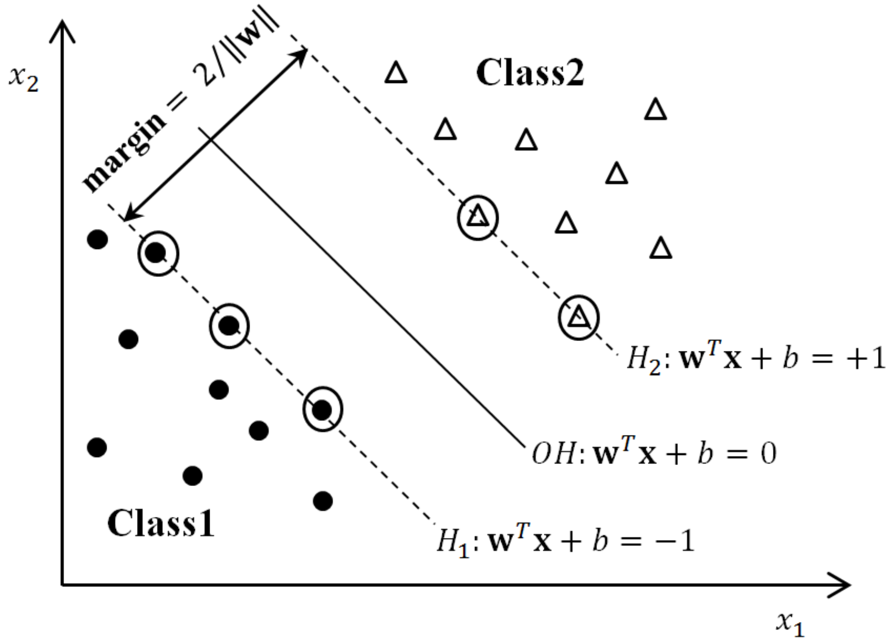

x_2
margin = 2/||w||
Class2
H_2: $\mathbf{w}^T\mathbf{x} + b = +1$
OH: $\mathbf{w}^T\mathbf{x} + b = 0$
H_1: $\mathbf{w}^T\mathbf{x} + b = -1$
Class1
x_1

Figure 3.4: *Canonical SVM for classifying two linearly separable classes. The decision boundary is shown by OH. Two hyperplanes H1 and H2 pass the support vectors that are circled inside the figure.*

a continuous symmetric kernel defined in the closed interval $t_1 \leq t \leq t_2$, the kernel can be expanded into series $\Sigma_(n = 1)^{\infty} = \lambda_n \varphi_n(x_i)\varphi_n(x_j)$, where $\lambda_n > 0$ are called eigenvalues and functions φn are called eigenvectors in the expansion. The fact that all the eigenvalues are nonnegative means that the kernel is positive semidefinite (Cortes, & Vapnik, 1995). To maximize the margin, H1 and H2 are pushed apart until they reach the support vectors on which the solution depends. To solve this optimization problem, the Lagrangian dual of Eq. (5) is used as follows:

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

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

3.3.4 Dialog System

Speech Recognition

http://doc.aldebaran.com/2-1/naoqi/audio/alspeechrecognition.html

Dynamic Oral Feedback

reason to design the dynamic feedback.

Chapter 4

X-Elophone: A New Instrument with New Experiment Design

In this section, a novelty instrument will be described in the following section.

reason why need this design. Due to the limitation of keys. This provides more possibility for different timber and major minor keys. That allows this system to play more customized song which kids love.

4.1 Xylophone Modification

4.1.1 Components Selection

A. Piezo Vibration Sensor: The LDT0-028K is a flexible component comprising a 28 μm thick piezoelectric PVDF polymer film with screen-printed Ag-ink electrodes, laminated to a 0.125 mm polyester substrate, and fitted with two crimped contacts. As the piezo film is displaced from the mechanical neutral axis, bending creates very high strain within the piezopolymer and therefore high voltages are generated. When the assembly is deflected by direct contact, the device acts as a flexible "switch", and the generated output is sufficient to trigger MOSFET or CMOS stages directly. If the assembly is supported by its contacts and left to vibrate "in free space" (with the inertia of the clamped/free beam creating bending stress), the device will behave as an accelerometer or vibration sensor. Adding mass, or altering the free length of the element by clamping, can change the resonant frequency and sensitivity of the sensor to suit specific applications. Multi-axis response can be achieved by positioning the mass off center. The LDTM-028K is a vibration sensor where the sensing element comprises a cantilever beam loaded by an additional mass to offer high sensitivity at low frequencies.

