12-2016

Psychophysiological analysis of a pedagogical agent and robotic peer for individuals with autism spectrum disorders.

Mohammad Nasser Saadatzi
University of Louisville

PSYCHOPHYSIOLOGICAL ANALYSIS OF A PEDAGOGICAL AGENT AND ROBOTIC PEER FOR INDIVIDUALS WITH AUTISM SPECTRUM DISORDERS

By

Mohammad Nasser Saadatzi

A Dissertation Submitted

to the Faculty of J. B. Speed School of Engineering

in Partial Fulfillment of Requirements

for the Degree of

Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering

J. B. Speed School

University of Louisville

Louisville, Kentucky

December 2016

PSYCHOPHYSIOLOGICAL ANALYSIS OF A PEDAGOGICAL AGENT AND ROBOTIC PEER FOR INDIVIDUALS WITH AUTISM SPECTRUM DISORDERS

By

Mohammad Nasser Saadatzi

A Dissertation Approved

on November 18, 2016

By the following Dissertation Committee:

_____

Dr. James H. Graham, Electrical and Computer Engineering Department
Dissertation Director

_____

Dr. Karla C. Welch, Electrical and Computer Engineering Department

_____

Dr. Robert C. Pennington, Special Education Department

_____

Dr. Tamer Inanc, Electrical and Computer Engineering Department

_____

Dr. Jeff L. Hieb, Engineering Fundamentals Department

DEDICATION

To My Beloved Wife

ACKNOWLEDGEMENT

ABSTRACT

PSYCHOPHYSIOLOGICAL ANALYSIS OF A PEDAGOGICAL AGENT AND ROBOTIC PEER FOR INDIVIDUALS WITH AUTISM SPECTRUM DISORDERS

Mohammad Nasser Saadatzi

November 18, 2016

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by ongoing problems in social interaction and communication, and engagement in repetitive behaviors. According to Centers for Disease Control and Prevention, an estimated 1 in 68 children in the United States has ASD. Mounting evidence shows that many of these individuals display an interest in social interaction with computers and robots and, in general, feel comfortable spending time in such environments. It is known that the subtlety and unpredictability of people's social behavior are intimidating and confusing for many individuals with ASD. Computerized learning environments and robots, however, prepare a predictable, dependable, and less complicated environment, where the interaction complexity can be adjusted so as to account for these individuals' needs.

The first phase of this dissertation presents an artificial-intelligence-based tutoring system which uses an interactive computer character as a pedagogical agent (PA) that simulates a human tutor teaching sight word reading to individuals with ASD. This phase examines the efficacy of an instructional package comprised

v

of an autonomous pedagogical agent, automatic speech recognition, and an evidence-based instructional procedure referred to as constant time delay (CTD). A concurrent multiple-baseline across-participants design is used to evaluate the efficacy of intervention. Additionally, post-treatment probes are conducted to assess maintenance and generalization. The results suggest that all three participants acquired and maintained new sight words and demonstrated generalized responding.

The second phase of this dissertation describes the augmentation of the tutoring system developed in the first phase with an autonomous humanoid robot which serves the instructional role of a peer for the student. In this tutoring paradigm, the robot adopts a peer metaphor, where its function is to act as a peer. With the introduction of the robotic peer (RP), the traditional dyadic interaction in tutoring systems is augmented to a novel triadic interaction in order to enhance the social richness of the tutoring system, and to facilitate learning through peer observation. This phase evaluates the feasibility and effects of using PA-delivered sight word instruction, based on a CTD procedure, within a small-group arrangement including a student with ASD and the robotic peer. A multiple-probe design across word sets, replicated across three participants, is used to evaluate the efficacy of intervention. The findings illustrate that all three participants acquired, maintained, and generalized all the words targeted for instruction. Furthermore, they learned a high percentage (94.44% on average) of the non-target words exclusively instructed to the RP. The data show that not only did the participants learn nontargeted words by observing the instruction to the RP but

they also acquired their target words more efficiently and with less errors by the addition of an observational component to the direct instruction.

The third and fourth phases of this dissertation focus on physiology-based modeling of the participants' affective experiences during naturalistic interaction with the developed tutoring system. While computers and robots have begun to co-exist with humans and cooperatively share various tasks; they are still deficient in interpreting and responding to humans as emotional beings. Wearable biosensors that can be used for computerized emotion recognition offer great potential for addressing this issue.

The third phase presents a Bluetooth-enabled eyewear – EmotiGO – for unobtrusive acquisition of a set of physiological signals, i.e., skin conductivity, photoplethysmography, and skin temperature, which can be used as autonomic readouts of emotions. EmotiGO is unobtrusive and sufficiently lightweight to be worn comfortably without interfering with the users' usual activities. This phase presents the architecture of the device and results from testing that verify its effectiveness against an FDA-approved system for physiological measurement.

The fourth and final phase attempts to model the students' engagement levels using their physiological signals collected with EmotiGO during naturalistic interaction with the tutoring system developed in the second phase. Several physiological indices are extracted from each of the signals. The students' engagement levels during the interaction with the tutoring system are rated by two trained coders using the video recordings of the instructional sessions. Supervised pattern recognition algorithms are subsequently used to map the physiological

indices to the engagement scores. The results indicate that the trained models are successful at classifying participants' engagement levels with the mean classification accuracy of 86.50%. These models are an important step toward an intelligent tutoring system that can dynamically adapt its pedagogical strategies to the affective needs of learners with ASD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by ongoing problems in social interaction and communication, and engagement in repetitive behaviors. According to Centers for Disease Control and Prevention, an estimated 1 in 68 children in the United States has ASD [1]. According to Centers for Disease Control and Prevention, an estimated 1 in 68 children in the United States has ASD [2]. Autistic children typically demonstrate delays in language development and impaired ability for imitation, imaginative play, and non-verbal communication [3]. Early intervention, however, has been proven to bring about a positive long-term outcome. Therefore, one of the challenges is identification and development of appropriate rehabilitation methods for ASD. To address this need, technological facilities, including computers and robots, have been utilized as rehabilitation/assistive devices for autistic children [4]. Mounting evidence shows that many of these individuals display an interest in social interaction with these machines and, in general, feel comfortable spending time in such environments [5, 6].

Articles [7, 8] have shown that an interactive robot is more engaging for children with ASD than is an inanimate toy showing repetitive behavior. It is known that the subtlety and unpredictability of people's social behavior are intimidating and confusing for these individuals [9, 10]. Computerized learning environments

and robots, however, prepare a predictable, dependable, and less complicated environment, where the interaction complexity can be adjusted so as to account for these individuals' needs, and can be planned according to their improvement. Additionally, they have the potential to furnish resilient systems to automate the time-consuming, repetitive behavioral therapy sessions.

Around one half of the ASD population does not develop any form of language, and the rest typically show delays in the onset and rate of development [11-13]. There have been several studies investigating the use and efficacy of computer-based instruction (CBI) of reading skills to individuals with ASD [14-20]. Although the results of these studies vary in terms of their positive gains, the overall results are quite favorable. They suggest that CBI may be highly motivating and effective for enhancing vocabulary acquisition for children with ASD. These studies also indicate that CBI typically results in benefits such as increased attention and enjoyment. It also has the potential to reach needy populations such as ASD individuals (ASDIs) at a broader scale because of the savings due to automation, resulting in a reduced demand for highly trained, costly therapists and allowing for broader dissemination of treatment, training, and education.

Advances in artificial intelligence (AI) have provided new avenues to facilitate human learning through interaction with animated virtual characters known as pedagogical agents [21-23]. Instructional applications that embed pedagogical agents have the capability to deliver instructional content while simulating social interaction. In PA–based applications, a student acquires target skills/behaviors through interaction with a PA. A number of studies have indicated

the positive instructional effects of PAs on cognitive and motivational factors [14, 24-28]. The results of these studies, albeit preliminary, represent important evidence that PA-based CBI is superior to less complex, auditory-alone CBI for language training. They suggest that students learn faster and remember better what they learn in the bimodal (voice and face) condition than in the unimodal (voice-only) condition [14, 27].

This dissertation research presents an AI-based tutoring system which uses an interactive computer character as a PA (Figure 1). This agent simulates a human tutor teaching sight word reading while demonstrating body and hand gestures, facial expression, and head nods. In order to facilitate learning through natural language, the agent delivers instructions through synthesized speech and receives learner's responses via microphone and automatic speech recognition.

Figure 1. Screen capture of PA, from left to right (a) PA gazing at the student, (b) PA pointing to the presented word, (c) PA clapping for the student

The most widely used instruction arrangement for individuals with ASD is direct instruction from an adult in a highly structured and well-planned format delivered in a one-to-one (1:1) student-to-teacher ratio [29-31]. In direct 1:1 intervention, the instructor delivers prompts to students to teach specified behaviors in a contained setting and isolated from peers, and the student responds to the prompts and is then differentially reinforced based on his response [32]. The

benefits of this type of intervention regarding behavior development and social functioning of children with ASD have been well-documented [29, 33-35].

This arrangement, however, consumes teachers' time [36-38] and restricts instructional and social integration of these children because the child is removed from peers for instruction [39]. Another major concern is that it ignores natural instruction environment and does not offer opportunities for student-to-student interaction. Relying on 1:1 format may leave these individuals with deficiencies in social interaction skills which hinders their functioning in inclusive environments and general education classrooms where group instruction is the norm [40, 41]. Besides, the skills learned through this arrangement often fail to generalize to other settings and their peers [42-44].

An alternative arrangement to 1:1 direct instruction is small-group instruction (SGI) during which several students, usually two to five, are taught in close physical and temporal proximity [45-47]. SGI enables more efficient use of teacher's time since more than one student is instructed at the same time [30, 41, 48-52]. SGI also better prepares students to function in less restrictive environments which frequently use group arrangements [50, 51]. SGI facilitates normalization as it more closely resembles regular instruction settings [45]. In SGI, students may learn appropriate student-to-student interaction and improve their social and behavioral skills as it provides a context with higher chances of interactions among peers [41, 53-58]. Furthermore, students receive multiple and varied forms of the target behavior within this arrangement, and the skills acquired may better generalize to different people [45, 54, 59, 60].

SGI provides a context in which a student can observationally learn a new response that is not directly instructed to him by observing teacher's instruction to another student, the student's response, and the subsequent teacher's differential reinforcement according to the student's correct or incorrect response. Therefore, simply by group participation, each student in the group has the opportunity to acquire skills targeted to other students even when he is not directly reinforced for his/her learning [61].

In this dissertation, a humanoid robot serves the instructional role of a peer for the student, augmenting the use of a PA. In this tutoring paradigm, the robot adopts a peer metaphor, where its function is to act as an emulated peer. With the introduction of a robotic peer, this dissertation suggests to augment the traditional dyadic interaction in tutoring systems to a novel triadic interaction (Figure 2), to enhance the social richness of the tutoring system, and to facilitate learning through observation.
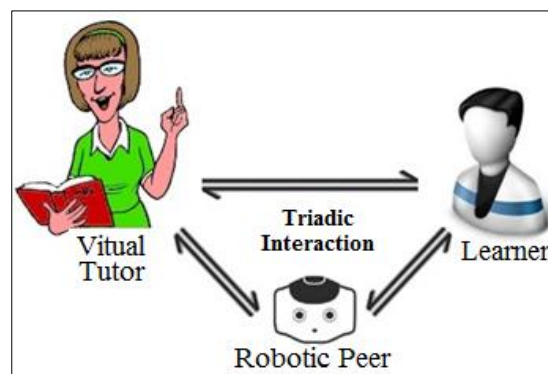
Figure 2. Triadic interaction among tutor, learner, and robotic peer

Expert human tutors are proficient at detecting the emotional state of students, and taking actions accordingly in order to optimize their learning. Recent decades have seen increased recognition of the interplay between affect and

learning [62-64]. It has been argued that learning is accompanied and influenced by a set of affective states such as boredom [65, 66], confusion [64, 67], flow/engagement [65], and frustration [64, 68]. For example, boredom negatively influences learning, whereas confusion and engagement in the learning process are positively correlated with learning [69].

Several recent tutoring systems, designed for typically developing individuals, are increasingly incorporating affect responsiveness into their pedagogical strategies [70-73]. These tutoring systems were reported to lead to better learning outcomes and to higher levels of engagement than their non-affect-responsive equivalent counterparts. To date, however, little effort has been made to investigate this approach and include affect responsiveness in technology-based intervention of individuals with ASD. Recent trends emphasize the need to understand heterogeneity in ASD from an emotional perspective. Mazefsky and colleagues [74] have argued for the benefits of integrating traditional ASD emotion research with emotion regulation frameworks more widely applied to normative populations.

In order for a tutoring system to respond to different affective states of learners, it must first provide a means to recognize those states. A number of modalities such as facial expression [75], voice prosody [76], body gestures [77], and physiology [78] have been investigated to assess affective states for typically developing individuals. However, ASDIs often have communicative difficulties regarding the expression of affective states [1] which limit traditional auditory and

visual approaches. Physiology, on the other hand, is not necessarily directly impacted by the impairments in emotional expressions in ASD [79-81].

Inspired by this idea, a wearable multi-sensorial biofeedback device, EmotiGO (Figure 3), is developed in this dissertation which is a pair of modified eyeglasses with three biosensors. The parameters measured are Galvanic Skin Response (GSR), Photoplethysmography (PPG), and Skin Temperature (SKT); a set of measurements useful for affect sensing. The collected signals are transmitted to a remote computer via Bluetooth Low Energy (BLE) communication in real time.

Figure 3. EmotiGO

Using the signals collected by EmotiGO during interaction with the developed tutoring system, a number of machine-learning-based computational models are trained in this dissertation that enable the tutoring system to track the affective experience of an individual with ASD based on his/her autonomic activity. This is an important step toward an intelligent tutoring system that can dynamically adapt its pedagogical strategies to the emotional needs of learners with ASD.

The remainder of this dissertation is organized as follows. **Chapter 2** presents the state of the art and a survey of relevant literature on computer-based

and robotic interventions for ASD. Chapter 2 also covers several important building blocks of social interaction such as joint attention, imitation, and turn taking addressed by robotic intervention of ASD to date.

Chapter 3 presents an instructional package which is comprised of an autonomous pedagogical agent, automatic speech recognition, and constant time delay (CTD). The pedagogical agent simulates a human tutor teaching sight word reading while demonstrating body and hand gestures, facial expression, and head postures. In order to facilitate learning through natural language, the agent delivers instructions by synthesized speech and receives learner's responses via microphone and automatic speech recognition. **Chapter 4** examines the effects of the instructional package during the instruction of reading sight words aloud to young adults with ASD. A concurrent multiple-baseline across-participants design is used to evaluate the efficacy of intervention. Additionally, post-treatment probes are conducted to assess maintenance and generalization. The results suggest that all three participants acquired and maintained new sight words and demonstrated generalized responding.

Chapter 5 introduces EmotiGO which is a Bluetooth-enabled eyewear for unobtrusive acquisition of a set of physiological signals, i.e., galvanic skin response, photoplethysmography, and skin temperature, which can be used as autonomic readouts of emotions. EmotiGO integrates multiple sensors in a single device which is lightweight and small enough to be worn comfortably without interfering with the users' usual activities. Thanks to its optimal low-power design, EmotiGO can be used for long-term in situ collection of PPG, SKT, and GSR

signals. Chapter 5 presents the architecture of the device and results from testing that verify its effectiveness against an FDA-approved system for measuring physiological signals.

Chapter 6 describes the augmentation of the AI-based tutoring system developed and evaluated in Chapters 3 and 4 with an autonomous humanoid robot that serves the instructional role of a peer for the student. With the introduction of the robotic peer, this dissertation augments the traditional dyadic interaction in the tutoring system to a novel triadic interaction, to prepare a technology-based small-group instruction arrangement and to facilitate observational learning. Chapter 6 evaluates the feasibility and effects of using PA-delivered sight word instruction, based on a CTD procedure, within a small-group arrangement. To evaluate the efficacy and efficiency of the learning modes, Chapter 6 employs a multiple-probe across-word-sets design replicated among three participants. The findings illustrate that all three participants acquire, maintain, and generalize all the target words they receive direct instruction for. Furthermore, they observationally acquire a high percentage (94.44% on average) of the non-target words exclusively instructed to the robotic peer.

Chapter 7 attempts to model the students' engagement levels using their physiological signals wirelessly collected with EmotiGO during naturalistic interaction with the tutoring system developed in Chapter 6. The physiological signals are first preprocessed in order to remove measurement noise and any motion-induced artifact in the signals. Several physiological indices are then extracted from each of the signals. The students' engagement levels during the

interaction with the tutoring system are rated by two trained coders using the video recording of the instructional sessions. The extracted features along with the engagement scores are subsequently input to two supervised pattern recognition algorithms, i.e., k-nearest neighbors and naïve Bayes. The results indicate that the trained classifiers are successful at classifying participants' engagement levels. Additionally, in order to identify the most diagnostic physiological signal, a systematic analysis is conducted on multichannel physiological data using two classifiers and three physiological signals as well as their fusion.

Finally, **Chapter 8** concludes this dissertation research, discusses the limitations of the developed systems, and outlines future research directions.

CHAPTER 2

LITERATURE REVIEW

This chapter is divided into two parts to cover background material supporting the technology-mediated instructional aspect of this research. Section 2.1 contains a survey about computer-based instruction for autistic individuals. Section 2.2 reviews and discusses justification of the use of robotics as an intervention method for ASD, as well as the most important social interaction forms addressed by this technology to date.

2.1 Computer-based Instruction

Computer-based instruction (CBI) is perhaps the most studied technology-based intervention for ASDIs, with a lengthy corpus of literature. To preserve readability, only studies related to the current dissertation, that is, vocabulary identification and reading skill, will be reviewed in the rest of this section. A variety of other skills, however, have been taught using CBI to these individuals, such as how to recognize other's emotions [82-84], enhance problem solving [85], and enhance vocal imitation [86].

In one of the first endeavors in this field by Heiman and colleagues in 1995, [17], the use of an interactive computer program to teach basic reading skill to 11 children with ASD (6:9 to 13:8 years of age[1]) was investigated. In the interactive computer program, reading and sentence syntax were instructed to the learners

---

[1] year: month

using voice, animated figures, and video. The program also provided immediate corrective feedback to the users. The data demonstrated that the participants displayed significant improvements in reading ability. Strong positive changes in verbal expression were also reported. That is, the children talked significantly more during the final session than in the first lesson. In the posttest, it was also observed that the participants expressed positive feelings and enjoyment almost 100% more than in the pretest.

In the study by Moore and Calvert in 2000, [19], the effects of CBI and teacher instruction on language acquisition of 14 children with ASD were compared. The computerized instruction included features such as sound effects and actions to attract the participants' attention. The CBI group was observed to spend more time on task than those in the teacher-instructed group. The CBI group also showed stronger desire to continue with the program after study completion than the teacher-instructed group.

Additional support for increased efficacy of CBI over more traditional methods was offered by [87]. In this study, development of reading skills in eight ASDIs was evaluated in two conditions; CBI and book-based instruction. The investigators found that at the end of CBI, five of the participants reliably identified at least three words when they failed to do so in the other condition. The participants were also reported to spend more time on and were less resistant to use reading material during CBI.

In another study by Bosseler and Massaro in 2003, [14], a three dimensional (3D) computer-animated talking head was evaluated on the improvement of

vocabulary in eight ASDIs (7-12 years old). In an initial experiment, significantly more words were identified by the participants at post-test than at pre-test. Besides, 85% of the new vocabulary words were correctly identified by the participants one month after program completion, suggesting that gains maintained over time. Additionally, high levels of enjoyment while working with the talking head were observed in seven of the participants.

In a follow-up study, [14], the same investigators evaluated the impact of the talking head program on language and skill generalization using a single-subject, multiple-baseline design across six of the eight participants from the first study. The participants were reported to learn faster and remember better the words they had learned in the voice-and-face condition than in the voice-only condition. The results of this study provide preliminary but important evidence that audio-visual CBI has superior impact over auditory-alone CBI in language training.

The study by Coleman-Martin and Heller in 2005, [16], reports an experiment on CBI conducted with three students one of which is a 12-year-old with ASD). This study included three conditions consisting of CBI only, teacher instruction only, and teacher plus computer-assisted instruction. The student with ASD reached criterion in each of the three conditions but had the fastest acquisition rate in the CBI only condition comparing to the other two conditions.

In the study by Hetzroni and Shalem in 2005, [18], six children with ASD were trained to identify words taken from commercial logos of food items. The effects of the program were evaluated using a multiple-probe design across participants. At the end of the study, all the participants were able to identify the

target words. Additionally, they successfully generalized their knowledge to classroom settings. The investigators, after study completion, collected probe data on a weekly basis to test identification accuracy over time. The data indicated that the participants maintained what they had learned over time. In a more recent study, [20], the effects of a computer-based early reading program on accurate reading were evaluated over four word sets by a nine-year-old child with ASD. The results, in the context of a multiple-baseline design across the word sets, indicated improved reading accuracy.

In a study on children with ASD [28], Tartaro and Cassell used an embodied software agent whose activities were manually controlled by an experimenter (Note – this is sometimes called Wizard-of-Oz approach). The experimenter observed the session from a different room and then chose the speech and gestures of the agent in an online manner, but manually, from a pre-recorded set of options. Two different conditions were studied, interaction with another child, or with the agent. The results indicated that the study participants increased their contingent discourse in the agent condition but failed to do so in the other one. It is noteworthy that this study is neither about sight words nor did it used any automated technology. The findings, however, suggest that embodied conversational agents can be useful tools for ASD intervention.

2.1.1 Discussion

The studies described in this section provide evidence for the potential of CBI to develop language and communication skills in ASDIs. This dissertation research takes advantage of the benefits and potential of CBI to teach sight words

to ASDIs. This follows studies such as [14, 27] which utilize embodied PAs, as they have demonstrated some superior gains for these individuals.

The novelty of this part of the dissertation, however, is the implementation of a computer application for teaching sight words reading which operates autonomously. This computer application delivers instructions through synthesized speech and receives learners' responses through automatic speech recognition. This autonomy can save teachers' instruction or preparation time while providing an engaging learning environment for autistic learners. It also enables learners to work on their reading materials on their own (even in the comfort of their home) while their reading performance is recorded in a computer file which can be accessed and analyzed by their teachers later on. The detailed information about this PA-based tutoring system will be provided in Chapter 4.

2.2 Robotic Intervention

The origin of robotic endeavors in ASD therapy can be traced to the 1970s through a seminal work done by Weir and Emanuel in [88], where an autistic boy was introduced to a remote-controlled turtle. However, it was not until the late 1990s and early 2000s, as ASD rates continued to escalate, that multiple labs adopted this topic for research. A socially interactive robotic system used in the field of ASD intervention is Robota doll (Figure 4). Robota dolls are a family of mini humanoid robots that can engage in interactions with humans, involving speech, vision, and body gesture [89]. They have existed since 1997 and were developed in order to investigate how a human can interact with a robot. The main body of the doll that contains the electronic boards and the motors is the plastic component

of a commercially available doll. The robot can react to touch by detecting passive motion of its limbs and head.

Keepon (Figure 4) is a simple creature-like robot, which is capable of expressing its attention by orienting its face to a certain target, and expressing emotional states by rocking its body from side to side and by bobbing its body up and down. They are reported to prompt spontaneous play in children with developmental disorders, and to develop social communication of these individuals with the robot and a co-present adult [90].
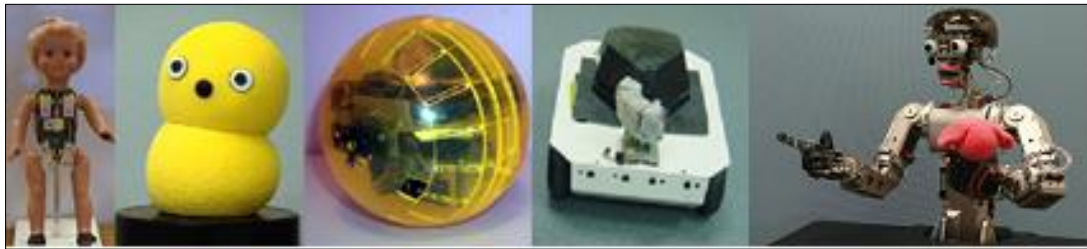
Figure 4. From left to right: Robota, Keepon, Roball, Labo-1, and Infanoid (Images taken from [4, 91-94], respectively)

Roball (Figure 4) is a spherical mobile robot that can generate various interplay situations with children in playroom environments. It autonomously wanders around in the environment and tries to initiate interaction with the child through some vocalizations [93]. Roball and a set of other robotic toys were designed to study effectiveness of different interactive capabilities of robots on engagement of children with ASD in playful interactions.

Labo-1 (Figure 4) is a non-humanoid mobile robot which consists of infrared sensors for obstacle avoidance and pyro-electric sensors, using which it can engage autistic children in chase-and-escape plays. Also, by means of a voice

production device, it can make utterances in certain situations and communicate with the children [4, 8, 95, 96].

Infanoid is a robot which was designed as a possible naturalistic embodiment for human-robot interaction (Figure 4) Infanoid possesses a complex mechanical design and a humanoid physical appearance. This robot has been used in contingency-detection games, where the robot engages in social interaction with autistic children, reacting to any social cues, as well as trying to elicit these from children by producing social cues [94].

KASPAR (Figure 5) is a minimally expressive child-sized humanoid robot. KASPAR's head, arms, and face can perform different movements, while its legs do not move. The face is made from a silicon rubber mask that covers a metal frame containing electromechanical parts. The eye lids and mouth can open and close. KASPAR is able to show minimal bodily expressions (e.g. waving or drumming a toy tambourine), facial expressions (e.g. happy, neutral, sad, and surprise), and gestures to interact with a human [97]. Case study evaluations suggest that KASPAR can encourage children to engage in social interactions with peers or adults [98-101].

Bandit (Figure 5) an upper-torso humanoid with two arms, a pan-tilt neck, an expressive face, and a mobile base has been used during free-play observations with children with ASD [102]. The robot moves autonomously around the room, and is able to make gestures, non-verbal vocalizations, and blow bubbles. The behavior of the robot is programmed to foster social interaction and communication of autistic children by turning the head, making

encouraging/disappointed utterances, blowing bubbles, moving towards the child, and rotating in place [103, 104].

IROMEC (Figure 5) is composed of a mobile platform, an application module, and a number of additional components that modify the appearance and behavior of the robot. The body has a main digital screen that displays graphical interface elements. The head has a smaller digital display that shows basic facial expressions like happiness and fear. The robot can engage in a number of play scenarios that are developed according to the needs of the target users group [98, 99]. Also, NAO, the humanoid robot used in this dissertation as the robotic peer, is reported to have the capability to engage autistic children in interactive behavior through speech, body movements, and touch. NAO has been recently used within researches in ASD [105].

Figure 5. From left to right: KASPAR, Bandit, IROMEC, and AIBO (Images taken from [98, 101, 103, 106], respectively)

AIBO (Figure 5) is a robotic dog widely used for education and research purposes. The effectiveness of using AIBO to improve social skills of eleven children with ASD was evaluated in [106]. The examiner compared the effects of playing with AIBO versus a toy dog. The results revealed that the robotic dog stimulated longer play interactions and more spoken communication. Furthermore,

the results showed that dyadic interaction (between child and object) and authentic triadic interaction (between child, object, and examiner) occurred more frequently in the AIBO condition compared to the toy condition.

The researchers investigating robotic intervention for ASDIs envisage design and construction of systems that can guide these individuals from very simple forms of interaction towards more complicated ones, as those in social human-human interactions, by slowly increasing the system's behavior repertoire as well as its complexity, duration, frequency, and unpredictability. At first, in a very simple context, the system demonstrates a small set of behavior plans, specifically for those individuals with intense deficiencies. After a while, in a gradual basis, the system features a larger behavior repertoire and plans its behavior, based on an agenda that is therapeutically relevant. The most important forms of social interaction are joint attention, imitation, and turn taking. In the following, the literature in the field of robotic intervention for ASD is reviewed according to these interaction forms.

2.2.1 Joint Attention

Joint attention (JA) is defined as the shared action of two people looking at the same object or pointing at it. JA skills are deemed to be the most essential building block of social communication [107]. In social life, the competency to participate in activities needing shared attention is a fundamental element to understand other people and to cooperate upon this understanding [108]. These skills appear between 9 and 18 months of age in typically developing children. Impairment in JA skills, however, is one of the first symptoms observed in children

with ASD [96, 109, 110]. JA deficits have serious consequences in development and learning language in these individuals [107].

Due to their inherent attractiveness and salience, robots or computer characters can be served as foci for JA, catalyzing interaction between autistic child and a co-present adult (e.g. teacher, parent) or child (e.g., peer, typical child) [91, 111]. Keepon has been utilized to improve basic JA skills in autistic children through certain play scenarios. Keepon, which is controlled manually (i.e., WOZ mode), alternates its gaze between a child's face, the caregiver's face, and sometimes a close toy. It produces a positive emotional response by bobbing its body and producing a happy-sounding utterance in reaction to any positive activity by the child, e.g., eye contact, touch, or vocalization. During the trials, autistic children were able to approach Keepon and gradually establish physical and social contact with it. Some of the children, furthermore, extended their dyadic interactions with the robotic toy into triadic ones by including a co-present adult into the interactional context [90, 94, 112].

In the study by Robins et al. in 2005, [91], JA skills in triadic interactions involving Robota, a child, and a second person were studied. The children were brought into the room two at a time, to participate in approximately five-minute sessions. The trials took place over several months and were designed to progressively move from very simple to more complex exposure to the robot. During the later trials, the investigator verbally encouraged the children to show each other how they could interact with the robot. In one play scenario, Robota mediated physical contact between the children, which is very unusual amongst

children with autism. In another play scenario, to improve the combination of JA and imitation (IM) skills, in addition to moving according to a request, the children sometimes positioned themselves appropriately even when not asked to do so. Sometimes, as a response to the robot's movement, the children simultaneously imitated the actions of Robota. The robot encouraged a full body experience for the children, stretching themselves, and exploring their own balances, while interacting with each other.

In another study by the same research group, [113], Robota operated as a preprogrammed dancing toy. In this mode, it moved its arms, legs, and head to the beat of pre-recorded music. Three autistic children, 5-10 years old, participated in nine three-minute-long trials during a period of 12 weeks. The children were brought to the room one at a time in the presence of the investigator. During the trials, the investigator did not initiate communication or interaction with the child, but did respond when addressed by the child. Results highlighted different ways where the robot provided a context in which the autistic children displayed embodied triadic JA interaction.

The study by Robins and Dautenhahn reported in [114] provides a case study where Robota mediated both indirect and direct triadic interactions between three children with autism and the experimenter. The robot was connected to a laptop and operated in WOZ mode, i.e. as a tele-operated puppet. In this mode, the investigator is the puppeteer and moves the robot's arms, legs, or head by a simple press of buttons on the laptop. The children engaged in an imitation (IM) game with the robot wherein the investigator mirrored their movements with the

robot. After a few turns of correct IM, the experimenter introduced a mistake in the robot's IM, which caused the children to giggle and talk to the robot with affection that this is wrong. When the children realized that the experimenter was operating the robot, it became a game between the experimenter and them. The children were knowingly sharing their enjoyment with the experimenter in JA framework.

Three children who participated in a study by Robins et al. reported in [100] suffered from severe ASD and verbal disabilities. The robot used in this study was KASPAR, operated via a wireless remote control. The trials lasted over several months for two of the children, but took place as a one-shot session for the third. The first child was reported to stretch out her hand, reaching for the experimenter's hand. The second child, in later stages of the study, started to touch and explore his own eyes, KASPAR's eyes and eye lids, as well as his teacher's eyes and face. The third child also started to share his excitement with his teacher, reaching out to her, and non-verbally asking her to join in the game. During the games, he learned to look at his therapist to see how she imitated KASPAR. Finally, the children succeeded in using the robot as an item of shared attention in order to play simple IM games together.

2.2.2 Imitation

Imitation, ranging from vocal responses for learning language to bodily mimics for learning movements and other physical activities, plays an essential role in child development and adulthood. It is a powerful means of signaling interest in another person, used for purposes of communication and interaction [115]. However, generally, autistic individuals are less capable of imitating actions and gestures [116]. The research carried out in [117] shows significant correlation

between IM and positive social behavior. Furthermore, [118] suggests that autistic children improve their social responsiveness, when they are being imitated.

In autism research, several studies have investigated the use of interactive technologies in teaching IM skills to children with ASD. In [4], Labo-1 is utilized by Dautenhahn and Werry to engage such individuals in simple, imitative interaction games. Four children who participated in the study were between 8–12 years of age, and included non-verbal children. The children played chasing game with the robot. They reached out and touched the robot, thereby making it approach them. When the robot was very close, the child took several steps backwards and again waited for the robot to follow them. Through these simple IM games, the interaction levels of the children with the robot are reported to significantly increase.

The research by Robins and Dautenhahn reported in [7] provides a case study wherein four autistic children, 5-10 years of age, participated in an average of nine trials each. The average duration of trials was approximately three minutes. By moving their limbs, the investigator showed the children how Robota could imitate their movements. The children, by themselves, could then continue the interaction with the robot, which was operating in the puppet mode hidden from them. In the last couple of trials, the children were not given any instructions or encouragement to interact with the robot, and were left to interact and play IM games on their own initiative.

In the study by Stribling et al. reported in [119], a high-functioning child with ASD was encouraged to interact with a robot that imitated the child's behavior. Although it was not the major focus of the study, the investigators collected a

sufficient number of samples of perseverative speech from the child-robot interaction to conduct conversational analysis on the interchanges. The preliminary data suggested that robot-child interactions might be useful for eliciting characteristic behaviors such as perseverative speech.

The study by Iacono et al. reported in [98] was carried out over six months with ten autistic children, with the average age of eight years. The robots used in this study are IROMEC and KASPAR. Each child performed the same play scenarios with each robot separately. The length of each session was approximately 20 minutes. The movements in interaction with KASPAR, which was operated by the investigator via remote control, consisted of touching the face, opening and closing the eye lids, and raising the arms in sequence or together. Also the experimenter sometimes gave the remote control to the child; in these cases, the child chose which of KASPAR's behaviors she would like to imitate. During the IM game with IROMEC the child was asked to look at IROMEC's movements and to imitate them. IROMEC performed a sequence of movements of traveling right, left, and straight. The scenarios based on IM seemed to work better with KASPAR than with IROMEC.

Two other studies found increased imitation speed to robot models in comparison to human models [120, 121]. The study by Bird et al, [120], reported a speed advantage in adults with ASD imitating robotic hand movements over human hand movements. In a rather similar study, [121], Pierno and colleagues found that children with ASD had significantly faster movements to grasp a ball when they saw a robotic arm perform the movement first than when they viewed a

human arm first. These studies suggest that ASDIs might benefit from tasks that involve imitating robots in comparison to imitating humans.

The study by Duquette et al. reported in [122] examined the use of a humanoid robot to help a group of four children with ASD practice imitation behaviors in a series of intervention sessions. In this experiment a human mediator or the humanoid robot performed role of a partner. The interaction partner performed a behavior, asked the child to imitate the behavior, and if the child was successful, provided positive reinforcement by raising arms and positive auditory feedback. Proximity and interactions were found more frequently in children paired with the robot mediator than in children paired with the human mediator.

2.2.3 Turn Taking

The main educational objectives addressed with turn-taking (TT) scenarios in ASD therapy are improvement of spatial and body awareness, sense of self, creativity, leadership and the taking of initiative in autistic children. Moreover, it has been suggested that TT and IM games allow infants to identify people separate from other objects, and distinguish between different persons [123].

The study done by Wainer and Dautenhahn reported in [101] describes an experiment in which six children with ASD played a dyadic video game alternatingly with an adult and KASPAR, in four 25-minute sessions over a period of three weeks. The two players had to stand on opposite sides of a single flatbed computer monitor, while taking turns and communicating with each other, either verbally or by pressing buttons, to decide which of the many shapes on the screen both of them would select. Each player would then use their handheld wireless controller to move one of two orthogonal lines about on the screen, making their

line intersect with the agreed-on shape. The players would then have to collaborate with each other to synchronize pressing the buttons on their controllers. When this was done successfully, the players were briefly rewarded by their controllers vibrating, a pleasant sound playing from nearby speakers, and the selected shape flashing and spinning around. During this experiment, it is observed that the children's actively collaborative behaviors between the first and second sessions of playing with the human partner increases significantly.

In the study by Iacono et al. reported in [98], several TT scenarios were conducted with IROMEC and KASPAR, in addition to IM games discussed in the previous subsection. In the first scenario, TT was carried out using IROMEC. While playing with the moving robot, the two players sat on the floor or stood a certain distance away from each other. The first player turned IROMEC toward the second player and touched the digital fur on its main visual display causing the robot to move forward. When IROMEC reached the second player it stopped and gave him/her an appropriate feedback (visual and sound). The second player then turned the robot towards the first player and sent the robot back to the first player. This procedure was repeated multiple times.

The second TT scenario in the study by Nadel et al, [117], was performed with IROMEC stationary while the two players sat on the floor close to it. During the game two different colors (red and yellow) appeared on IROMEC body display as two sets of ladybugs. There were two buttons (red and yellow), one for each player. The goal of the game was to summon animated digital ladybugs onto the body display and make them flap their wings by taking turns in pressing buttons.

Each player selected a different button, and pressed this button once during their turn. When a button was pressed a ladybug of the same color appeared and began to flap its wings. The game continued until three yellow ladybugs and three red ladybugs appeared on IROMEC's back. At the end of the game an audio reward (short melody) and a visual reward (flying ladybugs) were given.

The third TT scenario carried out with KASPAR, in the study by Michaud et al. in [93], is a collaborative play activity based on playing a drum. In this scenario, two identical drums were used, one placed on KASPAR's lap and the second given to the child. Both the child and the experimenter sat in front of KASPAR. The experimenter operated KASPAR with the remote control device. The child was asked in their turn to play their drum in the same way KASPAR had played before. During the scenario the experimenter sometimes asked the child to switch roles with KASPAR and to show it how to play the drum. In this case, the child was the one who initiated the drumming for the robot to imitate. Furthermore, the experimenter sometimes gave the remote control to the child to operate KASPAR. For some children, waiting for their turn to control the robot was an important incentive to participate in this TT activity.

The study by Lehmann et al. reported in [99] describes the use of KASPAR and IROMEC in a six-month-long study with autistic children. In this study the two players were a child and an adult. The objectives covered the improvement of the awareness of one's own body, spatial awareness, the understanding of cause and effect, the improvement of the ability to focus on a single task, the ability to establish JA, and the ability to understand and apply the rules of interaction. This

scenario with IROMEC consisted of clapping the hands to make IROMEC move around in the space. By clapping hands once it goes straight. To change the direction of IROMEC, left or right, or to keep the robot moving, the child has to clap their hands two or three times. This play scenario is played in the WOZ modality. A second experimenter operates the robot's movements using a button.

Also in the same study, [99], a TT scenario with KASPAR was examined wherein the goal was to clap hands to raise KASPAR's arms. The children were asked to clap their hands once, twice, or three times for KASPAR to lift up its right arm, left arm, and both arms respectively. The experimenter controlled KASPAR's movements in the WOZ mode. In some cases, the experimenter gave the remote control to the child. The experimenter instructed the children to do the same movements KASPAR did or to clap their hands in order to reinforce the feedback.

In the study by Kozima et al. reported in [92], a series of longitudinal observations were made to examine how autistic children interacted with Keepon in TT as well as IM scenarios. Keepon was placed in a playroom at a day-care center where autistic children, 2.4 years old on average, their parents, and nursing staff interact with each other. In the playroom, Keepon was one of the toys scattered over the floor. During the remedial session of about three hours, the children played with Keepon whenever they wished. This study observed interactions of autistic children with Keepon during 500 sessions, over 18 months of study. This study reports emergence of dyadic social interactions with Keepon that gradually converted to triadic forms with their parents as well as other children

after several months. These social interactions emerged through unconstrained individual or organized group activities of touching and playing with Keepon.

2.2.4 Discussion

The studies described in this section suggest that robots have the capacity to play important roles as therapeutic tools for children with ASD. They can enhance their quality of life and social integration, and bear the potential to furnish real-time, multi-modal, and embodied social interactions. Interaction with a robot lies between the complexity of interacting with a software agent and that of interacting with a real person. Robots foster interaction dynamics that closely reflect the real time nature of human-human interaction and naturally support multi-modal interaction.

This dissertation research studies the effects of using a humanoid robot in facilitating reading sight words for individuals with ASD. If ASDIs show preferences for robots, then one prediction would be that performance on tasks could improve if the same actions were modeled by a robot in comparison to a human. In this scheme, along with learner with ASD, a humanoid robot will attend the tutoring sessions as an RP. Learning through such triadic interaction requires the learner to apply and practice social interaction skills such as JA, IM, and TT. According to the literature reviewed in this section, robots can play important roles in fostering JA, IM, and TT skills in ASDIs.

For example, when RP is addressed by PA, it serves as a focus of JA between PA and the learner. Utilizing RP can catalyze interaction between the autistic learner and PA (because of its intrinsic attractiveness) and bring about higher engagement levels and time-on task which, in turn, result in higher learning

gains. Similarly, TT skill can be practiced when the learner and RP attend a learning session in which PA addresses them in turns. This approach also provides a medium in which the learner repeatedly mimics a behavior (IM) while actually learning how to read and identify words. The robot in this instance is directly active in modeling a specific behavior for the child.

RP can also provide feedback (reinforcement, etc.) or produce a positive emotional response through gesture and utterance to positive social or cognitive activities by the learner while involved in learning through interaction with PA. The use of a robot rather than a human to serve as a peer here fits with technology-oriented preferences of ASDIs.

CHAPTER 3

ONE-TO-ONE INSTRUCTIONAL ARRANGEMENT: AUTONOMOUS PEDAGOGOCAL AGENT

This chapter presents an AI-based tutoring system which uses an interactive computer character as a pedagogical agent (PA). This agent simulates a human tutor teaching sight word reading (SWR) while demonstrating body and hand gestures, facial expression, and head postures (i.e., nods and shakes). In order to facilitate interaction using natural language, the agent delivers instructions through synthesized speech and receives learner's responses via microphone and automatic speech recognition (ASR).

Reading ability is an essential component of language competency [124], and a fundamental skill for independent living [125, 126]. When learning to read, however, many individuals with ASD find difficulties that are attributed to specific cognitive deficits [127, 128]. Limited ability to interact during and lower time in literacy activities, as well as lack of self-confidence, attention, and motivation may also contribute to this deficiency [127, 129]. It may, as well, be due to ineffectiveness of conventional instruction methods for this population. ASD individuals (ASDIs) usually have deficits in phonemic awareness and word decoding which renders phonetic-based reading instruction less effective for these individuals [127, 130, 131].

Some researchers, hence, have advocated supplementing phonetic-based instruction with sight word instruction [132-134]. Sight words are frequent

31

vocabulary that one can read without any decoding strategies and phonetic analysis [135]. SWR can help these individuals survive and be safe within society [136], and improve their functioning with school- and work-related tasks [136-139]. SWR can also boost their confidence in reading ability, and improve their affective experience (e.g., by reducing anxiety) and, thereby, their attitude towards learning to read [132, 140, 141]. Sight word instruction for individuals with ASD has been reported effective in the literature [131, 142].

Many individuals with ASD find more comfort and joy in interacting with computers than with humans, and feel an affinity with them [5, 143, 144], suggesting that computer-based instruction (CBI) can be an effective intervention method for them. There have been several studies investigating the use and efficacy of CBI of language skills to these individuals [15-20]. Although the reported positive gains vary, the overall results are quite favorable. CBI can provide learning environments that are individualized, highly controlled and structured, and can include a variety of modalities, such as text, sound, and images. CBI enhances motivational and attentional factors, and increases engaged time and enjoyment in individuals with ASD [5, 87]. CBI also has the potential to reach needy populations, such as people on the autism spectrum, at a broader scale because of the savings due to automation.

In a recent endeavor in using computers to instruct SWR, Mechling et al. [145] used PowerPoint slides on a Smart Board with three participants with intellectual disability (ID). The successful results of this research strongly support effectiveness of computer-assisted instruction (CAI) of SWR to individuals with ID.

The proposed method, however, still involved the teacher in the process of instruction for advancing the PowerPoint slides, delivering task directions, verbally modeling the target words, as well as providing praise and feedback for their correct and incorrect responses, respectively. The researchers also mentioned that the program required the teacher to have some computer skills such as the use of PowerPoint and Smart Board.

Yaw et al. [142] used PowerPoint slides to teach SWR to a sixth-grader with ASD. In their proposed method, the target words appeared on the screen, one at a time, for two seconds. The student was to read the word before it disappeared and the word was modeled (pre-recorded using the experimenter's voice). The successful results of this study encourages the use of computers in SWR instruction to individuals with ASD. Coleman et al. [146] compared the efficiency of CAI and teacher-directed instruction of SWR to three students with ID. In the CAI condition, the words were taught through teacher-developed PowerPoint slides including the target words as well as pre-recorded audio (using the teacher's voice) providing task directions, controlling prompts, verbal models of the words, and non-contingent verbal praise at the completion of each session. CAI of SWR was proved effective, though just slightly less efficient, compared to the teacher-directed condition, in terms of number of trials to criterion.

The involvement of teachers in the proposed instruction methods of Yaw et al. [142] and Coleman et al. [146] is relatively less than that of Mechling et al. [145], although it comes at the expense of contingent reinforcement and feedback to students. Such reinforcement and feedback, however, play vital roles in instructing

individuals with ASD. With timely and accurate feedback, learners realize their mistakes and can correct them immediately [147]. A non-contingent response should reduce the cost-effectiveness of CBI against human instruction.

Recent advances in ASR can play a big role in automation of instruction to students with and without disabilities. ASR technology has the potential to furnish a learning environment where students read to the tutoring software and their reading performance is assessed automatically. ASR can provide a cost-efficient method to automate instruction, facilitate data collection, and expedite the turnaround period for receiving scores. ASR is not a new technology, and it has been used in navigation systems and customer service applications, just to name a few. In the field of special education, Raskind and Higgins [148, 149] compared the impact of general computer instruction against computer instruction augmented by ASR to students with Dyslexia. The results indicated significant improvements in word recognition, spelling, and reading comprehension.

Most current CBI software rely on keyboard and mouse usage for their users, while people with ASD often have low tactile sensory tolerance and poor development in both fine and gross motor skills. Incorporating ASR into CBI could address these issues and accommodate for sensory preferences, thereby catering to more users. In the current chapter, the author sets out to integrate ASR into the tutoring software that teaches SWR to students with ASD. The software evaluates the students' responses on the reading material in order to collect data on their performance as well as to provide contingent, immediate reinforcement and feedback, which are essential in treating children with ASD [150]. Besides, the

developed tutoring software talks back to the students via a text-to-speech (TTS) engine. The TTS engine enables the tutoring system to deliver the instructional tasks, controlling prompts, verbal praises, and feedback to the students using natural language.

The use of ASR and TTS in the educational software brings about an autonomous interaction that flows based on human terms. This autonomy can save teachers' instruction and preparation time while providing an engaging learning environment. It also enables students to practice independently while their performance is recorded in a computer file which can be accessed and analyzed by their teachers later on.

Social interaction, a deficient area in individuals with ASD, can be promoted while working on computers, provided that proper social arrangements are made [151]. Advances in computer graphics have opened up new avenues for human learning through interaction with pedagogical agents (PAs) rather than disembodied voices. PAs are virtual characters incorporated in learning environments [22, 152]. What makes PA-based learning environments unique from conventional ones is that they support multimodal human interaction. In PA-based environments, a student acquires the target skill while interacting with a PA equipped with social and emotional manifestations (e.g., gaze, joint attention, and facial expressions). PAs can provide this population with learning environments that are less socially intensive than human-based instruction [153, 154].

Positive effects of PAs on motivational, attentional, and cognitive factors in individuals with ASD have been demonstrated in the literature. Tartaro and Cassell

[28, 155] used a virtual agent, Sam (Figure 6), to enhance social skills, such as turn-taking, gaze behavior, and social discourse in children with ASD. Sam is a virtual agent whose activities were manually controlled by an experimenter via a control panel (i.e., Wizard of Oz approach). The experimenter observed the session from a different room and then chose the speech and gestures of the agent in an online manner from a pre-recorded set of options.
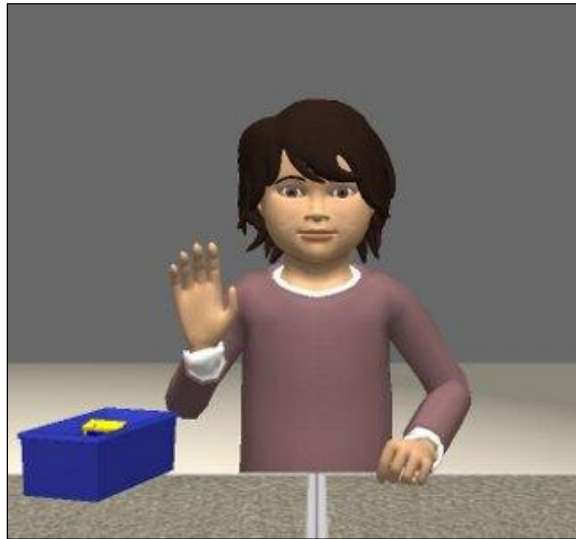
Figure 6. Screen shot of Sam educational environment (courtesy of Tartaro and Cassell [28, 155])

Two different conditions were studied for building collaborative narratives; interaction with another child or interaction with Sam. The authors reported that Sam was intuitive and motivating for the study participants. After interaction with Sam for an extended period of time, the participants scored higher on language tests, and displayed improvement in social behaviors. The results also indicated that the participants increased their contingent verbalization over the course of interaction in the agent condition but failed to do so in the other one. The participants were, more importantly, able to transfer the acquired skills to realistic settings and with their human peers. It is noteworthy that this study is neither about

reading instruction, nor did it use any automated technology. The findings, however, suggest that embodied virtual agents can be useful tools for ASD intervention.

Bosseler and Massaro [14, 27] examined a computer-animated talking head, Baldi (Figure 7), to instruct vocabulary recognition and grammar to children with ASD. In an initial experiment, significantly more words were recognized by the participants at post-assessment than at pre-assessment. Besides, 85% of the new words were correctly identified by the participants one month after program completion, suggesting that gains were maintained over time. Further, the participants displayed high levels of enjoyment while interacting with Baldi. In a follow-up study, the investigators evaluated the impact of Baldi on language and skill generalization. The study participants were reported to learn words faster and maintain them better while interacting with the tutoring software in his voice-plus-face mode than in the voice-only mode. The results of this investigation are preliminary but a promising indication that PAs may have superior impact over auditory-alone CBI in language training.
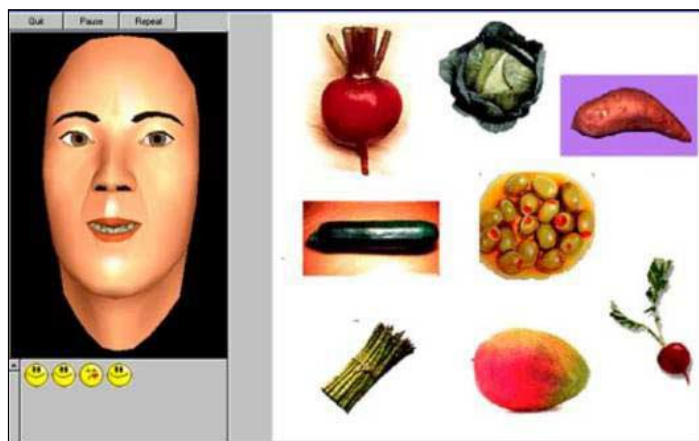
Figure 7. Screen shot of Baldi educational environment (courtesy of Bosseler and Massaro [14, 27])

Milne et al. [156] created another instance of PA-based educational software, Thinking Head, which has life-like agents with the capability to demonstrate facial expressions. They developed two modules for this agent platform (Figure 8); one to instruct how to deal with bullying and one for basic conversation skills. The autistic participants of this study were reported to find the agents enjoyable and useful, and their post-test scores on the knowledge of those two topics were significantly higher than the pre-test ones. They also found the interaction with the agents positive and non-threatening.
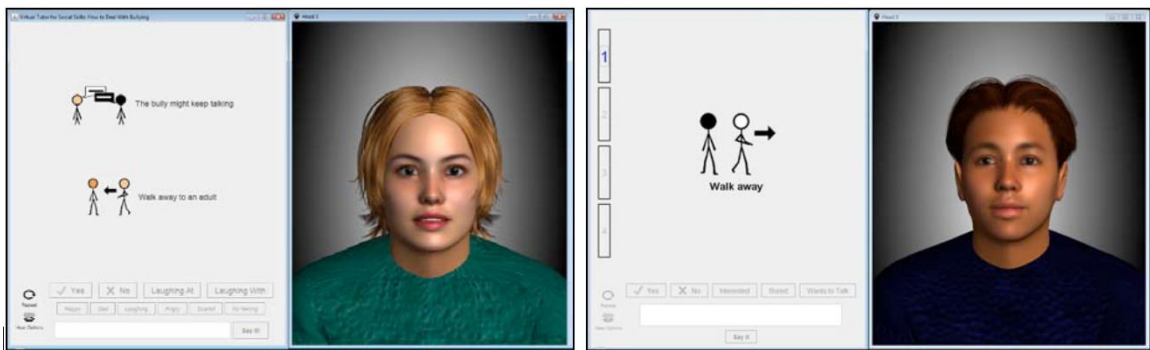
Figure 8. Screen shots of Thinking Head software (courtesy of Milne et al. [156])

Mower at al. [157] developed an interactive educational software, Rachel system (Figure 9), which included an agent along with emotionally evocative scenarios to elicit affective and social behavior in children with ASD. The agent which had a peer-like appearance was manually operated in the Wizard-of-Oz mode. The authors used the Rachel system to instruct the participants about emotional causes as well as to coach them through empathetic exercise. The findings of this study illustrated that the system was successful at eliciting socio-emotional interaction from children with ASD.