B. Op-Amp: An operational amplifier (often op-amp or opamp) is a DC-coupled high-gain electronic voltage amplifier with a differential input and, usually, a single-ended output.[1] In this configuration, an op-amp produces an output potential (relative to circuit ground) that is typically hundreds of thousands of times larger than the potential difference between its input terminals. Operational amplifiers had their origins in analog computers, where they were used to perform mathematical operations in many linear, non-linear, and frequency-dependent circuits. The popularity of the op-amp as a building block in analog circuits is due to its versatility. By using negative feedback, the characteristics of an op-amp circuit, its gain, input and output impedance, bandwidth etc. are determined by external components and have little dependence on temperature coefficients or engineering tolerance in the op-amp itself. Op-amps are among the most widely used electronic devices today, being used in a vast array of consumer, industrial, and scientific devices. Many standard IC op-amps cost only a few cents in moderate production volume; however, some integrated or hybrid operational amplifiers with special performance specifications may cost over US 100 in small quantities.[2] Op-amps may be packaged as components or used as elements of more complex integrated circuits. The op-amp is one type of differential amplifier. Other types of differential amplifier 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).

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.[1] 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.[2] 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.[1] 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.[1] An electronic multiplexer can be considered as a multiple-input, single-output switch, and a demultiplexer as a single-input, multiple-output switch.[3] 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.[4] The schematic on the right shows a 2-to-1 multiplexer on the left and an equivalent switch on the right. The *sel* sel wire connects the desired input to the output. The 74HC4051; 74HCT4051 is a single-pole octal-throw analog switch (SP8T) suitable for use in analog or digital 8:1 multiplexer/demultiplexer applications. The switch features three digital select inputs (S0, S1 and S2), eight independent inputs/outputs (Yn), a common input/output (Z) and a digital enable input (E). When E is HIGH, the switches are turned off. Inputs include clamp diodes. This enables the use of current limiting resistors to interface inputs to voltages in excess of VCC.

D. Arduino UNO: The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc.[2][3] The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits.[1] The board has 14 Digital pins, 6 Analog pins, and programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable.[4] It can be powered by the USB cable or by an external 9-volt battery, though it accepts voltages between 7 and 20 volts. It is also similar to the Arduino Nano and Leonardo.[5][6] 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.[1] The Uno board is the first in a series of USB-based Arduino boards,[3] and it and version 1.0 of the Arduino IDE were the reference versions of Arduino, now evolved to newer releases.[4] The ATmega328 on the board comes preprogrammed with a bootloader that allows uploading new code to it without the use of an external hardware programmer.[3]

While the Uno communicates using the original STK500 protocol,[1] it differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it uses the Atmega16U2 (Atmega8U2 up to version R2) programmed as a

USB-to-serial converter.[7]

show block diagram of the code:

4.1.2 Circuit Design

4.2 Sound Design

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

The computer has long been considered an extremely attractive tool for creating, manipulating, and analyzing sound. Its precision, possibilities for new timbres, and potential for fantastical automation make it a compelling platform for expression and experimentation - but only to the extent that we are able to express to the computer what to do, and how to do it. To this end, the programming language has perhaps served as the most general, and yet most precise and intimate interface between humans and computers. Furthermore, "domain-specific" languages can bring additional expressiveness, conciseness, and perhaps even different ways of thinking to their users. This thesis argues for the philosophy, design, and development of ChucK, a general-purpose programming language tailored for computer music. The goal is to create a language that is expressive and easy to write and read with respect to time and parallelism, and to provide a platform for precise audio synthesis/analysis and rapid experimentation in computer music. In particular, ChucK provides a syntax for representing information flow, a new time-based concurrent programming model that allows programmers to flexibly and precisely control the flow of time in code (we call this "strongly-timed"), and facilities to develop programs on-the-fly - as they run. A ChucKian approach to live coding as a new musical performance paradigm is also described. In turn, this motivates the Audicle, a specialized graphical environment designed to facilitate on-the-fly programming, to visualize and monitor ChucK programs in real-time, and to provide a platform for building highly customizable user interfaces. Show block diagram here.

Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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