The results of these studies, albeit preliminary, provide important evidence that PA-based tutoring systems may be superior to auditory-alone systems. They

are effective in eliciting social and emotional behaviors in these individuals, and have the potential to provide engaging, motivating, and enjoyable learning environments for this population. Particularly exciting evidence in most of these studies is that they report skill generalization to novel contexts, an area of difficulty for ASD intervention.

Figure 9. Screen shot of Rachel system (courtesy of Mower et al. [157])

The results of these studies, albeit preliminary, provide important evidence that PA-based tutoring systems are superior to auditory-alone systems. They are effective in eliciting social and emotional behaviors in individuals with ASD, and have the potential to provide engaging, motivating, and enjoyable learning environments for this population. Particularly exciting evidence in most of these studies is that they report skill generalization to novel contexts, an area of difficulty for ASD intervention.

In the next section, the details of the developed educational software and the adopted instructional strategy are presented.

3.1 Instructional Strategy and Software

The developed tutoring system relies only on proven instructional strategies for individuals with ASD. Various instructional strategies have been examined by

researchers to find the most effective and efficient ones for this population, including time delay [158], system of least prompts [159, 160], and least-to-most prompting [161], among others. A procedure is considered efficient when it results in high amount of learning in a short amount of time and provides high density of reinforcement.

The developed tutoring system, embeds an evidence-based instruction procedure called constant-time delay (CTD), as it has been shown to be an effective and efficient strategy for this population [158]. Successful application of the CTD procedure in ASD intervention includes instruction of social phrases [162], social play skills [163], and sight words [164, 165]. Furthermore, the structured format of CTD lends itself nicely to be implemented as a computerized procedure. Previous research has reported successful implementation of computerized CTD to teach multiplication facts [166], grocery words [145], functional sight words [146], and Dolch sight words [142].

CTD involves stimulus presentation, task delivery, controlling prompt, reinforcement, and feedback. Initially, the stimulus and task are immediately followed by the controlling prompt to ensure correct response by the student. After several trials of this drill, the controlling prompts are systematically faded on a time dimension by introducing a constant delay before the prompts in order to allow the learner to emit the response on their own. The instructor delivers reinforcement and corrective feedback for correct and incorrect responses, respectively [167]. Reinforcement and feedback play important roles for learning the target skill in the CTD procedure.

The developed software which is a desktop virtual reality (VR) application integrates a PA, a virtual classroom, a TTS engine, and an ASR engine. The application was created in a commercially-available VR design package, namely Vizard from Worldviz [168]. The PA is a 3-D full-bodied agent, rather than a face-only, shoulder-up, or waist-up character, for two main reasons. First, a virtual agent in full body can use its arms and body orientation to model pointing gestures more easily than in other demonstrations. Pointing has a great potential for capturing attention and directing one's gaze to a location in space.

Individuals with ASD often exhibit extreme deficiency in gaze following and joint attention. Joint attention is the coordination of gaze among two people on an object. In responding to one's joint attentional bids, the respondent is required to follow the gaze direction of the initiator [107]. Alcorn et al. [169] studied how gaze following and joint attention can be best stimulated, if at all possible, from individuals with ASD by an embodied virtual agent. The study's participants were to pick a target flower, among three, that the agent indicated using its gaze with or without pointing gestures (Figure 10). The participants were successful in following the agent's joint attentional bids, and it was the conjunction of gaze and pointing cues that resulted in significantly higher joint attention elicitation compared to the single cues. The employed full-body PA uses the combination of pointing cues, eye gaze, and head orientation in order to better capture users' attention and help direct their gaze toward the reading material. The author believes that a full-bodied, animated agent with gestural cues can become a part of the teaching scenario, and better draws students' attention to text which is to be read.

Figure 10. Agent uses different joint attentional bids to direct participant's gaze (courtesy of Alcorn et al. [169])

Second, social distance, characterized by the distance between two people, plays an important role in their social interaction [170, 171]. While interacting with a PA, a close view of the agent's face and upper torso may simulate an invasive social distance, which is considered an intimate space [170, 172], and has been reported in the literature to elicit uncomfortable feelings, attempts to increase the distance [170], and anxiety in these individuals [173]. In this educational software, by using a full-body representation of the agent, it is intended to simulate a social distance that provides a more comfortable conversation and avoids inducing anxiety, as anxiety is not conducive to learning [64] and perhaps is a hallmark of ASD [174].

The developed tutoring software simulates a classroom setting with a chalk board on which the sight words are shown in written form, one at a time (Figure 1). A real-time behavior and character algorithm is implemented to produce verbal and expressive attributes of the agent. When the student answers correctly (captured by ASR), the PA nods its head, smiles, and provides verbal praise (using the TTS engine) in order to deliver non-verbal and verbal reinforcement. On the other hand,

it shakes its head and provides the correct answer when the learner is incorrect. When there is no answer given by the student, it merely models the word without any expressive movements.

To ensure robust speech recognition, an advanced, proprietary engine with high recognition accuracy was used. Since this is a feasibility study and has an exploratory nature, instead of purchasing a desktop license for an ASR engine, the author readily used the accessible engine within a NAO robot available at our laboratory. NAO, which is a humanoid robot from Softbank Robotics [175], has microphones, and comes with a built-in ASR engine from Nuance [176]. The educational software was programmed in a way that it communicates with NAO in real time during the instructional sessions over a wireless network, to receive the recognized speech by its engine from the participants. In instructional sessions, NAO was inconspicuously placed behind the participants, while they were interacting with the PA, in a way that their voices can be reliably captured by NAO's microphones without distracting them. The participants were led to believe that they were conversing with the PA, while the actual speech recognition was performed with the robot behind them.

Further, as the author intended the interactions to be very human-like, a TTS engine from Sitepal [177] was utilized. This engine produces synthesized speech highly similar to human speech, rather than a machine-like voice. A natural human voice ensures that words are modeled with correct pronunciations. Therefore, although the developed tutoring system can utilize Microsoft's TTS engine, the PA's speech (including prompts, words, verbal praises, etc.) was

43

recorded using Sitepal's trial version which is freely available online. The author particularly intended the speech to be synthesized, rather than by, for example, his voice, in order to investigate its effectiveness in such a context. Realistic tutoring software will eventually need to rely on a TTS engine so as to provide a completely hands-off method for teachers and parents.

In the following chapter, details of a case study with three autistic participants, using one-to-one computer-based sight word instruction are presented.

CHAPTER 4

CASE STUDY FOR ONE-TO-ONE SIGHT WORD READING INSTRUCTION TO STUDENTS WITH ASD USING AN AUTONOMOUS PEDAGOGICAL AGENT

The current chapter examines the effects of the instructional package developed in the previous chapter, comprised of an autonomous PA, ASR and TTS technology, and CTD during the instruction of reading sight words aloud to young adults with ASD. A concurrent multiple-baseline across-participants design was employed to evaluate the efficacy of intervention and conduct post-treatment probes to assess maintenance and generalization. The findings suggest that all three participants acquired and maintained new sight words and demonstrated generalized responding.

4.1 Method

4.1.1 Participants

Participants were recruited from a public school transition program on the university's campus. Consent forms were sent to parents of students with ASD based on clinical diagnosis and the teacher's recommendation. Criteria for selection were (a) being verbal, (b) ambulatory (as they had to commute from their classroom to the experiment room), and (c) being on the sight word reading level. Having some a priori sight word reading skill is necessary, as it assures that the person is not completely impaired in that aspect, and could potentially learn more words. Three students were permitted by their parents to participate. They were

all male, with age range of 19-20 years, demonstrated delays in all academic areas including literacy, and had previous general exposure to computers in their classroom and at home. Student J was a 20-year-old African-American male with Functional Mental Disability (FMD), Multiple disabilities (MD) of FMD, and Orthopedic Impairment. He demonstrated autism-like characteristics and his IQ was 53. Student A was a 20-year-old white male with ASD and MD (FMD and Attention Deficit Hyperactivity Disorder). His full scale IQ was 49. Student M was a 19- year-old African-American male with ASD and FMD. His full scale IQ was 40. All three participants demonstrated delays across all academic areas including literacy, and had previous general exposure to computers in their classroom and at home.

Before approaching the students and their teacher, approval from the university's institutional review board (IRB) as well as the county public schools' IRB were obtained.

4.1.2 Settings and Materials

Sessions with the educational software were conducted in a one-to-one instructional arrangement in an experiment room in our laboratory. The virtual classroom and the PA were displayed on a 26-inch computer screen. During the sessions, the participant sat on a chair and directly faced the screen, which was located on a table, with properly-adjusted height for his eye level. Words were displayed with a 90 pt. Arial style font and with high contrast to insure visibility. Words were presented in lower case letters since that is how they most frequently occur in books and reading materials. As the developed educational software is equipped with ASR technology, and communicates through natural language with

users, neither a mouse nor a keyboard were provided for the participant. A webcam was mounted on the screen top to record the interactions throughout the experiment for the purposes of offline analysis, inter-observer reliability, and procedural integrity. The computer's sound volume was set at a comfortable level via a built-in loud speaker at the back of the screen.

Connected to the same computer, there was a second screen placed one meter away from, and perpendicular to, the first screen, in a way that the participant could not see its content. At the beginning of each session, using this screen, an experimenter launched the educational software (to be displayed on the first screen), and an application which was responsible for recording the video/audio stream coming from the webcam. Along with that stream, the researcher simultaneously recorded the content of the first screen (the PA and virtual classroom) in a picture-in-picture format, which enabled post-analysis of the PA's activities as well as the participant's reactions and answers. On these recordings, the reading material presented to the participants and their response could be observed at the same time.

4.1.3 Screening

The target words were chosen from Dolch and Brigance functional word lists [178, 179]. Prior to the experiment, each participant went through a series of tests to screen out words that he was already able to correctly read. The first stage of screening was performed by the participants' teacher in a one-to-one format. She presented the words, printed on flash cards, one-by-one to him, and asked "what word?", and noted down the words that the participant either read incorrectly or did not read at all. During screening, the participants did not receive corrective

47

feedback or reinforcement for their answers. For each participant, the words were selected from word lists higher than their reading ability level.

The second stage of screening was performed in the experiment room by one of the researchers. The words identified by the teacher as unknown to each participant were again presented to them, similar to the previous stage. The only differences were that the words were displayed on the computer screen, instead of flash cards, and the task ("what word?") was delivered by the researcher. Similarly, no feedback was provided. The words order was random each session. This stage was repeated three times on three different days to ensure the words were unknown to the participants. Repeated assessment is important since one may know a skill but for whatever reason does not show the knowledge in that specific occasion. From the words that were never read correctly, the researcher randomly chose four for each participant to be included in the study. At the end of screening sessions, each participant had a unique set of four words appropriate to his reading level that he could not read correctly. The teacher agreed not to instruct these words to the participants for the duration of this experiment.

4.1.4 Experimental design

To evaluate the efficacy of the proposed educational software, a concurrent multiple-baseline across-participants design [180] was employed. In this experimental design, the researcher started by collecting baseline data on reading performance by the participants. After three data points across three days, the tutoring software was presented only to one of the participants, while still collecting baseline for the other two. As soon as the first participant reached the criterion and the performance was stable, the researcher began delivering instruction by the

software to the second participant, and continued collecting baseline data from the third. This procedure was carried on until the third student reached the criterion as well. The criterion was defined as scoring 100% accuracy for three consecutive sessions.

This experimental design proves the experimental control when the performance improves only with the introduction of instruction, while the performance remains stable at the baseline level without intervention. Such staggered introduction of instruction and continuous measure are meant to control threats to internal validity of the intervention, and attributes any significant improvement in the target behavior to the intervention itself rather than other external sources to acquire the skill.

4.1.5 Baseline sessions

Each baseline session consisted of four trials, one trial for each target word. During baseline, the actual PA and virtual classroom were first presented to the participants. The experimenter launched the application and entered the participant's name. The PA started by greeting the participant with his name, and telling him "we are reading some words today. I need your attention." After 3 seconds, it presented one of the words in random, and asked "what word?" The word was displayed for 5 seconds on the board, in order to give enough time for the participant to emit his answer, and then disappeared. No feedback or reinforcement were provided to the participants. Before the PA advanced to the next word, there was a 1-second inter-trial interval. This procedure continued until all four words were presented to the participant. While so, the experimenter silently collected data on the participant's reading performance, and did not comment

about accuracy. The recorded videos were also checked at the end of each session to insure reliability.

4.1.6 Instructional sessions

Instructional sessions were implemented based on a 0-5-second CTD procedure. Each session had 12 trials, 3 trials for each target word; one 0-second trial and two 5-second trials. Similar to the baseline sessions, the PA started with greeting and obtaining the participant's attention. After a 3-second pause, the PA started the first round of trials which included a 0-second trial for each target word. The procedural steps in the first round were (a) task delivery ("what word?") and presentation of stimulus (appearance of the target word), (b) immediately modeling the word to solicit student practice, (c) up-to-5-second latency, (d) corrective feedback (modeling the word again) or reinforcement ("good work," "excellent," "nice job," etc.), and (e) 1-second inter-trial delay. In the first round, this procedure occurred exactly the same for all four words, one-by-one. During the second and third rounds of trials, the target words were not modeled immediately after the presentation of stimulus. Therefore, the steps were (a) task delivery and stimulus presentation, (b) up-to-5-second latency, (c) corrective feedback or reinforcement, and (d) 1-second inter-trial delay. Order of words were randomized across sessions in order to prevent participants learning the order. Each session concluded by the PA providing non-contingent praise, such as "you are doing great!" or "keep up the good work", while clapping for the participant, to promote general motivation.

During the instructional sessions, the researcher sat in front of the second screen to observe performance and did not intervene in any way, since participants

were expected to work independently. He silently collected data on the participant's reading performance and double checked with the recorded videos after each session. An answer was only recorded as correct if the participant read the word correctly in the provided 5-second time slot.

4.1.7 Assessment sessions

After each participant reached criterion, an assessment phase was conducted which lasted at least three sessions for each participant. Each assessment session was exactly identical to the baseline sessions, where the participants were required to read the presented stimuli and did not receive any sort of word modeling, feedback, or reinforcement from the PA. This phase was specifically carried out in order to investigate whether the reading performance remained as accurate without further training. In other words, the researcher intended to see if they can actually read the words on their own without receiving models from the PA. Since the participants got used to receiving praise from the PA when they read correctly over the instructional sessions, at the beginning of the assessment phase, in order to avoid any confusion, the researcher notified the participant that "she" (the PA) was only checking which words he could read. Again, the words were presented randomly across sessions.

4.1.8 Maintenance sessions

An additional follow-up maintenance probe was conducted eight weeks after completion of the assessment sessions. The procedure was identical to the baseline and transfer assessment sessions. The goal was to examine whether the participants maintained the acquired words.

4.1.9 Generalization session

A final probe was conducted for each participant at their regular classroom and with their teacher in order to test whether the acquired words would transfer outside the experiment room. This session was identical to the screening session performed by the teacher before the experiment began. The teacher presented the words printed on flash cards, one-by-one, and recorded correct and incorrect responses.

4.1.10 Dependent measures

Two variables were defined as dependent measures: percentage of words read correctly (i.e., percent reading accuracy) and number of sessions to reach criterion. These two measures are indications of effectiveness and efficiency of the proposed tutoring software, respectively. Data on the first dependent measure were collected by the experimenter in an online manner, and were double checked after each session using the recorded videos in order to guarantee reliability.

4.1.11 Procedural reliability

In experiments that involve human instructors, a procedural reliability (also known as treatment fidelity or intervention integrity) evaluation is performed on the instructors' behavior to examine their adherence to the planned intervention. A researcher scores their step-by-step behavior using a checklist on items such as task delivery, presentation of the target stimulus, waiting the outlined latency, presentation of the controlling prompt, etc. This procedure is calculated as the percentage of correct behaviors out of planned behaviors [181].Inspired by this idea, an evaluation was performed for the tutoring system to ensure that the computer program ran properly. The video recordings were inspected to see whether the tutoring software implemented the procedure correctly according to

the instructional hierarchy. Not surprisingly, because it is a computerized procedure, almost all (all but one) items received 100% agreement between the planned and performed behaviors. A main characteristic of computer algorithms is that, in following a step-by-step procedure, they deliver a very high fidelity. This fact is actually one of the main incentives of using CBI as it can deliver highly-controlled, reliable procedures. The only item that the educational software did not receive an impeccable 100% agreement was, expectedly, due to the ASR engine. The speech recognition accuracy was 94.5%.

4.1.12 Inter-observer agreement

Twenty percent of the recorded videos were judged independently by a second observer. Results, calculated by inspecting point-by-point comparison between the two observers' judgements, indicated an agreement of 100%.

4.2 Results

The percent reading accuracies of participants across baseline, instructional, and transfer assessment sessions are displayed in Figure 11 Visual analysis of data indicates that participants succeeded in reaching the criterion although the number of instructional sessions varied. Intervention averaged 7.33 instructional sessions to criterion. Further, it is noted that performance showed a quick improvement immediately after introduction of the intervention until asymptotic knowledge was obtained. As expected, the reading accuracies during baseline sessions remained stably zero, as it was made sure the participants did not know the words in two different settings and four screening sessions. Following intervention, the participants read consistently with 100% accuracy and maintained mastery of the skill.

Students J, A, and M reached criterion in 6, 5, and 11 sessions, respectively. Follow-up maintenance and generalization data could not be collected from student A because he stopped attending school due to a family issue a month before the maintenance session. The other two students maintained the words with 100% accuracy. Furthermore, they successfully generalized the acquired words to an independent assessment by their teacher in a classroom setting.
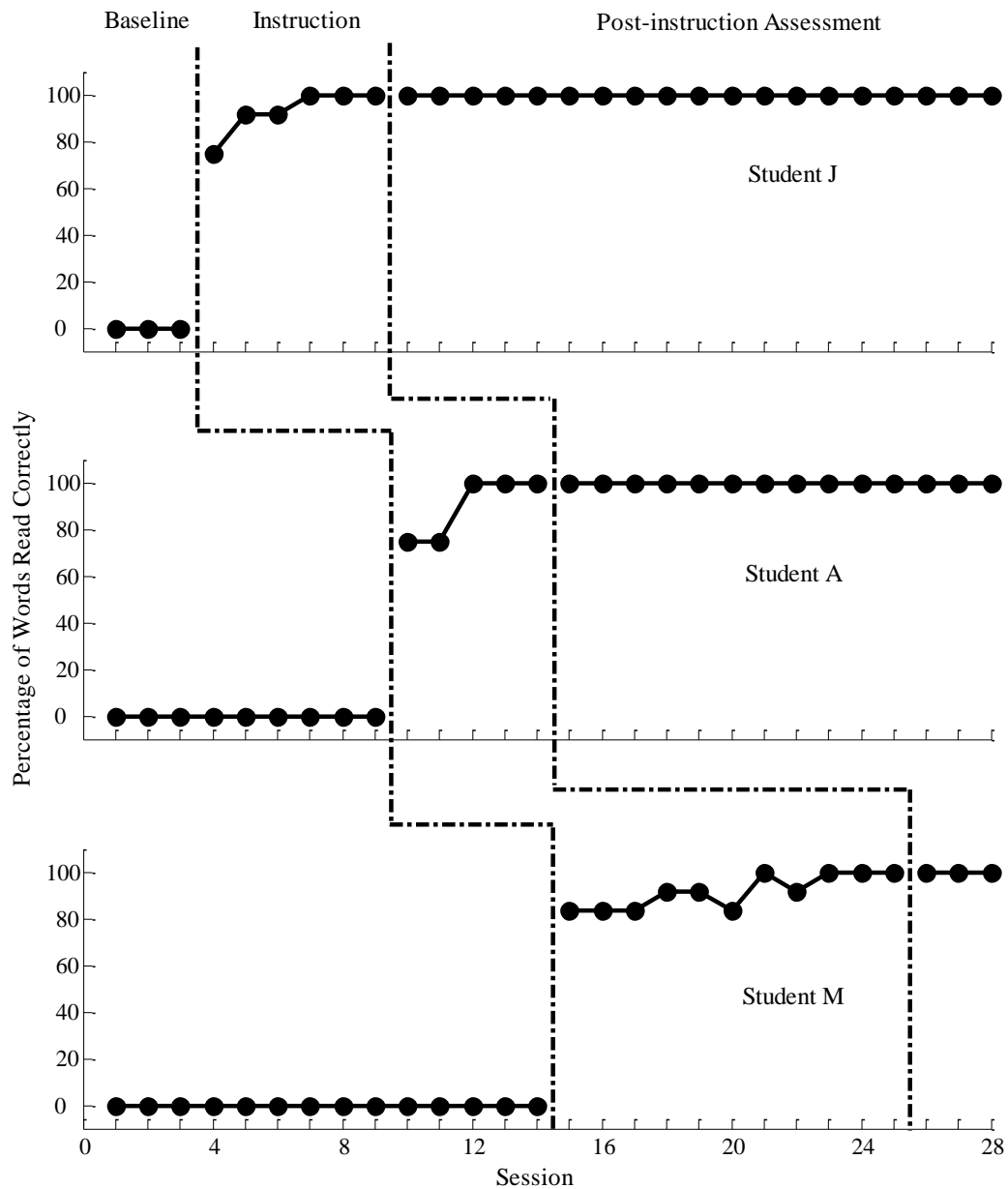
Figure 11. Percentage of correct responses during baseline, instruction, and assessment sessions.

4.3 Discussion

In this chapter, the effectiveness of an autonomous tutoring system developed for individuals with ASD was investigated. The software incorporates an autonomous PA that simulates a human tutor teaching SWR. The PA demonstrates body and hand gestures, facial expression, as well as head nods

and head shakes. The agent delivers instructions through synthesized speech and receives learner's responses via microphone and ASR. The proposed software allows for students to experience the words as spoken as well as written, and to hear themselves say the word.

The tutoring software was carefully constructed only based on proven techniques for individuals with ASD in order to optimally gain their attention (via the PA's gaze, head orientation, and pointing gesture), motivate them through timely reinforcement, and promote their learning via contingent corrective feedback. The results, obtained through a concurrent multiple-baseline across-participants experimental design, indicate that the tutoring software was successful in teaching SWR to the participants. The students also retained the obtained words after eight weeks, and generalized their skill outside the laboratory setting.

The tutoring software embeds an evidence-based instruction strategy, CTD, which is an effective and efficient method with a proven track record of success in ASD intervention [158, 162-165]. The CTD procedure includes immediate contingent reinforcement and appropriate feedback. In order to remain true to the original strategy, the developed software follows the steps with no changes. Too many deviations would result in an unreliable strategy which is incomparable with the original one.

A novelty of the proposed tutoring software is to integrate ASR technology to realize such contingency based on reading performance, doing so while furnishing a completely hands-off experience for teachers and parents. To the best

of the author's knowledge, the independent proficiency of the presented tutoring software is unlike all other CBI for individuals with ASD. The software autonomously deals with correct and incorrect reading responses, and provides appropriate, consistent verbal and non-verbal consequences by the PA. Such an autonomy means that instruction may be done by untrained providers even in the comfort of their homes, while collecting data on the student's performance. This data could be used internally by the tutoring software to plan future sessions, and to introduce new vocabulary by replacing words that the student mastered, thereby providing even more autonomy in instructing longer lists of words to the students. The data collected can also be accessed by teachers and caregivers for further analysis of a child's progress over time.

An issue had to be addressed in order to successfully integrate ASR in the current tutoring software. Children with ASD often lack the self-regulation to wait for the PA to stop talking and then emit their response. In a pilot study, children sometimes appeared to start reading the presented word before the PA finished the task delivery (i.e., "what word?"), especially when they got used to the program's predictability, and also when they had learned the word. In order to avoid simultaneous speech by the PA and the participant, which degrades the speech recognition accuracy, a possible solution could be to provide additional verbal prompts to help a participant respond at the appropriate time (e.g., "wait until she stops talking.") or perhaps the use of a beep sound (e.g., "wait until you hear the beep."). In order to provide a less confusing, natural interaction, though, the author used a more intuitive approach in which the word appeared on the chalk

board as soon as the task delivery finished. In this approach the PA verbally delivered the task while gesturally starting to point to the spot where the word would appear, by orienting its head, body, and hand. These two verbal and non-verbal cues were carefully synchronized in a way that they finished at the same time, and were immediately followed by the word appearing in written form. This approach not only resolved the simultaneous speech issue naturally but also may have made the sudden appearance of the word more gaze-capturing for the participants. In this method, the ASR engine launched its process for converting the participant's verbal response to text as soon as the PA stopped its speech and the word appeared, hence leaving no chance for speech interference. It was a simple but delicate timing of the task delivery and stimulus presentation, which proved completely effective.

In implementation of the computerized CTD procedure, another benefit of using ASR was that the tutoring software eliminated unnecessary pauses after each student's response. In other words, the latency provided by the software was not a fixed 5-second interval but rather the procedure was advanced to the next step (reinforcement, feedback, or a next word, accordingly) as soon as the participant emitted his response. This immediate reaction to the participant's response decreases the time-to-criterion measure by cutting the time that would be wasted otherwise and may result in the participants 's potential disengagement and boredom during the inactive period. It also brings about a more responsive PA, and, hence, a natural interaction.

Individuals with ASD often struggle to generalize skills acquired during an intervention to novel settings and people [182, 183]. They may become adept at a trained skill but fail to transfer that knowledge to everyday life. This makes development of interventions for these individuals extremely challenging. The ability to generalize skills, however, can be promoted by providing a naturalized learning environment to facilitate transition to natural situations [184]. In the proposed tutoring system, the instructions are delivered in a simulated classroom by a virtual tutor that communicates with students through speech, facial expression, and gesture resembling a human teacher. The PA creates mutual gaze by looking out from the screen into a viewer's eyes. It then initiates joint attentional bids by rotating its gaze direction, head orientation, and pointing towards the reading material in order to optimally capture a participant's attention.

The author believes that such multi-modal interaction (through gaze and body gesture as well as voice and facial expression) not only enhances the participants' engagement and attention but also promotes better generalization of the acquired words. The results of this study demonstrated that the participants transferred their knowledge, acquired through meaningful communication with the PA in a simulated naturalized environment, to their classroom.

In order to encourage generalization even further, other methods can be integrated in such a PA-based tutoring system that are worth future research, including changing the virtual environment and using multiple agents with different appearances and voices. With such a variety, learners are less likely to associate the tasks to a constant setting or a single presenter. It is important, however, to

perform these changes gradually over the course of several sessions, as these individuals have a preference for sameness. Sameness brings about predictability in which they find great comfort. On the other hand, randomization and variability help generalization but may cause anxiety.

Another challenge in developing interventions for individuals with ASD is to take into account sensory overload as it poses a major pressure on some of these individuals. To avoid sensory overload, the author designed the virtual environment to be simple, and avoided clutter and excessive use of sound. Further, the PA's motion, appearance, and expressions were kept intuitive and friendly, thereby avoiding confusion, distraction, and anxiety. The instructional strategy adopted, CTD, is a predictable and repetitive procedure, preferable to this population. Each session is less than three minutes long which is ideal for those with a short attention span, especially children who are likely to get bored relatively quickly.

Upon completion of the experiment, the participants completed a survey regarding their interaction with the tutoring software. According to this survey, they found the task easy to understand and interacting with the PA enjoyable. On the recorded videos, there were numerous instances of the participants smiling and showing excitement due to the reinforcement from the PA. At the end of each session, when the PA praised and clapped, one of the students verbally commented on his own performance such as "excellent job," "fantastic," "I did it," or "it sure was fun!", and shared his excitement with the experimenter. With a successful, friendly reciprocal social interaction, a PA may establish a good rapport

with the students. In the survey, in response to the question "what did you like most in this game?", he said "she loves to talk to me." Their teacher also mentioned several observations of the students talking with each other in the classroom about the tutoring system — a social practice not common among people on the autism spectrum. These observations demonstrate the potential of PAs for ASD intervention as engaging and rewarding tools to practice social interaction during the intervention and to mediate interaction with a social partner.

The current chapter illustrates the promise of CBI, and PAs in particular, to supplement conventional techniques in teaching SWR to ASDIs in a cost-effective manner. The developed autonomous software affords extensive one-to-one practice with no teacher supervision, thereby allowing more efficient use of the teacher's time. Further, with access to the software, students can practice repeatedly at home at their own pace, without potential embarrassment in front of classmates. Instruction is available any time of the day, which is helpful for those with irregular sleep patterns. The software also automatically records the data on the child's performance such as the amount of time spent on the task, errors, and correct responses. It is a useful time-saving feature as teachers often need to collect and report progress of students.

This study expands the literature in several ways, including the CTD procedure, sight word instruction, use of ASR technology in CBI, and PAs. A main limitation of this study, however, is the small number of participants and absence of female participants. Therefore, the findings, although very encouraging, should be deemed preliminary, and cannot be declared to be generalizable to all

individuals with ASD. Certainly, larger studies are needed to verify the results and to examine participants with a variety of disabilities. Besides, the relationship between individual differences (such as chronological age, language age, and IQ) and the system's effectiveness must be investigated. Additionally, longitudinal studies with longer word lists are needed to examine and better understand students' progress over long-term training.

CHAPTER 5

EMOTIGO: BLUETOOTH-ENABLED EYEWEAR FOR UNOBTRUSIVE PHYSIOLOGY-BASED EMOTION RECOGNITION

While machines (e.g., computers, tablets, cellphones, robots) have begun to co-exist with humans and cooperatively share various tasks; they are still deficient in interpreting and responding to humans as emotional beings. Wearable biosensors that can be used for computerized emotion recognition have great potential for addressing this issue. This chapter introduces a wearable sensor – EmotiGO – for unobtrusive acquisition of a set of physiological signals, i.e., galvanic skin response, photoplethysmography, and skin temperature, which can be used as autonomic readouts of emotions. This chapter presents the architecture of the device and results from testing that verify its effectiveness against an FDA-approved system for measuring physiological signals.

5.1 Introduction

Recent years have seen a large increase in research in the field of Affective Computing (AC) to fill the communicative gap between machine and the emotional human [185]. One main goal of AC is to develop computational systems that reliably detect humans' emotional cues, and alter machines' internal/external behavior based on their affective states (e.g., emotions). Affect-sensitive systems are being developed in a number of domains. For instance, in intelligent tutoring systems, an affective paradigm can be adopted to automatically recognize and pertinently respond to the student's emotional experience (e.g., frustration,

boredom) which, in turn, could optimize learning gains [69, 186]. AC has many applications such as in gamification [187], personalized robot companions [188], autism therapy [189], and healthcare [190].

The most prevalent techniques for automatic emotion recognition are based on facial expression [191], speech prosody [192], and more recently, physiology [193]. Less effort has been done to infer emotional states using physiological signals compared to the other methods. Reasons for this contrast are physiological sensors are relatively expensive and sensitive to motion artifacts. Physiological signals, however, have a number of promising advantages. They can be continuously collected during daily life which makes physiology-based emotion recognition suited for a wide range of real-world applications. Besides, they are more difficult to suppress compared to audiovisual emotion channels − hence could be a more reliable reflection of inner feelings [194]. There is a strong correlation among physiological indices and affective states. For instance, when someone is frightened, her heart rate increases; her muscles contract; she breathes faster and tends to sweat more. These physiological patterns can be monitored to automatically recognize emotions.

Most contemporary affective computing prototypes, however, make use of inflexible sensors, and still tether the user to their computers in an unnatural, obtrusive manner. Lack of comfortable, unobtrusive physiological sensors has limited widespread use of affect-sensitive systems. However, recent advances in the field of wearable technology can facilitate monitoring physiological signals in real life settings over extended periods of time.

This rest of this chapter presents a wearable multi-sensorial biofeedback device, EmotiGO (Figure 3), which is a pair of modified eyeglasses to include three biosensors. The parameters measured are Galvanic Skin Response (GSR), Photoplethysmography (PPG), and Skin Temperature (SKT); a set of measurements useful for affect sensing. The collected signals are then transmitted to a remote computer via Bluetooth Low Energy (BLE) communication in real time. The general layout of this chapter is as follows. The next section overviews related work and recently developed biosensors for affect sensing purposes. Section 5.3 discusses design challenges in wearable biosensors and how they are addressed in EmotiGO's design. In Section 5.4, the physiological signals measured by EmotiGO are described. Section 5.5 presents the general system architecture. Section 5.6 outlines the procedure and details on sensor testing and evaluation. And finally, conclusions are presented in Section 5.7.

5.2 Related work

In spite of enhancements in physiology-measuring technology, most of the research in the field of affect sensing has been limited to short-term studies in laboratory settings. There are a number of commercial physiological data acquisition systems which offer high-quality recordings such as the MP150 from Biopac Inc. [195]. MP150 is a general purpose acquisition system for collecting various physiological signals. The electrodes are attached to the subject's body and then wired to a central collecting unit. The collecting unit is connected to a computer by USB or Ethernet cable. MP150 lacks wireless capabilities to transmit data to the host computer, and tethers the subject with wired electrodes. It is accompanied by proprietary software only through which users can access and

65

analyze collected data. Additionally, its high cost, bulky form factor, and tethering nature make it impractical for large-scale and long-term applications.

A very important physiological signal in AC is GSR, also known as electrodermal activity (EDA) or Skin Conductance (SC), which is of particular interest in monitoring arousal levels of subjects. Researchers at the MIT Media Lab developed one of the first wearable biosensors, Galvactivator, which collects a wearer's GSR signal [196]. Galvactivator's components are embedded in a glove that is worn on the left hand, and two electrodes collect the GSR signal from the wearer's palm. Galvactivator has no wireless capabilities but there is a small jack that allows the collection of the amplified signal by a host computer in a cabled approach.

HandWave [197] is a Bluetooth wristband device that measures one's GSR signal from one's fingers or palm via external leads and electrodes. The external leads and the electrodes' position render HandWave encumbering and highly prone to motion and pressure artifacts. PET GSR is similar but is a commercial wireless GSR sensor from Brainquiry [198]. PET GSR suffers from the same drawbacks as the HandWave's. Furthermore, PET GSR's wireless equipment is only compatible with proprietary software.

MARSIAN [199] is a wrist-mounted device with a glove. It includes four physiological sensors of GSR, skin potential, heart rate, and SKT. The glove embeds the sensing elements, and the wrist device includes the electronics and battery. EREC-II [200] is also a glove-based biosensor which embeds SKT and GSR sensors. The analog circuitry and battery is placed inside a wrist pocket, and

the sensing elements are embedded in a cycling glove. Similar to MARSIAN, EREC-II's form factor is rather large, and the electrodes' positions are not suitable for pervasive mobile applications.

iCalm [201] is a wireless wearable set of physiological sensors designed for affect-sensing applications. This device can measure either heart rate or GSR, as well as environmental temperature and physical activity. In the version of iCalm which measures GSR, instead of typical metal electrodes, conductive fabrics have been used. Conductive material facilitates comfortable integration of the electronic parts and the sensing circuitry into various wearable garments such as a wrist band, ankle sensor band, or sock. Poh et al. [202] developed a wristband device that measures the GSR signal from the ventral side of a wearer's distal forearm. This device as well as iCalm are big steps towards unobtrusive affect sensing. However, they provide either heart rate data or a GSR signal, not both. Furthermore, their placement on body locations involving high physical activity (i.e., hand and leg) exposes them to vigorous motion, which inevitably introduces artifacts in the collected data.

E3, designed by Empatica [203], is a commercial wearable set of biosensors that has a form factor of a wristwatch. E3 includes four embedded sensors to measure PPG, SKT, and GSR signals as well as physical activity. E3 measures a wearer's PPG signal from the dorsal area of the wrist, while the GSR and SKT sensors' collection site is the ventral area of the wrist. E3 utilizes a proprietary technique including software and hardware to enhance the PPG signal

quality against motion artifacts in the data, although the GSR sensor is still not guarded against artifacts.

Having reviewed recently-developed wearable biosensors for affect sensing purposes, the next section discusses general challenges in the design of wearable biosensors for pervasive applications.

5.3 Design Challenges

Most current sensor systems, such as Biopac MP150, use electrodes attached to the subjects with Velcro fasteners or tape and are directly wired to the collecting unit. These constraints on data collection may be irritating to some subjects and can be a hindrance to their natural movement. EmotiGO is innovative in that it includes multiple biosensors in a single device all integrated through non-standard placement of measurement elements, and miniaturized to a level where the whole system is enclosed in an inconspicuous, lightweight, comfortable form factor. Such integration eliminates the need to use distributed multiple sensors (e.g., heart rate chest strap and finger-placed GSR sensor), thus resulting in a significant decrease of the overall system's size. Additionally, the novel collection site and the unencumbering enclosure facilitate monitoring of physiological signals without interfering with the users' usual activities.

EmotiGO uses BLE for real-time data transmission of the collected data. The BLE technology consumes only a small fraction of the power of the classic Bluetooth. This fact, along with the aggressive low-power design techniques applied, allow EmotiGO to operate for many hours on a small lithium-ion battery. BLE's advanced adaptive frequency hopping technology allows multiple EmotiGOs to reliably operate close to each other even in noisy environments [204].

EmotiGO's real-time data transmission is also an important advantage as it allows online collection and analysis of measured signals. SInce all the electronics and electrodes are invisibly integrated in a garment usually worn by people (i.e., eyeglasses), EmotiGO is user-friendly and is likely to be quickly accepted by users.

EmotiGO collects the signals from one's face. Hence, it is intrinsically less prone to motion artifacts, since head movements are much less, both in amplitude and frequency, compared to arm and leg movements. In general, GSR signal is measured via application of conductive gel on the skin and attachment of metal electrodes [205]. EmotiGO, however, does not require application of conductive gel, which makes it feasible for everyday usage.

Currently, most devices employ proprietary software which make them impractical for widespread use. Additionally, the collected data can only be accessed and displayed using the manufacturer's software and are not accessible instantly to other applications and third-party products. The hardware itself, also, can be commanded (such as on/off states and, if at all selectable, desired sampling frequency) through the provided proprietary and closed software. A desired feature of such measurement devices would be an open protocol which provides the capability of controlling the device and accessing data directly and instantly for external applications. EmotiGO's software is designed based on an application programming interface (API) which makes it very easy for developers to design external applications to interact with the device (i.e., externally control the device and immediately access the data) without worrying about the underlying architecture.

EmotiGO's API is a portable binary-based communication protocol designed specifically to simplify application development on various host environments using different programming languages. Through the API, application developers can scan for all EmotiGOs within a certain distance and connect to a specific one or all of them simultaneously; set the sampling frequency; select the signals' measurement range, and hence the resolution of sensors; start/stop data acquisition when desired; access data immediately; change the transmission power and range; read the battery level; and disconnect and make EmotiGO go to sleep. The open architecture and data format allows software developers to rapidly adopt EmotiGO and easily incorporate it into their systems as an input device. EmotiGO has been developed with friendliness in mind − from the end users' point of view as well as the developers'.

5.4 Physiological Signals

EmotiGO has been designed to measure bodily reactions to infer users' emotional states. Fusion of multiple channels, compared to a single measure, can increase the confidence level in emotion recognition. In this dissertation, the following set of biosignals are considered for affect inference: GSR, PPG, and SKT.

5.4.1 Galvanic Skin Response

Skin sweat is a weak electrolyte and a good conductor of electricity. When someone is exposed to cognitive, physical, or emotional stressors, eccrine sweat glands in the skin produce ionic sweat which elevates GSR [205]. This signal is a well-studied physiological measure, previously employed in monitoring emotional states [206, 207]. GSR signal has two components; tonic and phasic. The tonic

component, also known as Skin Conductivity Level (SCL), is a slowly-changing baseline level and is due to the general psychophysiological skin activation. The phasic component, also known as Skin Conductivity Response (SCR), superimposes on top of SCL and has a higher frequency nature [205].

The most common sites for recording GSR are the medial and distal phalanxes of the ring and index fingers, as well as the thenar and hypothenar eminences of palm. The magnitude of GSR depends on the density, size, and distribution of sweat glands [208] which varies between different measurement sites. The highest density of eccrine sweat glands is estimated to be on the palms and soles (600 to 700 glands/cm2) [209, 210]. In different body locations, lower densities of the eccrine sweat glands are found such as on the forehead (181 glands/cm2) and forearm (108 glands/cm2) [209, 210]. Several investigations have studied different collection sites such as forehead [211, 212], cheeks [211], and wrist [201, 202].

Tronstad et al. [211] investigated sweating patterns of the abdomen, the hypothenar area of the palm, the neck, and the forehead on 24 subjects during relaxation and intense physical activity. Interestingly, forehead showed higher GSR than the hypothenar area of palm (followed by abdomen and neck). This finding is not in line with the order of eccrine gland densities. However, other than the density, various factors including the distribution of dermatomes, and the innervations of sweat glands also influence GSR magnitudes and the similarity between readings at different locations [212]. In [212], GSR readings at 16 different body positions were compared on 17 subjects while watching emotional film clips.

Findings indicated that forehead, foot, finger, and shoulders had the highest SCL and SCRs. Surprisingly, none of the wrist collection sites were amongst the four locations with the highest SCL and SCRs. This fact is of a high significance as many recently developed wearable devices, such as iCalm and E3, tend to collect GSR from the user's wrist.

Since both hands are needed for manipulation, placing the GSR electrodes on the hands is impractical for a wearable device, as it interferes with daily activities and introduces motion artifacts. Therefore, embedding the GSR sensor in devices worn on highly active body parts, such as wrist, is not optimal either. Hence, the author sets out to design the wearable biosensor to collect GSR signal from the user's face. EmotiGO discreetly collects GSR from the user's temple. Findings in [211, 212] suggest forehead as a viable collection site for GSR signal with high SCL and SCR readings. Because temple and forehead are close, the author hypothesizes that the temple shows acceptable SCL and SCR readings as well.

5.4.2 Photoplethysmography

One's heart rate elevates in reaction to various stimuli such as physical activity, affective states, sexual arousal, and mental effort [213]. Using a well-acquired PPG signal, one can measure heart rate variability (HRV) as well as blood volume pulse. PPG sensors usually emit light to the tissue (using an LED), and measure the light reflected by the blood vessels. Unfortunately, PPG is notorious for being highly susceptible to motion artifacts. In most cases, these motion-induced artifacts have their highest energy within the same frequency band of PPG signal, rendering noise filtration very challenging. Therefore, for the same reasons

mentioned for the GSR sensor, recording PPG signal at one's wrist is not optimal due to vigorous hand movements as well as frequent local wrist motions.

Site selection for measuring the PPG signal is also a critical design consideration regarding the power consumption. Wrist has a relatively low density of superficial blood vessels, which calls for use of brighter LEDs with higher current consumption. The investigation in [214] compared the amplitude and quality of the PPG signals recorded from forehead and wrist. The investigators used LEDs with relatively low brightness in the sensor. They observed that a considerable amount of external pressure on the sensor was required in order to measure a discernable PPG signal from the wrist. On the contrary, their sensor obtained significantly stronger and less noisy PPG signal from the forehead using minimal contact pressure.

EmotiGO measures the PPG signal from the temple which is a highly vascular region due to the temporal artery. Temporal artery is readily accessible lying just 1-2 mm under the skin surface. Besides, the temporal artery is trapped between the skin and the skull. The temple bone, which is covered only by a relatively thin layer of skin, has a strong light reflection. Because of these advantages and the fact that the temple is very close to the forehead, the author hypothesizes that EmotiGO's PPG recordings will be of high quality.

5.4.3 Skin Temperature

The relation between SKT and affective states has been investigated in several studies. For instance, it has been shown that the finger temperature decreases with stress [215], while it increases in relaxation [216]. EmotiGO uses a micro temperature sensor that contacts the wearer's temple and measures SKT

on top of the temporal artery. Temporal artery, lying just underneath the skin surface, carries blood directly from the heart and is located a short distance from it, thereby providing a better indicator of the core body temperature [217] – compared to distal locations such as the fingertip. Moreover, because the temporal artery is not significantly influenced by thermoregulatory changes and has very little vasomotor activity, it is highly perfused and a steady flow of blood is assured [218].

5.5 System Architecture

The architecture of a wearable system that collects and wirelessly transmits physiological signals generally consists of three main components: a wearable device, a base station, and a host application. In the rest of this section, these main components are described.

5.5.1 Wearable Device

In the developed tutoring system, EmotiGO plays the role of the wearable device for real-time collection of the physiological signals from students. EmotiGO includes a BLE module, an analogue-to-digital converter (ADC), a PPG sensor, a GSR sensor, an SKT sensor, and a rechargeable battery (Figure 12). The main component of EmotiGO is its BLE module which is the BLE113 from Bluegiga [219]. BLE113 is based on a relatively powerful CC2541 chip from Texas Instruments (TI) [220]. BLE113 is a Bluetooth Smart module that integrates Bluetooth radio, software stack, and an 8051-compatible CPU core. The CPU only handles the communication protocol, hence, it is idle most of the time. In EmotiGO, the author uses this idle time to execute the application code. Therefore, EmotiGO is a standalone device without a separate microcontroller.
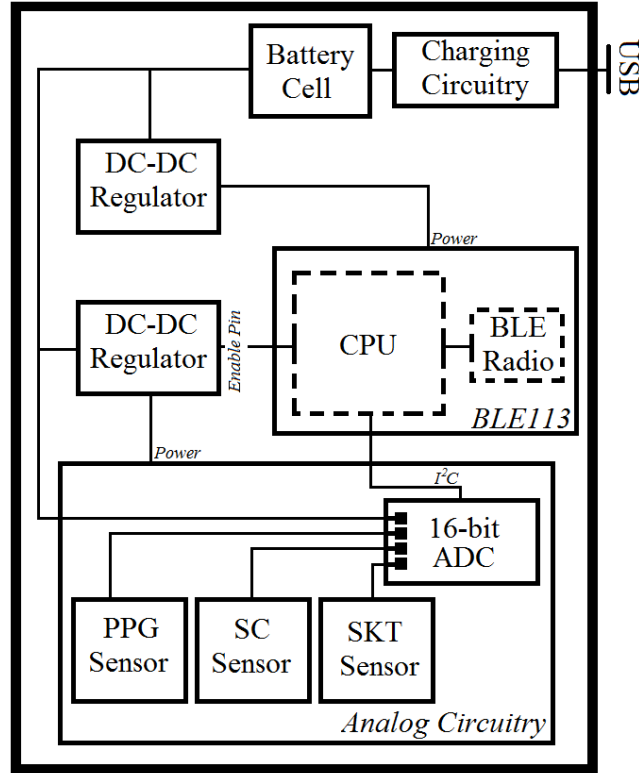
Figure 12. EmotiGO's architecture

Although BLE113 has internal 12-bit ADCs, the author uses an external low-power 16-bit ADC from TI, ADS1115, to achieve much higher resolution. ADS1115 features four input channels that EmotiGO uses for sampling PPG, GSR, and SKT signals as well as the battery charge level. Data are transferred to the BLE113 via an I^2C interface. An onboard programmable gain amplifier (PGA) on the ADC can be dynamically controlled which allows measurement of both large and small signals with high resolution. In EmotiGO, this gain can be set programmatically, which renders EmotiGO a nicely customizable measurement device. ADS1115 has a single-shot conversion mode that allows EmotiGO to sample the physiological signals very efficiently. In this mode, the ADC performs one conversion of the input signal and transfers it to the BLE113 module upon its

request. The ADC then automatically enters a low-power shutdown mode which only consumes 0.5µA.

To measure the GSR signal, EmotiGO applies a small constant voltage to the left temple using two small medical grade silver/silver-chloride disk electrodes. The small current that flows from one electrode to the other is then boosted by a low-pass filter with a 1-Hz cut-off frequency. The ADC samples the filtered signal, and sends it to the BLE113 module. EmotiGO can measure GSR signals between 0 and 40 µSiemens. Based on the high resolution of ADS1115 and its PGA, EmotiGO offers very high and customizable resolution of GSR measurement that can be set programmatically (through the API).

In the PPG sensor, a single infra-red LED with the peak wavelength of 850 nm is used. There is a small photo diode (PD) adjacent to the LED that converts the light reflected from the wearer's skin to a very small current. This current is then conditioned by a transimpedance amplifier. The amplifier is a low-pass filter with the cut-off frequency of 3 Hz. The output of the transimpedance amplifier is sampled by the ADC and sent to the BLE113 module. The skin temperature is measured using LM94022 which is an extremely small, precise, analog temperature sensor from TI. LM94022 has a typical supply current of merely 5.4µA. The output of this sensor is filtered by a passive low-pass filter with a cut-off frequency of 1 Hz.

As depicted in Figure 13, the components are placed very densely on two separate printed circuit boards (PCBs) which are of two different sizes; a smaller one and a larger one. The smaller PCB contains the sensing elements (i.e., the

GSR electrodes, the temperature sensor, LED and PD) which need to be in direct contact with the wearer's skin surface. The space surrounding the elements is covered with resin for electrical isolation. Furthermore, the resin is chosen to be black to prevent external light disturbances and multiple scatterings of the LED's light. All the elements on the smaller PCB and the resin coating are of equal height (1 mm), sitting smoothly against the skin. The rest of the components are tightly placed on the larger PCB. The two PCBs are electrically connected through a number of header pins.
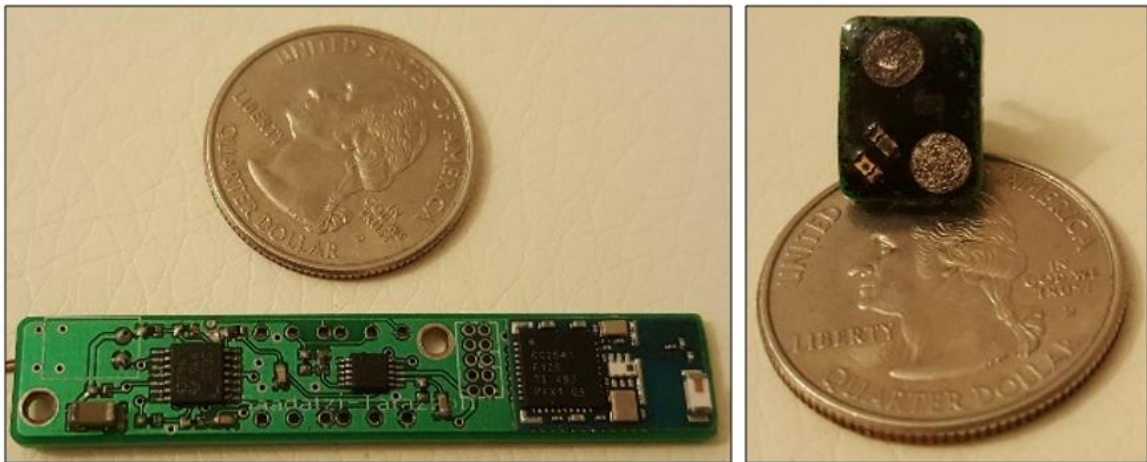
Figure 13. EmotiGO's PCBs

Because EmotiGO is to be worn as eyewear for long-term measurement, the battery has to be as lightweight as possible. EmotiGO has a rechargeable lithium-ion CR2450 coin cell with the nominal voltage of 3.6V and the capacity of only 120mAh. With the EmotiGO's small battery capacity in mind, rigorous low-power design techniques are adopted in all aspects of EmotiGO. EmotiGO's components are put to sleep when they are unused. For instance, the ADC enters a low-power shutdown mode after each conversion. This technique provides significant power savings during idle periods between the conversions. Depending

on EmotiGO's sampling frequency, the power consumption of the ADC can be a few hundred times more efficient using this technique. The BLE module itself is programmed to sleep after a certain amount of inactive time, where its current consumption is merely 500nA.

Choosing BLE technology for the wireless capability of EmotiGO contributes considerably in EmotiGO's consumption reduction. Such a low consumption level required by EmotiGO, however, cannot be obtained by the radio hardware alone; it also calls for smart design of the rest of the hardware as well as the software. In EmotiGO, there is no separate microcontroller to run the application code. The application code is efficiently executed right on the same module that handles the Bluetooth communication, thereby eliminating the need for an extra CPU. This elimination brings about drastic improvements in size, power consumption, weight, and cost of EmotiGO.

In a typical PPG sensor, the power consumption of the LED constitutes a major portion of the total consumption because the LED must be bright enough so that the light can reach the blood vessels beneath the LED. In EmotiGO, however, the LED does not need to be as bright because the human temple is a highly vascular area with a thin covering of skin as well as a bone with strong light reflection. In EmotiGO, the LED consumes less than 2mA. Furthermore, because the sampling frequency of EmotiGO as well as its transmission range are customizable, users can adjust them according to their needs, thereby avoiding waste of power. All this customization comes through the use of APIs, explained in the next subsection.

5.5.2 Base Station

In order to receive the data collected on EmotiGO, one remote computer needs to be BLE compatible. However, many computers still do not have built-in BLE transceivers. To address this issue, the author used a BLE dongle from Bluegiga, BLED112, which integrates Bluetooth Smart features. BLED112 is based on TI's CC2540 chip and has an 8051-compatible core as well as a USB interface.

BLED112 is a general-purpose BLE dongle that comes with a set of predefined APIs. Since BLE technology is not designed for high data rates, this set of pre-defined APIs is unnecessarily comprehensive and too heavy for this application. Therefore, the author erased the code on this dongle entirely (using TI's CC-debugger) and reprogrammed it with different software that handles a customized set of APIs optimally designed for communicating with EmotiGO. In the following, the modified dongle and the set of APIs are referred to as EmotiGO dongle and EmotiGOAPI, respectively.

The EmotiGOAPI is simply a transport protocol which is used to implement applications on a separate host such as a computer. The transport protocol is designed to control the Bluetooth stack on the dongle in order to transmit measurement parameters to and receive data from EmotiGO. The EmotiGOAPI is a binary-based communication protocol designed specifically for ease of host application development.

The Bluetooth communication between EmotiGO and EmotiGO dongle is based on a client-server architecture. EmotiGO, which is the server, exposes information including measurement parameters and collected data. EmotiGO

dongle, on the other hand, is the client through which a host application can set the measurement parameters and receive the data collected on EmotiGO. EmotiGO also acts as a slave which advertises and is connectable. EmotiGO dongle, on the other hand, is the master which scans for advertising EmotiGOs.

The EmotiGOAPI is a command-response-event protocol. First, a command is submitted by the host application to the EmotiGO dongle using an API code. Based on the command, some sort of activity occurs on the dongle and the response is sent back to the host application. Besides, some activities by EmotiGO trigger events on the dongle that are forwarded to the host for further decision making. A simple example is provided below.

- The user switches on the EmotiGO and puts it on. EmotiGO is programmed to start in advertising mode as soon as it comes on.
- Then an application is launched on a computer which commands the dongle, using an API code, to start scanning for the EmotiGOs in range.
- The dongle starts the scan process. Among all Bluetooth devices around, it searches for a specific universally unique identifier that is chosen for all EmotiGOs. The dongle keeps sending IDs of any advertising EmotiGOs back to the host application until it receives another API code to stop scanning.
- After the host application receives the IDs of available EmotiGOs, it commands the dongle to connect to a specific EmotiGO, via an API code that includes the desired ID.

- When the dongle receives the code, it initiates a connection request to the desired EmotiGO. As soon as the connection between the EmotiGO and the dongle is established, an event is triggered on the dongle itself which is forwarded to the host application as an indication of successful connection.

- Having established the connection, the host application can set parameters such as sampling frequency, measurement range of the GSR sensor, and Bluetooth transmission power via a set of API codes.

- As soon as all the parameters are set by the host application, it then commands the dongle to start the collection. EmotiGO, therefore, launches the collection and wirelessly transmits them to the dongle in real time. The dongle receives data and forwards them to the host in a certain format that can be parsed by the host application. As soon as new data comes in, the host application can store them, display them, perform online analysis, etc.

- This process goes on until the host application commands the dongle, via an API code, to stop the data collection.

This architecture provides a seamless communication link between the host application, the EmotiGO dongle, and the EmotiGO, while the whole process is transparent to the end application. Furthermore, the bi-directional wireless communication between the host and EmotiGO allows for greater flexibility during deployment.

5.5.3 Host Application

Host applications can be designed with any programming language using the EmotiGOAPI. The author developed a graphical user interface (GUI) to

streamline the process of real time collection of physiological signals (Figure 14). The GUI has been developed in MATLAB and allows researchers to interact with the wearable device in a friendly manner. Using menus, the GUI allows the researcher to select and attach to any EmotiGO dongle that is plugged into the computer USB ports. When a dongle is attached, by the host application, the researcher can scan for available EmotiGOs in range and connect to a specific one. The measurement options (e.g., sampling frequency, transmission power) can also be selected on the GUI. When all the parameters are set, the researcher can initiate the data collection. The GUI then records the incoming data in real time and displays them in dynamic graphs. The graphs are zoomable and scrollable. After the collection is over, the data can be saved in different formats for future analysis. On the GUI, pre-stored data can also be displayed for manual inspection.
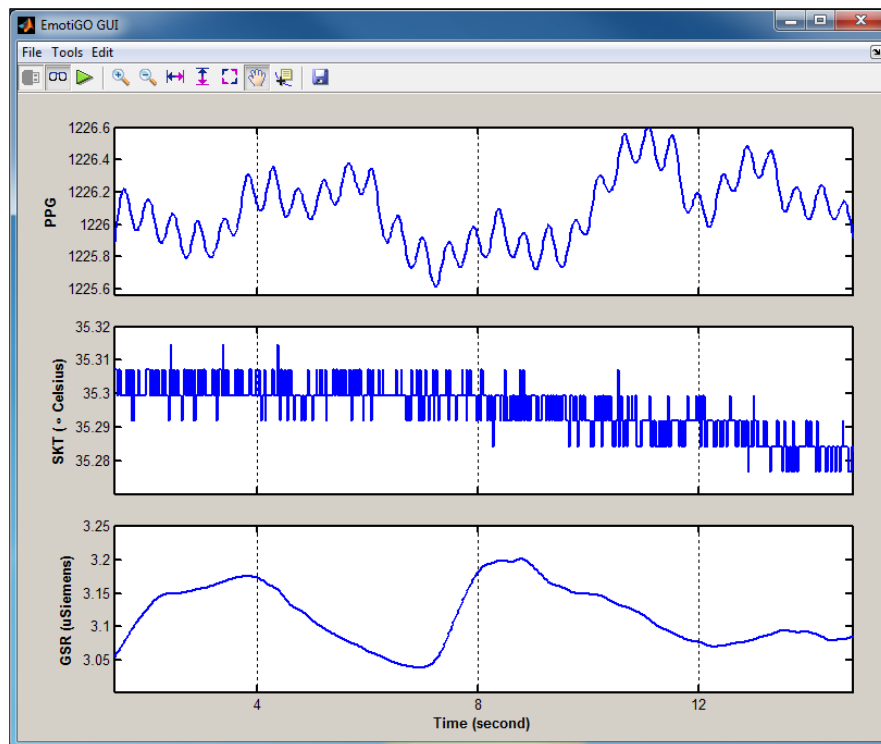
Figure 14. EmotiGO GUI

5.6 Sensor Testing and Validation

In this section, several trials are performed in order to validate EmotiGO's functionality. The goal is to examine the feasibility of simultaneous collection of PPG, GSR, and SKT signals using one centralized measuring unit from the wearers' temples. The author intends to confirm that the temple is a viable collection site for the aforementioned physiological signals. There are a number of factors about EmotiGO that call for extensive validation and testing. First, the temple is not a standard location to collect GSR signal. GSR readings depend on several skin characteristics such as concentration, distribution, and size of sweat glands at each specific measurement site. To the best of the author's knowledge, there is no study that investigates the measurement of GSR signal at this location. Another factor that renders EmotiGO's GSR sensor even more unusual is that EmotiGO is intended to function without conductive gel. This characteristic considerably contributes to EmotiGO's usability in real-life applications although it warrants precise examination and validation.

Second, PPG and GSR signals are susceptible to motion artifacts. Because EmotiGO is developed for natural applications, it is required to evaluate its performance while partaking in high physical activities. Besides, the temple and palm belong to different body regions regarding the skin thermoregulatory functionality. Therefore, the temple's SKT must be compared to that of palm in order to have a better understanding of this signal's variations at different body locations.

In this section, the quality and characteristics of the signals collected by EmotiGO are evaluated. The validity tests were carried out against Biopac MP150,

which is an FDA-approved system, while the participants engaged in a physical task, startle test, and a cognitive task.

5.6.1 Participants

Three participants (two male and one female), average age 32 years, consented to complete the experiments. This study was approved by the University of Louisville Institutional Review Board (IRB).

5.6.2 Procedure

Before EmotiGO is worn by the participants, they washed their faces with lukewarm water and patted their skin dry. There was no pretreatment of the temple such as Ethanol or conductive gel. MP150's GSR electrodes were attached to the distal phalanxes of index and ring fingers with conductive gel. The SKT and PPG sensors were also attached to the distal phalanxes of thumb and middle finger, respectively. The wires were affixed to their wrist so that there was no wire drag. The sampling frequency of Biopac's sensors and that of EmotiGOs' were chosen to be 1000 and 62.5 Hz, respectively.

After the attachment of Biopac sensors and wearing EmotiGO, each participant sat on a stationary bike, and rested their left hand on a cart next to the bike. Because the collected signals on Biopac are the gold standard reference, the participants were asked to keep their left arms as still as possible, thereby avoiding introduction of motion artifacts into the gold reference. The participants were asked to sit quietly and relax with eyes open to collect five minutes of baseline data. After the baseline, as the physical task, each participant started to pedal on the stationary bike for five minutes while maintaining a speed of 12 mph at a fixed resistance level of eight. A three-minute recovery period followed the physical

activity. Right after the recovery period, the startle test started [221]. During this stage, a very loud noise (air-horn sound) was played on the computer speaker every one minute for five times. This test is to elicit a defensive response from the participants, known as startle response, which is associated with negative affect.

After the fifth burst, there was another one-minute recovery period, followed by a Stroop word–color matching test [222]. The Stroop test, which creates a conflict between an incompatible color and word (for instance, the word RED in font color blue), is a psychology test used to study cognition and emotion. This test continued for 2.5 minutes. After completion of the Stroop test, participants remained seated during a 10-minute recovery period.

5.6.3 Results

Several algorithms were written to investigate the performance characteristics of EmotiGO against Biopac MP150 for each of PPG, GSR, and SKT signals.

PPG signal: The algorithm that analyzes PPG signals from EmotiGO and MP150, applies a low-pass filter on both signals to remove measurement noise. After noise removal, the algorithm measures the time difference between every two consecutive systolic points in order to calculate instantaneous heart rates. The series of these instantaneous heart rates constructs a staircase signal known as HRV. The HRV signal collected from EmotiGO's PPG signal shows less than 2% error on average compared to that of Biopac MP150. By taking into account that the participants did not limit their head and body movements (other than their left hands) during the tests, the author believes the results are quite satisfactory. If necessary, one can apply further signal processing techniques to improve the

calculated HRV signals. The main goal, however, was to observe the performance of EmotiGO and study the quality of raw PPG signals collected at the temple during tasks that best resemble true daily activities.

GSR signal: The collected GSR signals from EmotiGO and MP150 are shown for one of the participants in Figure 15. At the first sight, the two signals look significantly different. This difference stems from a slow rise in the SCL signal collected at the temple after the beginning of physical exercise. This slope affects the general shape of the signal. To have a better understanding of the nature of the signals in shorter time intervals, an 80-second snapshot of the signals are depicted in Figure 16. Evident by this figure, the two signals are in rather high agreement, although collected at different locations.
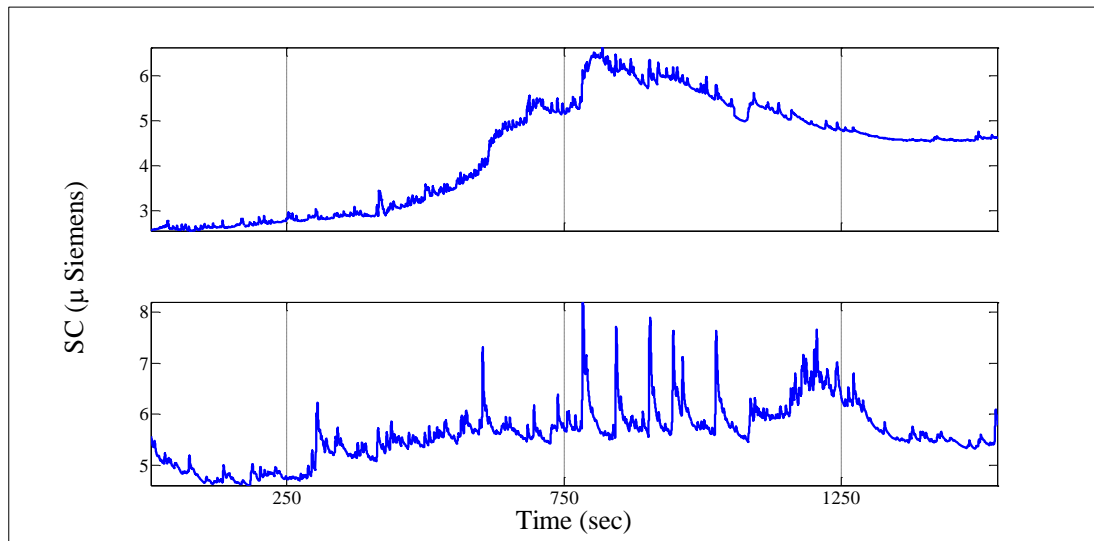
Figure 15. GSR signals collected by EmotiGO (top) and MP150 (bottom)

The increase in the SCL signal collected by EmotiGO is probably due to the use of dry electrodes. Dry electrodes may need longer times so that enough moisture accumulates between the electrodes and the skin, and establish stable electric connection. Since conductive gel is used at the fingertips, the connection

reaches steady state much faster; hence the increasing trend is not significant in the GSR signals collected at the fingertips. Despite the slow increase in the tonic level of EmotiGO's signal, the average Pearson's correlation coefficient for the filtered recordings is 0.71, indicating high similarity between the signals.
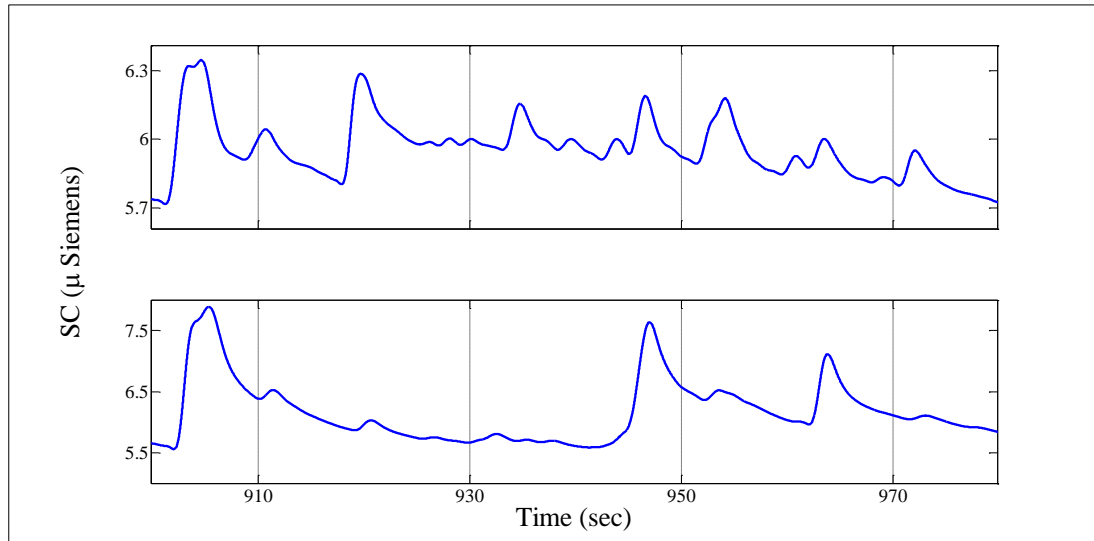
Figure 16. Typical GSR signals collected by EmotiGO (top) and MP150 (bottom)

Physical activity, stress, and mental tasks can increase SCL and elicit SCRs. In this experiment, the collected data at the fingertips and temple showed the increase of SCL from baseline in response to physical activity, startle and Stroop tests. Expectedly, many SCRs were also induced by the tasks. When the tasks ended, for all the participants, GSR signals slowly settled to a plateau during the final recovery period.

SKT signal: The SKT signals recorded by EmotiGO and MP150 (Figure 17) collected at the temple and the thumb, respectively, show weak correlation of 0.11 on average. This discrepancy, however, is in line with literature on the relation between SKT responses at different body locations. The literature shows that the measurement site for SKT signal substantially affects the magnitude and direction

of SKT changes. For instance, the participants in [223] were exposed to a stress-inducing activity while their skin temperature was being simultaneously collected at several body locations. The results showed that, due to stress, temperature significantly increased on the neck over time compared to the control condition, while it significantly decreased on the fingertip.
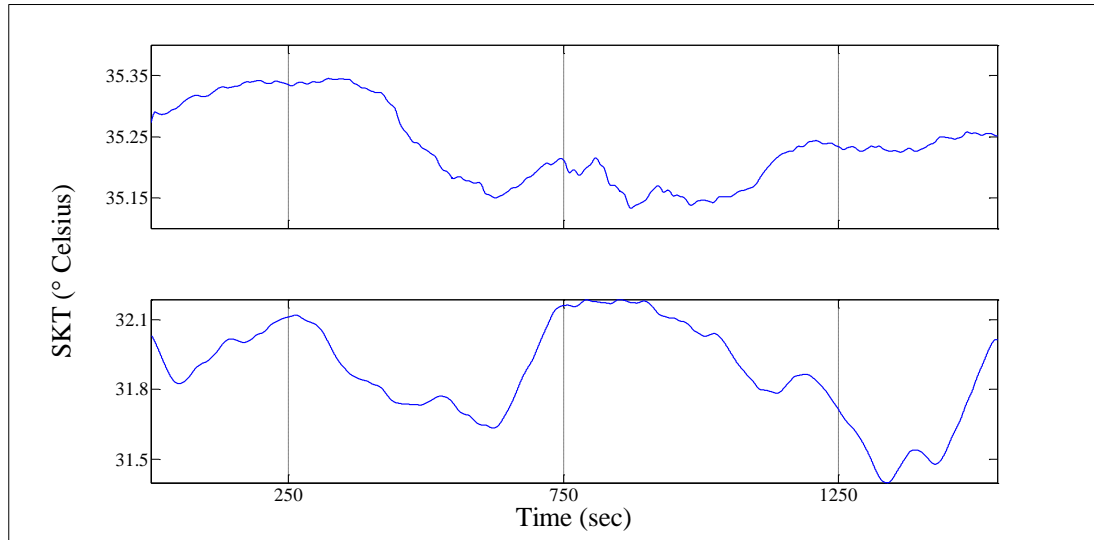
Figure 17. SKT signals collected by EmotiGO (top) and MP150 (bottom)

5.7 Conclusion

The current chapter introduced EmotiGO which is a wireless biofeedback device designed for unobtrusive affect-sensing applications. EmotiGO integrates multiple sensors in a single device which is lightweight and small enough to be worn comfortably without interfering with the users' usual activities. Thanks to its optimal low-power design, EmotiGO can be used for long-term in situ collection of PPG, SKT, and GSR signals. EmotiGO is relatively invulnerable to motion artifacts because of its novel collection site. EmotiGO's open software and API protocol provide an easy way for developers to immediately access data and design custom applications. Another novelty of this device is the use of the nontraditional temple

88

area as the collection site. In Chapter 7, EmotiGO will be used to wirelessly measure autonomic activities of three children with Autism while they interact with the developed tutoring system. The signals collected will be subsequently utilized to model engagement levels of the students.

CHAPTER 6

SMALL-GROUP INSTRUCTION: PEDAGOGICAL AGENT AND ROBOTIC PEER

In the current chapter, the traditional dyadic interaction in the PA-based tutoring system (described in Chapters 3 and 4), is augmented to a triadic interaction by including a robotic peer.

Individuals with ASD receive intervention in various teaching arrangements. The most widely used instruction arrangement for these individuals is direct instruction from an adult in a highly structured and well-planned format delivered in a 1:1 student-to-teacher ratio [29-31, 34]. In direct 1:1 intervention, the instructor delivers prompts to students to teach specified behaviors in a contained setting and isolated from peers. The student responds to the prompts and is then differentially reinforced based on his response [32]. The benefits of this type of intervention regarding behavior development and social functioning of children with ASD have been well-documented [29, 33-35]. This arrangement, however, consumes teachers' time [36-38] and restricts instructional and social integration of these children because the child is removed from peers for instruction [39]. Another major concern is that it ignores natural instruction environment and does not offer opportunities for student-to-student interaction. Relying on 1:1 format may leave these individuals with deficiencies in social interaction skills which hinders their functioning in inclusive environments and general education classrooms

where group instruction is the norm [40, 41]. Besides, the skills learned through this arrangement often fail to generalize to other settings and their peers [42, 44].

An alternative arrangement to 1:1 direct instruction is small-group instruction (SGI) during which several students, usually two to five, are taught in close physical and temporal proximity [45, 46]. SGI enables more efficient use of teacher's time since more than one student is instructed at the same time [30, 41, 48-52]. SGI also better prepares students to function in less restrictive environments which frequently use group arrangements [50, 51]. SGI facilitates normalization as it more closely resembles regular instruction settings [45]. In SGI, students may learn appropriate student-to-student interaction and improve their social and behavioral skills as it provides a context with higher chances of interactions among peers [41, 53-58]. Furthermore, students receive multiple and varied forms of the target behavior within this arrangement, and the skills acquired may better generalize to different people [45, 54, 59, 60].

As a result of exposure to other students' behaviors in the group, an advantage of SGI as compared to 1:1 arrangement is the opportunity for students to acquire new skills/responses through imitation of peers [224-226]. From a behavioral perspective, imitation, which is a type of observational learning (OL), is defined as a child's behavior that is contiguous in time and similar to that of a model [227-230]. For instance, if one shows a thumbs-up in front of a child for the first time and the child emits a similar gesture with his hand immediately following the modeled behavior, it would be considered an imitative performance. An imitative performance may be strengthened through direct reinforcement.

Another benefit of SGI arrangement is the opportunity for students to acquire novel skills or refine behaviors by observing the "consequences" delivered to the peers for their correct or incorrect responses/behaviors [165]. This type of learning, which is another category of OL, is called Vicarious Learning (VL). VL involves children observing others come in contact with stimulus-response contingencies [30, 165, 231, 232], and, thus, learn without directly receiving consequences. An important difference between imitation and VL is that the latter may cause behavior suppression as a result of observing non-preferred consequences received by a model due to a behavior. For instance, a child may vicariously learn not to throw a basketball at a peer after he sees the negative reactions of other students to another peer doing so [233]. Early studies done by Bandura et al. [234-238] demonstrated that children who observed a model received punishment following his aggressive behavior were less likely to match the response of the model and to engage in that kind of behavior. Imitation, on the other hand, can only include an observer's behavior matching that of a model[2].

SGI provides a context in which a student can vicariously learn a new response that is not directly instructed to him by observing teacher's instruction to another student, the student's response, and the subsequent teacher's differential reinforcement according to the student's correct or incorrect response. Therefore, simply by group participation, each student in the group has the opportunity to

_____

[2] Merriam-Webster dictionary defines vicarious as "experienced or felt by watching, hearing about, or reading about someone else rather than by doing something yourself." Historically, the terms "vicarious learning" and "observational learning" have been used interchangeably, although vicarious learning, originally coined by Bandura in 1962 is technically only a subcategory of observational learning which is a broader term also including imitation, delayed imitation, etc.

acquire skills targeted to other students even when he is not directly reinforced for his learning [61]. SGI can be made more efficient if instruction procedures are carefully designed to ensure VL occurs [52, 60, 61, 239-242].

SGI has been shown an effective and efficient arrangement for children with special needs. For instance, Favell, Favell and McGimsey [41] compared the relative effectiveness and efficiency of instructing sight words to participants with developmental disabilities (DD) in a 1:1 format versus in a group format. The results of the study showed that participants in the 1:1 arrangement acquired 11 words in 36.5 hours while the other group acquired 15 words in 15 hours on aggregate. The results showed that not only was the SGI arrangement effective but also more efficient than the 1:1 arrangement in instructing sight words to people with DD.

Stinson, Gast, Wolery & Collins [243] used a time-delay procedure in an SGI format to instruct sight words to four individuals with moderate mental retardation. The results of the study showed that all participants learned their own target words as well as at least half of the words targeted to other students in the group through vicarious learning. In another study in 2008 [165], Ledford, Gast, Luscre, and Ayres instructed sight words to six children with ASD. They delivered the instructions to the participants in groups of two via a CTD procedure. The results demonstrated that all participants acquired all of their target words and five participants vicariously learned all of the non-target words presented to the other participant in their group. Research supports efficacy and benefits of SGI arrangement to instruct a number of skills to individuals with ASD including sight

words [55, 165], expressive labels [244], social skills [245, 246], job functions [244], play [247, 248], inference [244], reading [249], food and drink preparation [250], picture to word/word to picture matching [249], and sign language [251].

For those students with ASD who have the prerequisite skills for group participation, SGI is an effective and efficient instructional arrangement. Certain prerequisite skills are required for a successful inclusion, efficient progress in group settings, and VL including attending to multiple cues such as the target stimulus, the model performing the task, and the corresponding consequences [252, 253]. Many individuals with ASD, however, show restricted interest [254], inattentiveness to surroundings [255, 256], and deficits in sustaining attention [233, 257-259]. These characteristics may narrow the pool of relevant details of the environment, observed behaviors, and consequences, thereby adversely influencing these individuals' proper advancement in such arrangement. These deficits could also result in social exclusion for those individuals in spite of increase in opportunities for social interaction and other potential benefits of group instruction [260, 261].

Robotic instruction has been shown to increase motivation, attention, and time on task in individuals with ASD, which can be particularly helpful in directing these individuals' attention to relevant cues in an SGI arrangement and improving their interest and engagement with the instructional material. In this chapter, the developed autonomous tutoring system, described in the third and fourth chapters, is augmented by inclusion of a humanoid robot that serves the instructional role of a peer for the student. In this tutoring paradigm, the robot adopts a peer metaphor,

where its function is to act as an emulated peer. With the introduction of the robotic peer (RP), the author suggests to augment the traditional dyadic interaction to triadic interaction, to prepare a technology-based SGI arrangement and to facilitate VL.

6.1 Problem Statement and Research Questions

In the remainder of the current chapter, a study is conducted in which a child with ASD as well as the RP attend instructional sessions together (Figure 18). The sessions are guided by the PA which teaches SWR to both group attendants, that is the student and the robotic peer, using a CTD procedure. A group of unknown words is selected for the student from a screening group of words. From this group, two lists of words, L1 and L2, each containing four words, are selected for the student and the RP, respectively. These lists are defined as L1 = {Γ1, Γ2, Δ1, Δ2}, and L2 = {Γ1, Γ2, Π1, Π2}, where Γ1 and Γ2 are two different words that are common between the student and the RP. Δ1 and Δ2 are unique to the student, whereas, Π1 and Π2 are unique to the RP. These two lists are comprised of six different words all unknown to the student.
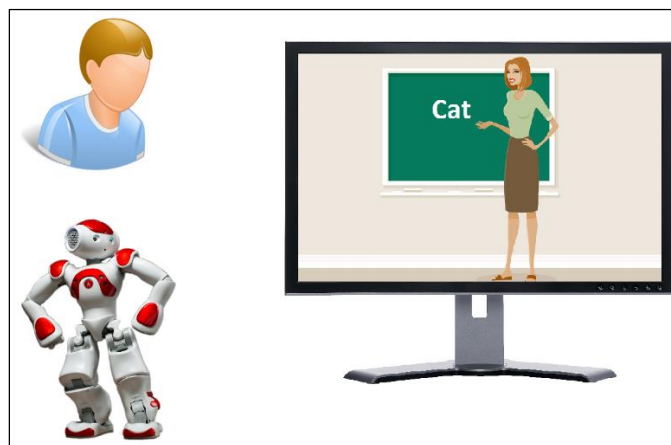
Figure 18. The experimental setup

The PA instructs the words in L1 when it addresses the student. Similarly, the words from L2 are exclusively instructed to the RP. During the instruction of these words, and because the student and the RP participate the tutoring sessions at the same time, three different modes of sight word learning may occur as discussed below.

Experiential Learning (EL): Δ words defined as {Δ1, Δ2} are unique to the student and serve as experiential words, in that, they are not instructed to the RP. Therefore, the student may learn these words only through direct dyadic interaction with the PA.

Vicarious Learning (VL): Π words defined as {Π1, Π2} are exclusive to the RP. The RP simulates learning of these words over time through dyadic interaction with the PA. Since both the student and the RP attend the sessions together, the student sees the words instructed exclusively to the RP, that is, Π words. The student can also hear the RP read those words, and observe the consequences (reinforcement or corrective feedback) delivered to the RP. Therefore, Π words may be vicariously learned by the student only while watching these words being instructed to the RP but not through direct experiential interaction with the PA.

Experiential plus Vicarious Learning (EVL): Γ words, {Γ1, Γ2}, are common between the student and RP. Therefore, the student may acquire these words both experientially through dyadic interaction with the PA, and vicariously through watching the PA instructing them to the RP. These words can be acquired by the student through triadic interaction with the PA and the RP.

This chapter addresses the following research questions. Would the use of PA-delivered instruction and a 5-s CTD procedure taught within a small group arrangement including one autistic student and the RP result in young children with ASD:

(a) reading the target words exclusive to themselves (Δ words) through direct interaction with the PA (EL mode);

(b) reading the target words exclusive to the RP (Π words) through vicarious learning (VL mode); and

(c) reading the target words common between themselves and the RP (Γ words) through direct interaction with the PA as well as vicarious learning (EVL mode)?

Additionally, the questions regarding the relative efficiency of EL and EVL modes, as well as the generalization and maintenance of words acquired though the three learning modes are addressed.

6.2 Method

6.2.1 Participants

Three children with ASD were recruited for this study. They were all male, with age range of 6-8 years. Student I was a biracial boy of 8 years and 4 months of age at the beginning of the study. He was administered, at the age of 4 years, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) and obtained a standard score of 75 for verbal IQ and 49 for performance IQ. He was also diagnosed with Anxiety Disorder Not Otherwise Specified and a moderate to severe receptive-expressive language delay.

Student J was a biracial boy of 8 years and 6 months of age. He was administered the WPPSI-III at the age of 3. His performance IQ score was 82. He demonstrated below average cognitive skill. On the development profile 3 (DP3), he scored 103, 61, and below 50 for adaptive behavior, social emotional, and cognitive tests, respectively. Student V was a white boy of 6 years and 11 months of age. He was administered the Autism Diagnostic Observation Schedule (ADOS-2) and the WPPSI-IV at the age of 5. He was diagnosed with high functioning autism and full scale IQ of 107. Approval to conduct the current study was obtained from the university's institutional review board.

6.2.2 Settings and Materials

The researcher conducted all sessions within a one-to-one instructional arrangement in a room within our laboratory. The experimental sessions occurred two to three times a week over a 4-month period. The virtual classroom and the PA were displayed on a 26-inch computer screen. During the sessions, the participant and the RP sat on chairs faced the screen with a little orientation toward each other to ensure the student can see both the screen and RP. The computer screen was located on a table and adjusted to an appropriate height for the participant's eye level. Words were displayed with a 90 pt. Arial style font and with high contrast to insure visibility. The Words were presented in lower case letters to simulate their most common occurrence in books and reading materials. As the proposed educational software is equipped with ASR technology, and communicates through natural language with users, neither a mouse nor a keyboard were provided for the participant. A webcam was mounted on the screen top to record the interactions throughout the experiment for the purposes of offline

analysis, inter-observer reliability, and procedural integrity. The computer's sound volume was set at a moderate level via a built-in speaker at the back of the screen.

At the beginning of each session, the experimenter launched the educational software and an application which was responsible for recording the video/audio stream coming from the webcam. Along with that stream, the researcher simultaneously recorded the content of the screen (the PA and virtual classroom) in a picture-in-picture format. On these recordings, one can observe the reading material presented to the participants and their response, at the same time.

Although the autonomous tutoring system was capable of automatically recognizing words read by the participants, a researcher monitored the sessions and advanced the instruction by listening and manually determining whether the participants read the words presented correctly or not. Since the author planned to compare the efficiency of EL and EVL modes, a high procedural integrity of the experiment was highly desired. This capability was added to the software developed previously in order to avoid any confounding variables in the data, since the accuracy of automatic speech recognition was less than perfect, as shown in the results in the fourth chapter. All other activities of the tutoring system, including the PA's activities (i.e., body gestures, facial expressions, delivering feedback and reinforcement, etc.) and the RP's activities (i.e., body gestures, reading words, etc.) were still autonomous as the tutoring system was capable of delivering perfect procedural integrity in those domains.

6.2.3 Screening

The target words were selected from the Dolch and Brigance functional word lists [178, 179]. Prior to the experiment, each participant went through a series of tests to identify words that would be targeted during intervention. The first stage of screening was performed by the researcher in a one-to-one format in the experiment room. The researcher presented the words, printed on flash cards, and asked "What word?", and noted the words that the participants either read incorrectly or did not read at all. During screening, the participants did not receive reinforcement or feedback for their responses.

In the second stage of screening, the words identified in the previous stage as unknown to each participant were again presented to them, similar to the previous stage. The only difference was that the words were displayed on the computer screen, instead of flash cards. Similarly, no feedback was provided. This stage was repeated three times on three different days to ensure the words were unknown to the participants. The words order was randomized each session. From the words that were never read correctly, the researcher randomly chose 18 for each participant to be included in the study. Since the experiment was conducted during summer, there was no concern about the participants receiving instruction of those words at school. Additionally, the parents agreed not to instruct these words to the participants at home for the duration of this experiment. The parents were also kept blind to the study hypotheses.

6.2.4 Experimental Design

To evaluate the efficacy and efficiency of the learning modes, a multiple-probe across-word-sets design, replicated among three participants, was employed [262]. Instruction was staggered across word sets. For each participant,

the researcher started by collecting baseline data on reading performance across all three word sets. After three data points across three days, the instruction started only on one of the word sets. As soon as the participant reached criterion on the first set, the researcher started collecting data on all three sets for three sessions. Upon completion of this stage, the researcher began delivering instruction on the second word set. This procedure was carried on until the participant reached criterion on the third word set. After the participant reached criterion on the third word set, the researcher went through another data collection on all three word sets. The criterion was defined as scoring 100% accuracy for three consecutive sessions. Additionally, a pre/posttest design was employed to assess participant's ability to maintain over time and to generalize the words acquired to a non-laboratory setting. Furthermore, to evaluate the acquisition level of words through vicarious learning (i.e., Π words through VL mode), a pre/posttest experimental design was designed.

6.2.5 Dependent Measures

Three primary dependent variables were collected: percentage of words read correctly (i.e., percent reading accuracy) acquired through EL, VL, and EVL modes, number of sessions to criterion, and number of errors made by the participant. A word was scored as correct, if the student stated the word presented by the PA within 5 seconds. Sessions to criterion were scored as the number of session from the first intervention session to the participants third consecutive days at 100%. Incorrect response and no response were both counted as error. Additionally, several secondary measures were collected such as generalization

and maintenance of words obtained through EL, VL, and EVL modes. Data was collected by the researcher in the experiment room.

6.2.6 Feet-wet Session:

Prior to the beginning of baseline sessions, to address the novelty factor, the actual PA, virtual classroom, and the RP were presented to the participant. The researcher described the procedures and the role of PA and RP to him. The participant was taken to the experiment room and introduced to the RP and its capabilities for the first time. A 10-minute long demonstration of the robot capabilities including speech, regular body movements, playing music and dance was prepared for the participant. The participant was then allowed to touch the robot and familiarize himself with it.

6.2.7 Baseline Sessions

Each baseline session consisted of 18 trials, one trial for each word. At the beginning of each baseline session, the researcher launched the application and entered the participant's name. The PA started by greeting the participant by name, and stating "we are reading some words today. I need your attention." After 3 seconds, it presented one of the words at random, and asked, "What word?" The word was displayed for 5 seconds on the PA's chalkboard and then disappeared. No feedback or reinforcement were provided to the participant. Before the PA advanced to the next word, there was a 1-second inter-trial interval. This procedure continued until all 18 words were presented. During the baseline sessions, the RP was seated beside the participant and programmed to do nothing.

6.2.8 Instructional Sessions

During instruction, the PA implemented a CTD procedure to instruct sight words to the participant and the RP in turn. In each session, the PA presented three trials per each of the four words dedicated to each group attendant (i.e., one 0-s delay trial, two 5-s delay trials). Therefore, each group attendant (i.e., the participant and the RP) received 12 trials in each session.

The PA started with a greeting and obtaining the RP's attention by asking "NAO, are you ready?" At this moment, the RP responded back "Yes." Then, the PA asks the same question from the participant by calling his name (e.g., Alex). When the participant responds, the PA emitted a group attentional cue by saying "Alright, everybody, look at me." After a 1-second pause the PA started the first round of trials (i.e., 0-s delay) for each word and group attendant. A word randomly chosen from list L2 (e.g., "help") appeared on the PA's chalkboard while the PA stated "NAO, what word?", followed by an auditory model, and waited 5 seconds for the RP to respond. The RP was programmed to randomly choose among 3 options, read the word correctly, read the word incorrectly, or say nothing. Rather than always correctly reading the words presented, the RP was programmed to demonstrate learning of the words over time to simulate the most natural situation in a real group setting. Therefore, the likelihood of reading the words correctly was increased over time. The PA responded to correct responses with verbal praise (e.g., "good work," "excellent," "nice job") and to errors or no response with corrective feedback (modeling the word again). In case of receiving reinforcement from the PA, the RP was programmed to demonstrate happy gestures. This differential reaction was programmed into the RP's behavior repertoire in order to

draw the participant's attention and to facilitate discrimination of consequences for the participant, and thereby vicarious learning.

After the first trial for the RP and a 1-second inter-trial interval, a word randomly chosen from list L1 appeared on the chalkboard and the PA stated "Alex, what word?" The PA then waited 5 seconds for the participant to respond. Subsequently, the PA reacted to correct responses with verbal praise and to errors or no response with corrective feedback.

The PA then carried on this procedure alternatingly with the RP and the participant until all the words of list L1 and list L2 were instructed to the RP and the participant, respectively. After a 1-second pause, at the beginning of the second round of trials, the PA emitted another group attentional cue by saying "Everybody, look at me." During the second and third rounds of trials, the PA used procedures identical to those used with the first round, with the exception that a 5-s delay interval followed the PA's presentation of the words. Across sessions, the presentation of words was randomized to prevent participants from memorizing the words by their order. Each session concluded by the PA providing non-contingent verbal praise to the participant, such as "You are doing great!" or "Keep up the good work", while clapping for him. The RP also delivered non-contingent praise to the participant by orienting his head toward him and saying "Good job."

During the instructional sessions, the experimenter observed the interaction and silently collected data on the participant's reading performance and advanced the instructional trials by determining whether the participant read the words presented correctly or not. He did not intervene in any way since the participant

was expected to work independently. A response was only recorded as correct if the participant read the word correctly in the provided 5-second time slot.

6.2.9 Post-instruction Probes

When the participant reached criterion on each word set, an assessment phase was conducted which lasted three sessions. Each assessment session was exactly identical to the baseline sessions, where the participant was required to read the presented stimuli and did not receive any word modeling, feedback, or reinforcement from the PA. This phase was specifically carried out in order to investigate whether the reading performance remained as accurate without further training. These probes were repeated after the participant reached criterion for each word set.

6.2.10 Maintenance Session

An additional follow-up maintenance probe was conducted two months after completion of the last post-assessment session for the third word set. The procedure was identical to the baseline sessions. The goal was to examine whether the participants maintained the acquired words over time.

6.2.11 Generalization Session

A final probe was conducted at the participant's home and with his parent in order to test whether the acquired words would transfer outside the experiment room. This session was similar to the screening session performed done by the researcher before the experiment began. The parent was instructed to present the words printed on flash cards, one-by-one, and to record correct and incorrect responses without delivering feedback or reinforcement.

6.3 Results

EVL and EL words: Percent reading accuracies for EVL and EL words across baseline, CTD instructional sessions, and post-instruction probes are displayed in Figure 19, Figure 20, and Figure 21 for students I, J, and V, respectively. Visual analysis of these figures reveal that prior to the intervention none of the participants verbally identified the words, neither the EL nor the EVL words. Upon introduction of the small-group PA-delivered instruction, the percent accuracies showed an immediate change in the therapeutic direction for all three word sets and all the participants.

The participants met criterion in relatively small number of sessions. Student I reached criterion for the first set of EVL and EL words in 4 and 5 sessions, respectively. He reached criterion for both EVL and EL words in the second set in 5 sessions. He also reached criterion for the EVL and EL words in the third set in 8 and 7 sessions, respectively. He made 11 errors in total in the EL mode and 20 errors in the EVL mode. Among the participants, student I was the only one that his EVL mode was worse than his EL mode, mainly because he was not learning one of the EVL words in the third word set for several sessions until finally he started reading the word correctly. Although at the beginning of the study words with equal difficulty were chosen, one could only hope that the words are actually equally hard for an individual.

Student J met criterion for EVL and EL words of the first set in 5 and 7 sessions, respectively. He did so for the second set in 4 and 6 sessions while it took 5 and 7 sessions for him to reach criterion for the EVL and EL words in the third set. He made 21 errors in total in the EL mode and 11 errors in the EVL mode.

Student V reached criterion for EVL and EL words of the first set in 4 and 5 sessions, respectively. He reached criterion for both word types of the second and third sets in 5 sessions although with a faster pace for the EVL words in both sets. He made 13 errors in total in the EL mode and 6 errors in the EVL mode. This pattern depicts experimental control in a multiple-probe across-word-sets design.

As depicted in these figures, 5-second CTD procedure delivered in the SGI arrangement using a PA and an RP was effective in instruction of sight words to three children with ASD. During the maintenance and generalization probes, all three participants read the words they acquired through EL and EVL learning (i.e., Δ and Γ, respectively) with 100% accuracy.

VL words: Table I, Table II, and Table III summarize the responding percentage of sight words by the participants presumably acquired through vicarious learning. Data in the white section and gray section represent reading accuracy before and after the instruction to the RP, respectively. The instruction (to the RP) on each word set began at the verge of white-gray shading. During the pretest, all participants consistently identified VL words (which were later taught exclusively to the RP) at 0% accuracy.
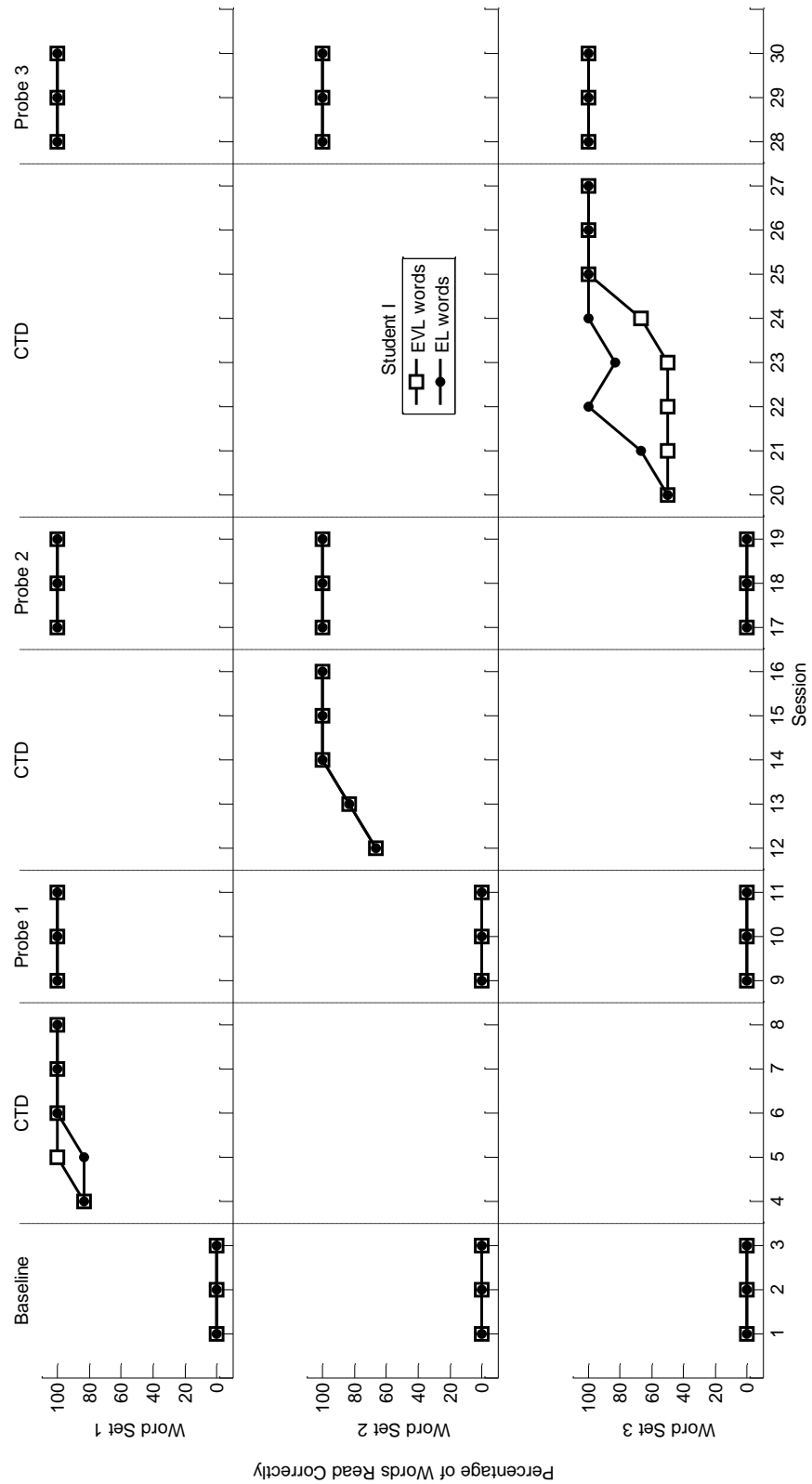
Figure 19. Percentage of student I's correct responses during baseline, CTD instructions, and assessment probes.
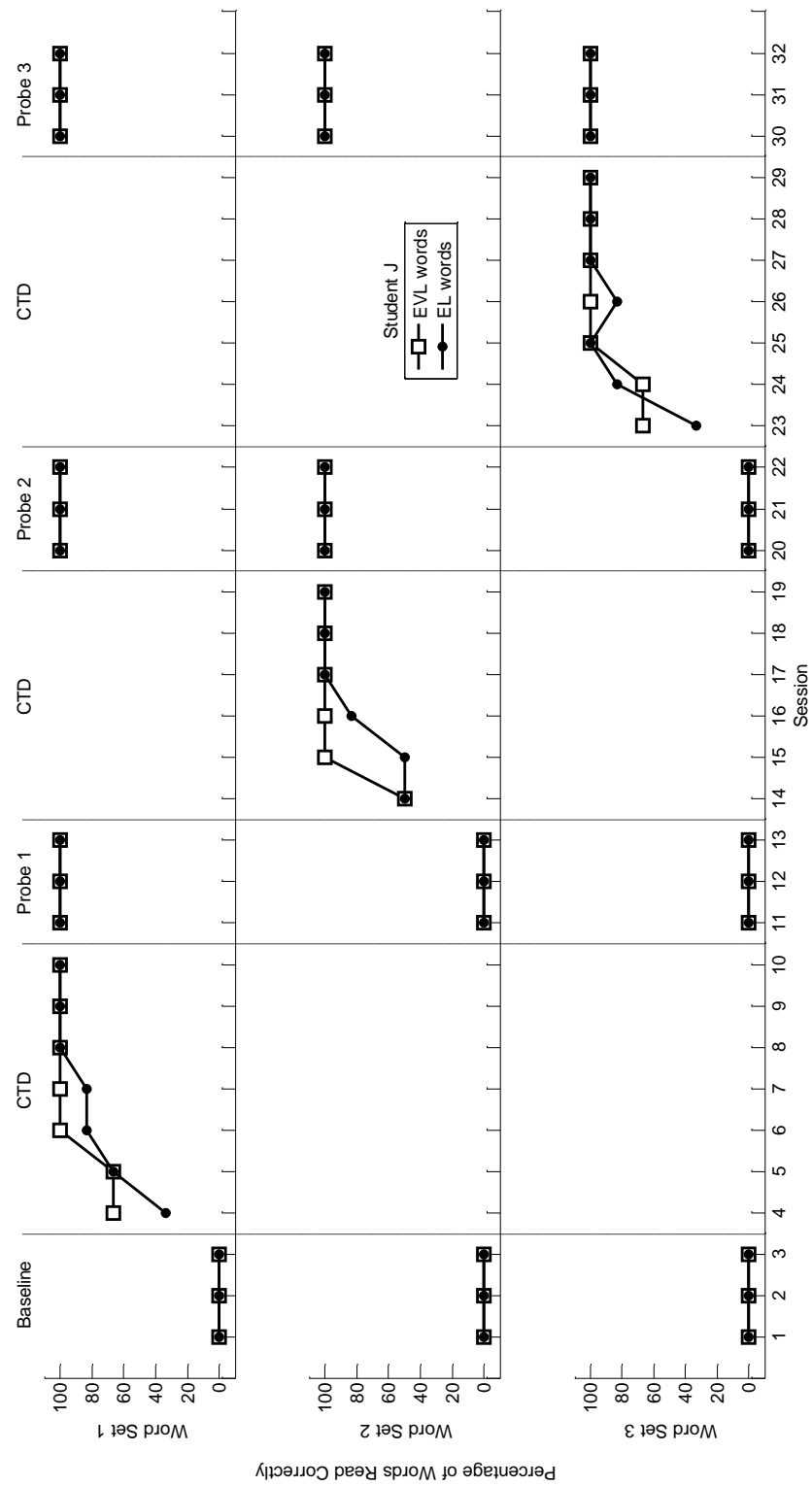
Figure 20. Percentage of student J's correct responses during baseline, CTD instructions, and assessment probes.
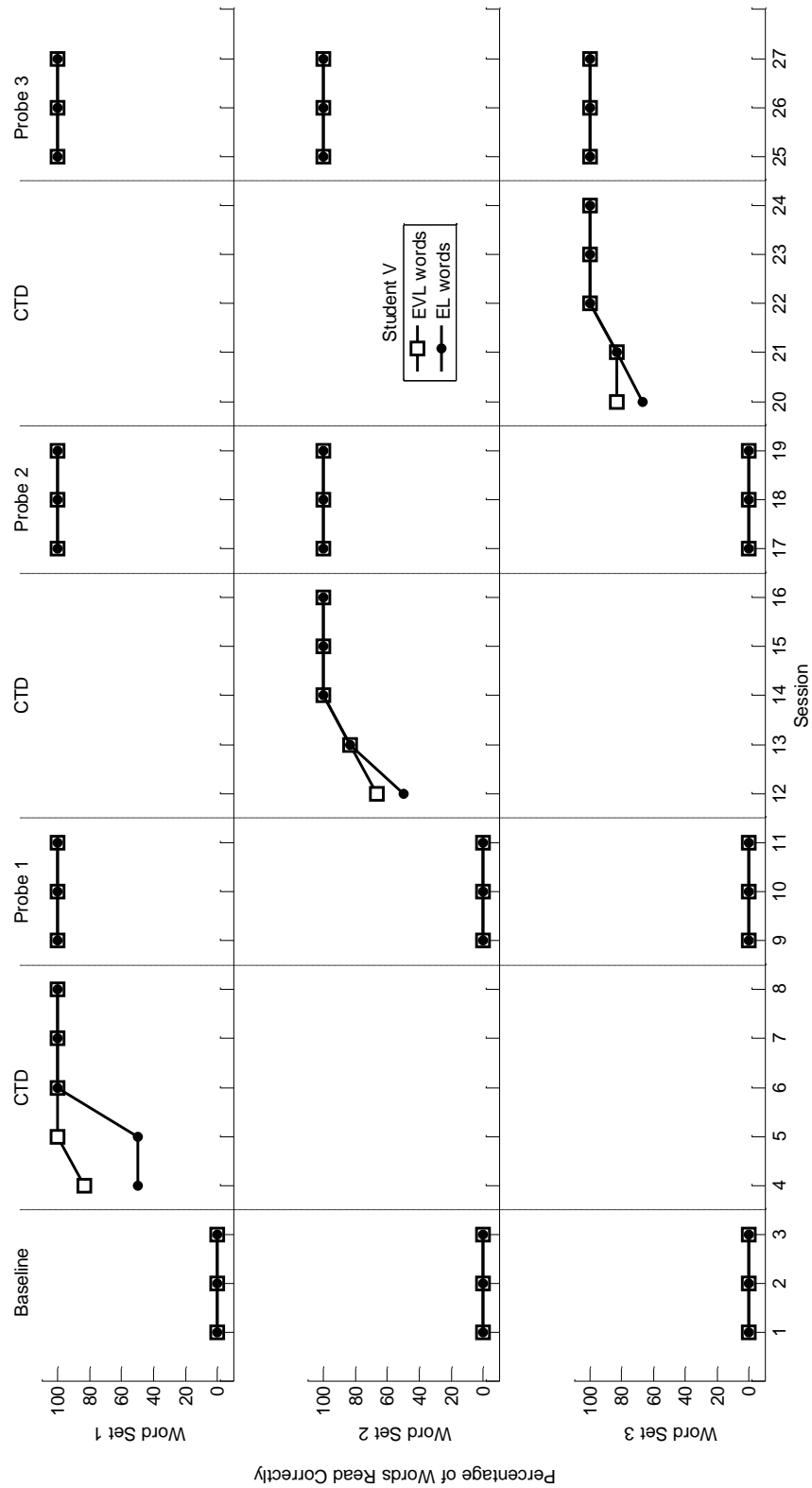
Figure 21. Percentage of student V's correct responses during baseline, CTD instructions, and assessment probes.

Table I. Student I's correct responding percentage of VL words

Word Set	Pre-test	Probe 1	Probe 2	Probe 3	Maintenance	Generalization
1	0%	100%	100%	100%	100%	100%
2	0%	0%	100%	100%	100%	100%
3	0%	0%	0%	50%	50%	50%

Table II. Student J's correct responding percentage of VL words

Word Set	Pre-test	Probe 1	Probe 2	Probe 3	Maintenance	Generalization
1	0%	100%	100%	100%	100%	100%
2	0%	0%	100%	100%	100%	100%
3	0%	0%	0%	100%	100%	100%

Table III. Student V's correct responding percentage of VL words

Word Set	Pre-test	Probe 1	Probe 2	Probe 3	Maintenance	Generalization
1	0%	100%	100%	100%	50%	50%
2	0%	0%	100%	100%	100%	100%
3	0%	0%	0%	100%	100%	100%

According to Table I, Table II, and Table III, all participants acquired some of the RP's exclusive words (i.e., Π words) through vicarious learning of PA-delivered instruction to the RP. Although student I did not do as well in his EVL words as in his EL words, he performed quite well in his VL words. Student I acquired 5 of the 6 VL words for which he demonstrated perfect maintenance and generalization. Student I's reading accuracy in maintenance session was 83.33% on average.

Interestingly, student J learned, maintained, and generalized all the Π words. Student V acquired all the VL words although he did not show generalization and maintenance for one of the Π words in the third set. Student J and V maintained their Π words (acquired through VL) with 100% and 83.33% accuracy, respectively. All three participants generalized, to their home setting, all the Π words they had maintained.

6.4 Discussion

A main goal of the study in this chapter was to determine the feasibility and effects of using PA-delivered SWR instruction, based on a CTD procedure, taught within a small-group arrangement consisting of one child with ASD and a robotic peer. The data illustrated that all three participants acquired, maintained for two months, and generalized to natural context, all the words they were instructed through EL and EVL modes.

Furthermore, although a functional relationship was not established (because of the nature of the experiment), all three participants were able to acquire a high percentage (94.44% on average) of the non-target words (Π words) that were exclusively instructed to the RP through vicarious learning. In addition, the results indicated that the participants successfully maintained and generalized a substantial amount of the non-target words (88.88%). It is noteworthy that, even though the participants did not learn all the non-target words, they may probably acquire these words more efficiently when those words are targeted for direct instruction to them.

Other than one EVL word for student I, the data show that not only did the participants learn extra words through vicarious learning but also sight words may

be acquired more efficiently and with less errors by addition of a vicarious component to experiential learning. The promising results of this study, and the positive effects of vicarious component of the instructional package in particular, may be explained by the addition of the robot which may have increased the saliency of the required response and reinforcement by drawing the participants' attention, interest, and motivation. This possibility is of paramount importance since children with ASD show deficits in attending to the relevant elements of the instruction.

During the instructional sessions, the participants followed the routine established by the PA, practiced turn taking, demonstrated patience while waiting for their turn without any inappropriate behavior such as speaking or reading the words directed to the RP. No behavior management problems were encountered. Although it is rather unusual for this population, upon completion of the session and when the RP verbally praised their performance, there were numerous times when the participants showed their appreciation by responding "Thank you" or "Nice job", etc. One of the participants almost always greeted the robot when he arrived at the experiment room, touched the robot eyes and fingers, and hugged and kissed the robot goodbye when he was leaving the room. Another participant gradually began imitating the RP's happy gestures with him and also when he, himself, was reinforced by the PA for his correct responding. His parent confirmed that he had not shown these gestures before this experiment. Observation of such an imitative behavior is particularly interesting since he was not instructed or reinforced for such behavior.

For those students with ASD who have the prerequisite skills for group participation, SGI is an effective and efficient instructional arrangement. Certain prerequisite skills are required for a successful inclusion and efficient progress in group settings including attending to the teacher and peers, imitation, tolerating intermittent attention, contingent turn-taking, waiting without disruptive behavior (e.g., leaving group and inappropriate verbalization), staying on-task and maintaining attention, and compliance [30]. Many individuals with ASD, however, show deficits in at least some of these skills [230, 239, 252, 263-270] which, in turn, could adversely influence proper advancement of not only themselves but that of other group members as well in such arrangement.

There is evidence that many individuals with ASD do not readily learn by imitation [230] and are deficient in many areas associated with VL [233, 271]. VL involves observing and imitating a model but also, importantly, requires simple and complex discriminations among consequences received by the model [61, 230, 233, 239, 252, 264, 265, 272-274]. Twenty-one studies reviewed in [230] concluded that some children with ASD appear to be deficient in imitative skills. They also show poor visual attending and may be deficient in making simple and complex visual and auditory discriminations [239, 264, 265, 275-277]. These deficits would result in poorer performance and social exclusion for those individuals in spite of increase in opportunities for social interaction and other potential benefits of group instruction [260, 261].

For example, Koegel and Rincover [54] studied the effects of transferring students with ASD from a 1:1 instruction format to a group instruction format. They

114

first instructed a number of basic group behaviors assumed necessary for group participation (such as attending to the teacher upon command and imitation) to eight children with ASD in a 1:1 format, and then transferred them to a group arrangement. The investigators found that the behaviors acquired in the 1:1 setting were minimally and variably performed in the group format, even in a group as small as one teacher and two students. Additionally, even prolonged group participation did not seem to produce much change as the students failed to acquire new behaviors during a month-long participation in the group setting. The investigators argued that this failure may be due to poor generalization skills and stimulus over-selectivity [278-282] in individuals with ASD.

Although group instruction can be an effective and efficient instructional arrangement for individuals with ASD, it seems necessary to pay careful attention to how to transit from 1:1 to group settings. Special procedures should be identified and be employed to ensure a smooth shift and avoid abrupt transition for many of these individuals. Added emphasis must be placed on developing strategies that prepare these individuals for group learning and equip them with skills required for multi-student contexts [283]. The developed system in this chapter presents an intelligent tutoring system that resembles a small-group instruction arrangement for children with ASD. Previous literature [182, 283] suggests that generalization is enhanced when instruction setting is similar to extra-training setting. Therefore, learning through interaction with this system may properly prepare these individuals for group participation. This system provides ample opportunity to

practice group skills which, in turn, increases the likelihood for generalization of those skills to group learning arrangements.

In this study the RP was programmed to show gradual acquisition of target words to resemble the most frequent situation in a natural group setting. Some interesting questions that may be addressed in future are (a) how would skill acquisition and vicarious learning be affected if the RP simulated a competent model which would result in higher number of reinforcement than corrective feedback, and (b) how, if possible at all, can a sense of competition be stimulated in the students while attending a small group arrangement including a robotic peer.

Future studies should investigate the feasibility and effects of other instructional procedures (e.g., other response prompting strategies) to instruct sight word reading as well as other skills/behaviors to children with ASD through such small-group interaction system. Furthermore, the effectiveness and efficiency of other responding methods such as choral responding can be studied and compared against individual responding. Future investigations should include larger number of participants and with wider spectrum of behavioral characteristics and different diagnoses to strengthen the extension of these findings to more versatile populations. With all the current questions regarding how to include children with ASD in regular instructional settings, and considering the importance of vicarious learning, an enhanced understanding of how these individuals acquire skills in such arrangements and through vicarious learning seems to be of substantial importance for future research.

CHAPTER 7

PHYSIOLOGY-BASED MODELING OF ENGAGEMENT DURING
NATURALISTIC INTERACTION WITH TUTORING SYSTEM

Recent decades have seen increased recognition of the interplay between affect and learning [62-64]. A number of researches have argued for a set of affective states that influence learning such as boredom [65, 66], engagement [65], and frustration [64, 68]. These affective states are believed to accompany and influence learning. For example, increased levels of boredom are negatively correlated with learning, whereas increased levels of engagement in the learning process are positively correlated with learning [69]. Given this link between affect and learning, a learning environment that's sensitive to a learner's affective states will presumably enrich learning by increasing motivation and engagement.

Several tutoring systems, designed for typically-developing individuals, are increasingly incorporating affect responsiveness into their pedagogical strategies [70-73]. An example is Affective AutoTutor [72] which is an affect-responsive dialogue-based tutoring system for computer literacy. Affective AutoTutor was designed to detect students' emotions and use this information to guide response selection to help children regulate their emotions during learning [71]. This tutoring system was reported to lead to better learning outcomes and to higher levels of engagement than its non-affect-aware counterpart.

An affect-responsive tutoring system can be more effective specifically for individuals with ASD who exhibit deficits in emotion regulation and expression

[275, 284]. To build an affect-responsive tutoring system, one must first enable the system to detect user's affective states. Several modalities such as facial expression [75], vocal intonation [76], gestures [77], and physiology [78] have been utilized to evaluate affective states for typical individuals. However, individuals with ASD often have communicative impairments (both nonverbal and verbal) particularly regarding the expression of affective states [79, 80, 275]. These vulnerabilities place limits on traditional auditory and visual methodologies. Physiological signals, however, are not necessarily directly impacted by the difficulties in emotional expressions in people on the autism spectrum [81, 285, 286].

Individuals with ASD often react outwardly in ways unlike their typically-developing counterparts. They, for example, might smile when they are actually in pain, or might show a neutral expression when they are enjoying an activity. These individuals usually have problems expressing their stress, frustration, and emotions in general. It is well accepted that these individuals may have different emotional expressions than what is truly happening on the inside. An autistic person can appear very calm outwardly while his/her heart rate is unusually high [287]. An individual with ASD can be highly aroused, according to his/her electrodermal activity measurements, without any outward signs of stress [288]. In the case of ASD, to know what is truly going on, relying on what is shown on the outside could be misleading. Monitoring physiological signals, however, is an avenue to receive emotional information that may not be apparent through outward expressions.

The literature supports creating affective models for individuals with ASD using their physiological signals [189, 289]. In [289], physiological signals from six children with ASD were used to infer high and low levels of three affective states, i.e. anxiety, liking, and engagement. Various features of several physiological signals including cardiovascular, electrodermal, and electromyogram activities were examined. The physiological signals were recorded when children interacted with computer tasks such as pong and anagram. The indices were derived by applying a series of signal processing algorithms on the collected signals. For training the models, the affective labels were derived from subjective reports provided by a therapist experienced in working with autistic individuals. After applying machine learning algorithms, the investigators reported reliable prediction with approximately 83% accuracy on average across all participants and affective states.

In a follow-up study [189], the same investigators designed an affect-sensitive robot-based basketball, wherein a robot adapted its behaviors in real time according to the preference of a child with ASD. An undersized basketball hoop was attached to the end-effector of a robotic manipulator, which could move the hoop in different directions with different speeds. The children were instructed to shoot a required number of baskets into the moving hoop within a given time. Three robot behaviors were designed with different speeds and directions of the manipulator and background music. Using the affective models, created in the previous study, the robot implemented a session of affect-sensitive robot-based basketball specific to each child with regard to the game configuration. The robot

kept learning each participant's preference over time using a reinforcement learning algorithm according to the records of the robot's actions and the consequent liking level of each participant. It selected appropriate behaviors to present in the task to maintain the child at a high level of liking during the session. The results indicated that the individual's preference was learned by the robot with high probability across all the participants. The robot chose the most preferred behavior 72.5% of the time on average across all the participants. The moderately preferred and least preferred behaviors were also chosen 16.7% and 10.8% of the time, respectively. Furthermore, the liking level of five of the participants was significantly higher in the affect-sensitive robot-based basketball sessions comparing to the non-affect-sensitive sessions (i.e., when the robot selected behaviors randomly).

In this chapter, the author sets out to extend the seminal work done by Liu et. al in [289] with the focus on the user's engagement level based on the physiological signals collected by EmotiGO during the triadic interaction with the tutoring system including the PA and RP. A main difference of the study in the current chapter and the work done in [289] is that, instead of using wired collection of data through a tethered interaction with the system, the author utilized the newly developed and validated measurement system, EmotiGO, to unobtrusively collect autonomic activity of the participants. Tethering participants to a computer station with sticky electrodes and dangling wires may in itself influence their affective experience [290]. During the instructional sessions, EmotiGO was worn by the participants and collected their physiological signals without interfering with their

usual activities while interacting with the tutoring system. Another major difference of the current study and [289] is that this chapter focuses on the feasibility of affective modeling and accuracy of the models based on a smaller set of physiological signals, i.e., PPG, SKT, and GSR signals only.

Finally, in the current chapter, for the purpose of training affective models, spontaneous (as opposed to experimentally-induced) affective experiences are used. These affective states arose in the participants as a consequence of naturalistic interaction with the developed tutoring system. Most of the previous literature in the field of affect classification rely on experimental methods to artificially induce emotion. It is, however, not clear whether the classification performances achieved using models trained based on experimentally-induced affect will generalize to naturalistic situations [291]. The present chapter attempts to collect such naturalistic corpus of affect/physiological data, and investigate the feasibility of and challenges involved in such affective modeling approach.

7.1 Procedure and Data Collection

The participant wore EmotiGO 15 minutes ahead of their instructional sessions, allowing enough time for EmotiGO sensors to acclimate and settle to a steady-state condition. After the adaptation period, the participant was seated in a chair in front of the screen adjusted to his eye level. The researcher asked the participant to relax quietly. He then launched the EmotiGO GUI for the collection of physiological signals from the participant, and collected a 5-minute baseline of physiological data. As soon as the baseline was over, at the beginning of the instructional session, the experimenter launched the educational software and the application responsible for recording the video/audio stream coming from the

webcam. The instructional sessions were videotaped using the webcam placed on top of the screen for the purpose of offline coding of the participants' engagement levels. All physiological data points as well as video frames were timestamped in order for data synchronization and post-processing.

7.2 Engagement Ratings

The definition of engagement is context-specific. However, in the learning domain, engagement is typically defined as sustained attention to the instructional material. Engagement is an important factor for students with ASD to acquire various communication and social skills, and is positively correlated with their learning gains in academic settings [292].

Assessing one's ability to read and recall a list of words is a relatively straightforward process. With somewhat more difficulty, one can also test maintenance of those words across time and their generalization to different settings. It is much harder, though, to assess how one feels during a certain task (for example, his/her engagement level in this case). Affective experiences are typically rated using questionnaires and self-reports (e.g., [293, 294]). Despite the general acceptance of questionnaires, for children with ASD, self-reporting affect is never considered highly valid due to their incapability in reliable identification and expression of their affective states [295]. Furthermore, questionnaires cannot be used unobtrusively and continuously as they require interrupting the learning experience.

An alternative approach to assessing engagement levels is through analyzing gaze behavior. For example, in the context of market research, gaze fixation duration has been used as an indicator of engagement level of people with

certain stimuli [296]. Similarly, in the context of conference meetings, head pose orientation (as an index of where the person looks at) has been used to identify the focus of attention [297, 298]. In the context of learning, gaze direction is an important signal to study the focus of attention and measure the learner's engagement level. When the learner is on task, the gaze is mainly directed toward the instructional material, whereas it usually wanders off from the instructional material in an off-task state. Other eye-related behaviors have also been studied and utilized as autonomic indices of engagement such as spontaneous eye-blink rate [299-301] and changes in pupil diameter [302]. In this study, the author used duration of gaze fixation on instructional material as an indicator of engagement level with the underlying assumption that participants are engaged with the material they are staring at.

For the purpose of rating gaze fixation duration, the video recordings of each instructional session were segmented to 1-second intervals. The sessions were coded by scoring on-task if the participant looked at the screen or the RP. It was, on the other hand, coded as off-task if the participant's gaze fell on anything else. The scores were then summed up over windows with the length of 20 seconds to produce the total number of occurrences of these behaviors and the total duration of the participant's engagement. The on-task durations were subsequently normalized over the total number of possible occurrences to yield the final ratings of the participant's engagement level during a specific window as a number in the interval of 0% to 100%. For more information about this approach refer to [303, 304].

Two human coders with relevant expertise in this domain, were recruited to rate the participant's engagement levels using the videos recorded from the instructional sessions. They were first trained to obtain perfect agreement over several videos recorded during a pilot study with a different group of people. All video recordings from the instructional sessions were then rated by one of the coders to train affective models. In order to verify the reliability of the coding and to ensure inter-rater reliability, 30% of the video dataset was selected (evenly randomized among the three participants), and independently coded by the second coder. The agreement level of the two coders was 95.08%.

7.3 Physiological Features

Quick sensor movements usually induce motion artifacts that are of high frequency nature. This kind of artifacts typically happens when the user touches the bio-sensor. For instance, there were a number of instances where one of the participants touched EmotiGO's frame while rubbing his eye. This kind of disturbances is usual in naturalistic settings where the participants are not required to limit their body movements. Therefore, prior to the analysis of physiological data, all three signals underwent low-pass filtration with cut-off frequency of four Hz to remove any motion-induced artifact in the signals as well as measurement noise. After noise and artifact removal, various features were individually computed from each physiological signal (i.e., PPG, SKT, and GSR) using the Augsburg Biosignal Toolbox [305] for MATLAB. Below, features extracted from each signal are described.

7.3.1 GSR

The Augsburg Biosignal Toolbox (AuBT) applies a number of preprocessing steps such as low-pass filtering and normalization. After the preprocessing stage, this toolbox decomposes the GSR signal to its tonic and phasic responses using the deconvolution approach proposed by Benedek and Kaernbach in [306]. The phasic component of GSR, also known as skin conductance response (SCR), is the quick changes in the signal. SCR is considered to carry affect information as it signifies responses to internal/external stimuli. The tonic level, also known as skin conductance level (SCL), on the other hand, changes slowly and exists in the absence of any specific environmental or external event. Figure 22 depicts a typical instance of GSR waveform consisting of both SCL and SCR components. The decomposition algorithm detrends the GSR signal by connecting every two consecutive local minima of the GSR signal together, and creating a piece-wise linear estimation of the SCL signal. By subtracting the computed SCL from the original signal, the SCR signal is obtained. In Figure 22, the decomposition of the GSR signal into its SCL and SCR components is illustrated.
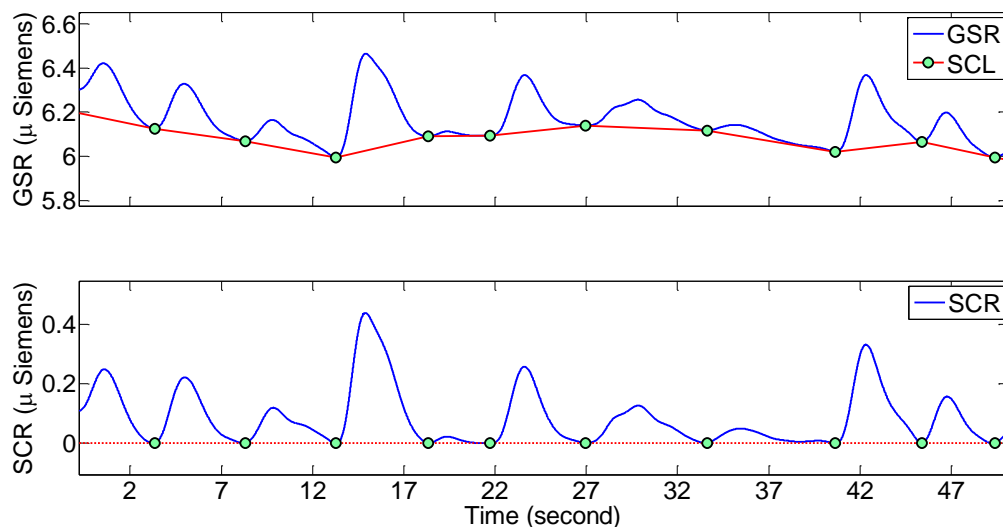
Figure 22. Decomposition of a GSR signal to its SCL and SCR components. (Top) GSR and SCL, (bottom) SCR

Upon decomposition of these two components, the toolbox computes a bunch of statistical features such as mean, median, standard deviation, minimum, maximum, and the number of peaks from the SCR signal as well as from its first and second differences. This toolbox extracts a total of 21 features from the GSR signal. For more information about this toolbox refer to [305].

7.3.2 PPG

AuBT does not specifically support PPG signal but rather has library support for Electrocardiography (ECG) signal. Both PPG and ECG signals are pseudo-periodic waveforms that carry heart-rate variability (HRV) information. It is therefore, straightforward to modify the ECG algorithms in AuBT in order to extract features from PPG.

To calculate HRV, AuBT first derives heart rate time series from the ECG signal by detecting the length of time between every two consecutive R-peaks in the signal. The modified algorithm, instead, measures the time difference between every two consecutive systolic points in the PPG signal for the calculation of instantaneous heart rates.

Upon computing the HRV signal, identical to AuBT, several features are extracted including the mean, median, standard deviation, minimum, maximum, and range of the derived HRV signal. A number of other features such as pRR50 were also extracted from the HRV signal following AuBT algorithms. pRR50 is the fraction of inter-beat intervals (IBI) that differ by more than 50ms from the previous IBI [307]. For more information regarding this toolbox refer to [305]. Similar to the GSR signal, the mean, median, standard deviation, minimum, and maximum of the

raw PPG signal as well as its first and second time differences are extracted. A total of 67 features was extracted from the PPG signal.

7.3.3 SKT

No special signal processing was necessary for the SKT signal. The author simply calculated the aforementioned statistical features from this signal as well as from its first and second time differences. A total of 21 features was extracted from the SKT signal.

7.4 Classification

Although the original goal of this study was to classify fine-grained engagement levels of the participants while naturalistically interacting with the developed tutoring system, a major complication with the dataset was encountered. The distribution of engagement ratings was very unbalanced, thereby rendering the classification problem extremely challenging. Since the tutoring system (including the PA and RP) was designed to be as engaging as possible for the participants, this was no surprise that most of the engagement ratings came out very high. The distribution of engagement ratings for students I, J, and V are depicted in Figure 23, Figure 24, and Figure 25 respectively. When a dataset is highly unbalanced, the classifier cannot appropriately model all the classes, and it ends up ignoring the class with least number of occurrences. As can be seen in these figures, the majority of the ratings, for all three participants, are close to perfect engagement (i.e., 100%). Although this is a positive sign indicative of the tutoring system being highly engaging for the participants, it renders the classification problem very complicated. This complication points out

a limitation of the present study but also highlights the great challenge of training affective models in naturalistic settings.
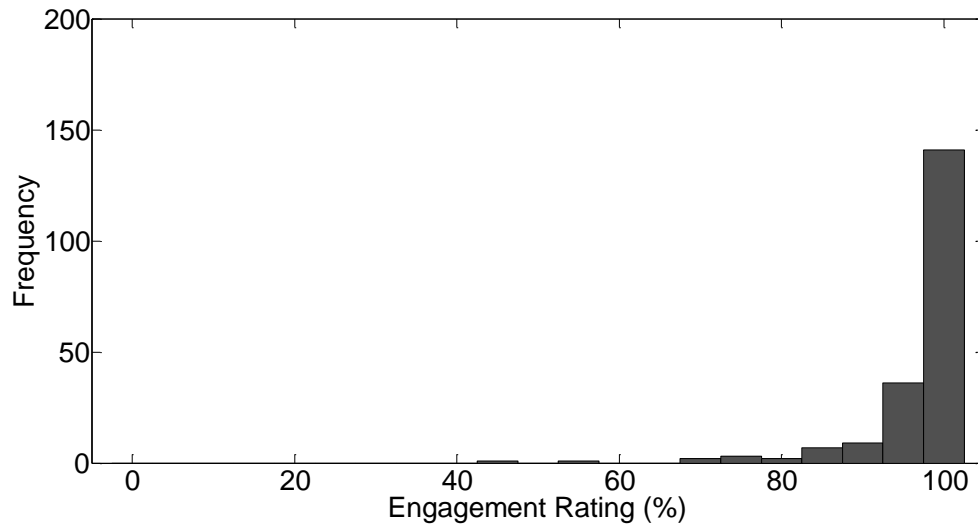
Figure 23. Distribution of engagement ratings for student I
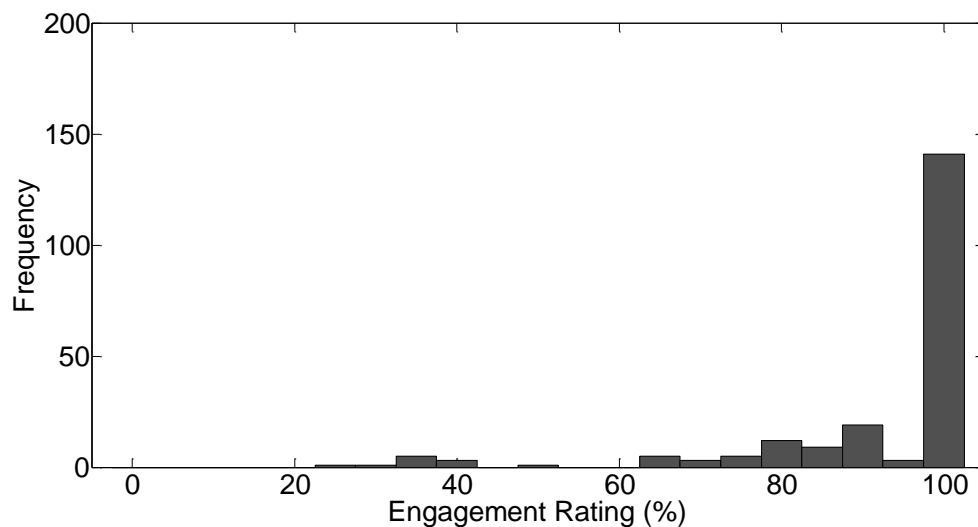
Figure 24. Distribution of engagement ratings for student J

In order to attenuate this issue and start exploring the feasibility of automatically classifying the participants' engagement level based on their physiological signals, the observations with less than 80% engagement scores were removed from the dataset. The remaining observations were then all labeled as "high engagement." Also, every 20-second window of the physiological data

collected during the baselines was labeled as "low engagement." This assumption

was made based on the fact that the participants were instructed to sit quietly and

do nothing during the baseline. It was assumed that this stage induced a state of

disengagement/boredom in the participants.
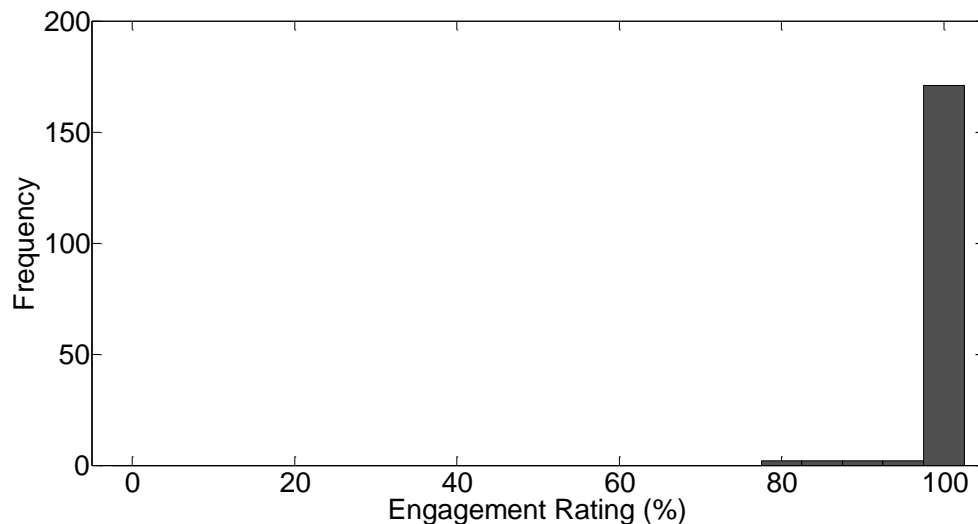
Figure 25. Distribution of engagement ratings for student V

This technique rendered the classification problem as a binary classification

with a near-balanced distribution of classes. Figure 26 shows the distribution of

engagement levels for each participant. The augmented datasets have 386, 386,

and 309 observations for students I, J, and V, respectively. This variation is a result

of the differences in the number of instructional sessions for each student (i.e., the

length of interaction with the tutoring system before they reached the criterion) as

well as the number of observations removed from each participant's dataset.
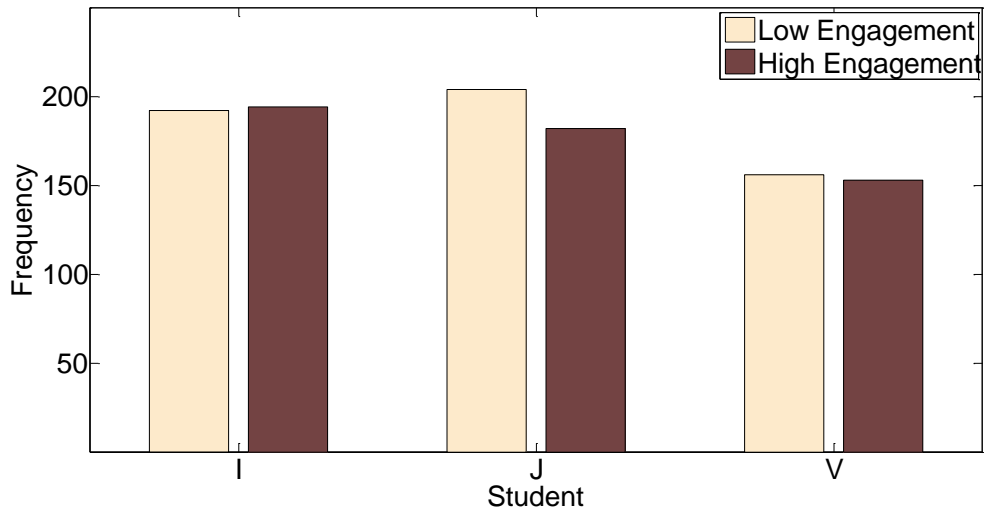
Figure 26. Distribution of engagement levels used in classification

For modeling the participants' engagement levels, a user-dependent approach was taken to account for the person-stereotypy phenomenon and the diverse affective characteristics of the individuals with ASD. The person stereotypy is due to the fact that different individuals may express the same emotion with different characteristic response patterns under the same contexts [308]. Individual differences in responding to emotional stimuli are attributed to factors such as age, sex, culture, and personality traits [309, 310]. Asendorpf and Scherer [311] showed that individuals with two contrasting personality traits differed in their subjective responses to an arousing situation. The user-dependent approach accommodates the differences encountered in affective expression.

For the classification problem, two different supervised pattern recognition algorithms were employed, namely k-Nearest Neighbors (k-NN) [312] and Naïve Bayes (NB). In addition, for the testing and training of classifiers a 10-fold cross-validation protocol was used. Also, classification accuracy (the percentage of

correct recognitions) was utilized as the performance metric since the datasets are fairly balanced.

In order to avoid overfitting and also reducing the computational intensity of the models, the author incorporated the sequential forward feature selection (SFFS) method [313]. SFFS is a wrapper which uses the classifier itself for the evaluation of a specific feature. SFFS performs a non-exhaustive search on the feature space by iteratively adding features. It usually starts with an empty set of features and then picks the feature which improves the classification performance the most in every iteration. SFFS stops iterating when either a desired number of features is reached or there remains no more features that improve performance. SFFS reduces the computational cost and improves generalization of the trained models to unseen events (by avoiding overfitting), but also improves the classification performance by removing noisy features [313].

To identify which physiological signal is more diagnostic of engagement levels, the author analyzed the performance of the models trained based on individual signals. The combination of all three signals was also studied to investigate whether the fusion of signals would result in better recognition performance.

7.5 Results

Table IV shows the classification performance achieved using each of the physiological signals as well as all three combined for the NB classifier. Similarly, Table V, Table VI, and Table VII tabulate the classification accuracies obtained using each of the physiological signals as well as all three of them combined, for 3-NN, 5-NN, 7-NN, and 9-NN classifiers for students I, J, and V, respectively.

Table IV. NB's classification accuracies for individual students

Physiological Signal	Student I	Student J	Student V
PPG	77.84%	68.13%	81.82%
SKT	71.65%	63.47%	68.88%
GSR	70.10%	62.18%	67.48%
PPG+SKT+GSR	79.12%	72.02%	84.97%

Table V. k-NN's classification accuracies for student I

Physiological Signal	3-NN	5-NN	7-NN	9-NN
PPG	77.58%	77.06%	76.55%	76.03%
SKT	75.00%	75.26%	75.52%	73.97%
GSR	78.35%	78.87%	78.35%	78.35%
PPG+SKT+GSR	86.08%	87.11%	84.54%	83.25%

Table VI. k-NN's classification accuracies for student J

Physiological Signal	3-NN	5-NN	7-NN	9-NN
PPG	67.62%	66.32%	67.36%	68.39%
SKT	66.58%	66.32%	67.88%	67.62%
GSR	77.72%	77.98%	76.17%	76.42%
PPG+SKT+GSR	85.23%	82.64%	81.35%	80.31%

Table VII. k-NN's classification accuracies for student V

Physiological Signal	3-NN	5-NN	7-NN	9-NN
PPG	82.87%	83.92%	82.52%	82.87%
SKT	70.98%	72.03%	72.03%	73.08%
GSR	90.21%	88.81%	86.71%	83.57%
PPG+SKT+GSR	93.70%	92.30%	91.61%	89.86%

A repeated analysis of variance (ANOVA) was performed on the classification accuracies in order to assess the effects of classifiers, physiological channels, and students. The ANOVA problem was a 2x4x3 (classifier, physiological channel, and student) randomized complete block design with four replications of each combination. The two levels of the classifier factor were k-NN and NB. The four levels for physiological channels were PPG, SKT, GSR, and

PSG (PPG+SKT+GSR). The three levels for student factor were I, J, and V. The author specifically searched for significant differences in different levels of the three factors with Tukey tests.

Significant differences were found for all three factors (p-value<0.001). k-NN had significantly higher accuracies than NB with the classification accuracy means of 78.64% and 71.67%, respectively. According to the ANOVA results, physiological signals were ranked in the order of PSG (with the mean of 82.14%), PPG (with the mean of 75.31%), GSR (with the mean of 73.60%), and SKT (with the mean of 69.59%). Also, the classifiers were more successful at classifying the engagement levels of the students in the order of V (with the mean of 79.23%), I (with the mean of 76.66%), and J (with the mean of 69.58%). The ANOVA results, furthermore, demonstrated that the combination of k-NN classifier and PSG physiological channels (i.e., using all the physiological signals) results in significantly the highest classification accuracy with the mean accuracy of 86.50% (p-value<0.001) across participants.

7.6 Discussion

The results indicate that the models were successful in discriminating between high and low levels of participants' engagement using their physiological signals collected by EmotiGO while they interacted with the tutoring system. The author believes this is a significant result at it signifies the feasibility of developing an automatic physiology-based classifier of naturalistic expressions using a comfortable, inconspicuous biofeedback device. The participants never appeared to be distracted by EmotiGO or to show any discomfort. In many occasions, they forgot to take off their glasses when the session was over, and the researcher had

to remind them to remove their EmotiGOs. This observation is particularly promising considering the fact that many individuals with ASD are extra sensitive to touch, and, yet, all the participants wore EmotiGO for the length of the instructional sessions without any complaints.

Steps were taken to ensure that the introduction to the novel instructional setting (the experiment room, PA, and RP) and the measuring equipment (EmotiGO) did not alter the participants' affective experience during the instructional sessions. Prior to the physiology and engagement data collection stage, the participants had several chances to visit the instructional setting and repeatedly try EmotiGO on during the word screening and word baseline sessions. Although EmotiGO was designed with robustness against motion artifact in mind, for the duration of physiological data collection, the participants were seated in a comfortable chair without any major body movements. The classification results, hence, are believed not to be influenced by any motion-induced disturbances in the collected signals.

In this study, in order to identify the most diagnostic physiological signal, a systematic analysis was conducted on multichannel physiological data using two classifiers and three physiological signals and their combinations. According to the classifiers' performances, all three physiological signals (PPG, SKT, and GSR) carried information regarding the participants' engagement levels. The results showed that the high and low levels of engagement can be recognized with accuracies significantly higher than random. Among the individual signals, PPG features were significantly more discriminating regarding engagement levels than

were GSR features, and GSR features significantly more than that of SKT signal. The fusion of the features extracted from all three signals, however, outperformed individual signals with substantially higher recognition performance. More specifically, the k-NN classifier along with the concatenated feature sets seem to perform significantly better than all the other combinations of classifiers and physiological signals. Additionally, the fusion approach provided the most robust results since the best individual signal differed among different participants while the fusion model outperformed every individual channel in any of the two classifiers and any participant.

The potential impacts by endowing the tutoring system with the capability of real-time tracking of an autistic learner's engagement level could be various. For instance, when the engagement level is drifted low, the tutoring system could change its pedagogical strategies (using its instructional agents – the PA and RP) or by introduction of interesting activities to retain the learner's attention. Additionally, system designers could optimize/redesign the whole pedagogical approach by analyzing the record of activities and the consequent student's engagement. The tutoring system, itself, could deliver a highly customized tutoring system by learning these patterns over time for each student, and automatically choose strategies that would result in the highest likelihood of student's engagement and, thereby improve learning gains.

In future studies, different fusion techniques such as decision-level fusion (as opposed to feature-level fusion) can be studied. In a decision-level fusion approach, one could train separate classifiers for each individual physiological

signal, and subsequently fuse the models together according to some rule (e.g., voting or Bayesian fusion) to produce an ensemble of classifiers for the whole set of physiological signals. In this study, the author adopted 20-second windows for the physiological signals. Since different signals may have different temporal dynamics, an alternative approach for sensory fusion could consider windows with different lengths, or even windows with varying lengths for each physiological channel. All these concepts warrant further investigation.

CHAPTER 8

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Autism spectrum disorders are neurodevelopmental disorders characterized by impaired social interaction and communication, and by restricted and repetitive behavior. The number of people diagnosed with ASD has increased dramatically since the 1980s. However, there is a lack of appropriate intensive intervention resources in most communities at the present time. Autistic children exhibit chronic and robust deficits like delays in language development and impaired ability for imitation, imaginative play, and non-verbal communication. Early intervention, however, has been proven to bring about a positive long-term outcome. Therefore, a challenge is identification and development of advanced rehabilitation methods for ASD. In this dissertation, the suthor developed two technology-based instructional packages to teach sight word reading to individuals with ASD, a wearable biofeedback device to log their autonomic activity, and a number of computational algorithms to model their engagement levels while interacting with the developed tutoring system. In the reminder of this chapter, summaries of these endeavors as well as insight into future directions are provided.

8.1 Dyadic interaction: Autonomous Pedagogical Agent

In Chapters 3 and 4, the author examined the effects of an instructional package comprised of an autonomous pedagogical agent, automatic speech

recognition, text-to-speech capability, and constant time delay during the instruction of reading sight words aloud to young adults with ASD. The tutoring software was carefully constructed only based on proven techniques for individuals with ASD in order to optimally gain their attention (via the PA's gaze, head orientation, and pointing gesture), motivate them through timely reinforcement, and promote their learning via contingent corrective feedback.

The findings indicate that the tutoring software was successful in teaching SWR to the participants. The participants also maintained performance for eight weeks, and generalized performance outside of the laboratory setting to written stimuli within a classroom setting. These findings are compelling in that the participants met criterion with a limited amount of time spent in receipt of instruction.

The novel application of autonomous instructional technology may offer many potential benefits to teachers of and students with ASD. First, participants made gains in the absence of a human instructional agent. This is critical in that many students with ASD and other disabilities may require higher teacher to student ratios than available in some educational contexts [314]. The use of a PA during periods of instruction may assist teachers in increasing these ratios. In addition, this technology may be used for instruction outside of classroom settings, whereas a trained teacher may not be available. Parents, community support personnel, and other providers might use the technology to teach new skills germane to their specific environments. Second, it is important to note that though the use of computer-based instruction for students with ASD is not new, previous

technologies required students to use a keyboard, mouse, or switch to respond to software directives. In the current investigation, students responded to the PA queries by speaking their responses. This use of ASR technology may increase access to computer-based instruction by students with motor impairments or deficits in keyboarding/mouse skills.

The autonomous tutoring system was designed to resemble the participants' naturalistic instructional settings. Commonly used discrete trial training procedures were employed that typically occurred within the participants' educational program. Since the students' teacher was female, the author programmed a female PA to deliver commonly-used instructions and praise statements. Furthermore, the PA exhibited typical body and hand gestures, head nods, and changing facial expressions. The use of instructional stimuli common to the natural environments has long been recommended as a way to promote generalization [283], and may have contributed to students' generalized responding in the current study. Future research on automatic tutoring software might incorporate other generalization strategies, such as the use of multiple PAs, or thinning schedules of reinforcement to those present in natural environments.

The study conducted in Chapters 3 and 4 illustrates the promise of computer-assisted instruction, and PAs in particular, to supplement conventional techniques in teaching SWR to students with ASD. Several existing technologies were combined to create an effective package that reflects a new direction in automated instructional software. The autonomous software afforded extensive one-to-one practice with the capacity to collect student data with minimal teacher

supervision. This innovation may help teachers increase their instructional efficiency and support students' active engagement during literacy instruction or students may use the software to practice at home, at their own pace, and in the absence of potential negative feedback from peers.

The preliminary findings, although encouraging, should be viewed in the context of several limitations. First, a multiple-baseline design across a small number of participants was employed. This design allowed for a single demonstration of effect for each participant (i.e., inter-subject replication) and in light of small number of stimuli taught, the author acknowledges the limited generalizability of the findings. Furthermore, the participants in the current study had extensive experience in the SWR instruction using CTD. It may have been the case that students less experienced in discrete trial SWR instruction would not have performed as well. Finally, participants spent a limited time interacting with the PA. It is not known whether extended period of PA delivered instruction might enhance or negatively impact student motivation and ultimately responding.

Several other features of this innovative program warrant further discussion. Since the software was able to detect and react to different student vocal responses, it also has the capacity to collect data and monitor student progress. In the current study, the tutoring software was able to discriminate between correct and incorrect responses with high levels of reliability. This finding suggests that teachers may be able to rely on the tutoring software to collect data during similar instructional contexts. Furthermore, the software may be able to make decisions about whether a student has mastered a skill and subsequently introduce new

stimuli. This automated process might reduce teacher errors around data-based decision making. Future research should assess the efficacy and feasibility of such automated data collection systems.

8.2 EmotiGO: Wireless, Unobtrusive Logging of Autonomic Activity

Chapter 5 introduced EmotiGO which is a wireless biofeedback device designed for pervasive affect-sensing applications. EmotiGO is innovative in that it includes multiple biosensors in a single device all embedded through non-standard placement of measurement elements, and miniaturized to a level where the whole system is enclosed in an inconspicuous, lightweight, comfortable form factor. EmotiGO integrates three sensors into a single centralized device eliminating the need to use distributed multiple sensors (e.g., heart rate chest strap and finger-placed GSR sensor), thus resulting in significant decrease of the overall system's size. Additionally, the novel collection site and the unencumbering enclosure facilitate monitoring of physiological signals without interfering with the users' usual activities.

EmotiGO can be used for long-term in situ monitoring of bio-signals thanks to its low power consumption and wireless nature which permits users' mobility. EmotiGO's real-time data transmission is also an important advantage as it allows online collection and analysis of measured signals. Since all the electronics, electrodes, and battery are invisibly integrated in a garment usually worn by people (i.e. eyeglasses), EmotiGO is user friendly and is likely to be quickly accepted by users.

EmotiGO collects the signals from one's face. Since head movements are much less (both in amplitude and frequency) compared to arm and leg movements,

141

EmotiGO is intrinsically less exposed to motion artifacts. Novel design of the enclosure guarantees robust contact of the sensing elements to the user's skin. It also prevents uncontrolled sensing conditions (e.g., loose contact and electrode slippage, variable electrode pressure and position) through fixed positioning of the sensing elements on the collection site. In general, GSR signal is measured via application of conductive gel on the skin and attachment of metal electrodes. EmotiGO, however, does not require application of conductive gel, which makes it feasible for everyday usage, and eliminates the uncertainty caused by different amounts of conductive gel applied on different days.

EmotiGO uses BLE for real-time data transmission of the collected data. The BLE technology consumes only a small fraction of the power of the classic Bluetooth. This fact, along with the aggressive low-power design techniques applied, allow EmotiGO to operate for many hours on a small coin cell battery. BLE's advanced adaptive frequency hopping technology allows multiple EmotiGOs to reliably operate close to each other even in noisy environments. The use of popular Bluetooth technology allows for portability and rapid market adoption due to easy communication with the preexisting devices equipped with a Bluetooth transceiver.

Currently-available devices employ proprietary software and protocols which make them impractical for widespread use. Additionally, the collected data can only be accessed and displayed using the manufacturer's software and are not accessible instantly to other applications and third-party products. The hardware itself, also, can be commanded through the provided proprietary and

closed software. EmotiGO's software is designed based on an API set which makes it very easy for developers to design external applications to interact with the device (i.e., externally control the device and immediately access the data) without worrying about the underlying architecture. EmotiGO's API is a portable binary-based communication protocol designed specifically to simplify application development on various host environments using different programming languages. The open architecture and data format allows software developers to rapidly adopt EmotiGO and easily incorporate it into their systems as an input device. EmotiGO was developed with friendliness in mind - from the end users' point of view as well as the developers'.

In Chapter 5, the author benchmarked EmotiGO against an FDA-approved device for measuring autonomic activity, and showed that the physiological signals can be reliably and simultaneously measured from the temple. At this point, an EmotiGO dongle can be connected to one EmotiGO at a time. In other words, to work with two EmotiGOs, two dongles are required, although several EmotiGOs can operate simultaneously in close vicinity of each other. This, however, does not mean that a specific dongle is dedicated to a specific EmotiGO. An un-connected dongle can connect to any un-connected EmotiGOs. A path to extending EmotiGO's capabilities is to enable EmotiGO dongles to handle simultaneous communications with several EmotiGOs.

8.3 Triadic Interaction: Pedagogical Agent and Robotic Peer

Chapter 6 investigated the feasibility and efficacy of using PA-delivered SWR instruction, based on a CTD procedure, taught within a small-group arrangement consisting of one child with ASD and a robotic peer. The results

indicate that the participants acquired their target words in a limited amount of time spent in receipt of instruction. They also maintained the acquired target words for two months, and generalized performance outside of the laboratory to written stimuli within their home setting.

Furthermore, although a functional relationship was not established (due the nature of the experiment), all participants vicariously acquired a high percentage (94.44% on average) of the non-target words that were exclusively instructed to the robotic peer. In addition, the results indicate that the participants successfully maintained and generalized a substantial amount of the non-target words (88.88% on average). The comparison of the efficiency data (i.e., number of sessions to criterion) for the words acquired through EL and EVL reveals that only 11.11% of target words were acquired faster through EL than through the EVL mode. These data show that not only did the participants learn extra words through vicarious learning but also sight words acquisition may become more efficient by addition of a vicarious component to direct instruction. These promising findings may be explained by the addition of the RP which may have increased the saliency of the required responses and reinforcement by drawing the participants' attention, interest, and motivation. This possibility is of paramount importance since children with ASD show deficits in attending to the relevant elements of instruction.

Although group instruction can be an effective and efficient instructional arrangement for individuals with ASD, it seems necessary to pay careful attention to how to transit from 1:1 to group settings. Special procedures should be identified and be employed to ensure a smooth shift and avoid abrupt transition for many of

these individuals. Added emphasis must be placed on developing strategies that prepare these individuals for group learning and equip them with skills required for multi-student contexts. The system developed in Chapter 6 presents an intelligent tutoring system that resembles a small-group instruction arrangement for children with ASD. Previous literature suggests that generalization is enhanced when the instructional setting is similar to the extra-training setting. Therefore, learning through interaction with this system may properly prepare these individuals for group participation. This system provides ample opportunity to practice group skills which, in turn, increases the likelihood for generalization of those skills to multi-student contexts. The goal of this small-group technology-assisted instruction is not to only instruct sight words to children with ASD but also to engage them in socially appropriate behaviors such as eye contact, joint attention, and turn taking while doing so. The students can practice group/classroom skills while learning sight words in an environment free from peer teasing and without fear of failure.

In Chapter 6, the RP was programmed to show gradual acquisition of target words to resemble the most frequent situation in a natural group setting. Some questions of interest that may be addressed in the future are (a) how would skill acquisition and vicarious learning be affected if the RP simulated a competent model which would result in a higher number of reinforcement than corrective feedback, and (b) how, if possible at all, can a sense of competition be stimulated in the students while attending a small group arrangement including an RP.

Future studies should investigate the feasibility and effects of other instructional procedures (e.g., other response prompting strategies) to instruct

sight word reading as well as other skills/behaviors to children with ASD through such a small-group interaction system. Furthermore, the effectiveness and efficiency of other responding methods such as choral responding can be studied and compared against individual responding. In Chapter 6, the PA delivered instruction to the RP and student in a predictable order (i.e., one after the other). In future investigations, the effects of unpredictable order should also be studied. Unpredictable order of instruction might result in higher attention of each student to the instructional material which would, in turn, bring about higher acquisition rate at least in vicarious mode.

Future investigations should include a larger number of participants, especially female students, and with a wider spectrum of behavioral characteristics and different diagnoses to strengthen the extension of these findings to more versatile populations. With all the current questions regarding how to include children with ASD in regular instructional settings, and considering the importance of vicarious learning, an enhanced understanding of how these individuals acquire skills in such arrangements and through vicarious learning seems to be of substantial importance for future research. Future studies may also attempt to instruct other skills (including functional academic and social behaviors) to children with ASD using a small-group technology-assisted arrangement.

8.4 Physiology-based Affective Modeling

Chapter 7 attempted to model high and low engagement levels of the participants using their physiological signals. The participants wore EmotiGO while they interacted with the tutoring system described in Chapter 6. The physiological signals collected wirelessly by EmotiGO were first preprocessed in order to remove

measurement noise and any motion-induced artifact in the signals. Several features were then extracted from each of the signals. The participants' engagement levels during the interaction with the tutoring system were rated by two trained coders using the video recording of the instructional sessions. The extracted features along with the engagement scores were subsequently input to two supervised pattern recognition algorithms, i.e., k-nearest neighbors and naïve Bayes.

The results indicate that the trained classifiers are capable of discriminating between high and low levels of participants' engagement with recognition accuracies significantly higher than chance. The author believes this is a significant result as it signifies the feasibility of developing an automatic physiology-based classifier of naturalistic expressions using a comfortable, inconspicuous biofeedback device.

In Chapter 7, in order to identify the best physiological signal, a systematic analysis was conducted on multichannel physiological data using two classifiers and three physiological signals (PPG, SKT, and GSR). Among these signals, PPG features were significantly more discriminating regarding engagement levels than were GSR features, and GSR features significantly more than that of the SKT signal. The fusion of the features extracted from all three signals, however, outperformed individual signals with substantially higher recognition performance. More specifically, the k-NN classifier along with the concatenated features seem to perform significantly better than all the other combinations of classifiers and physiological signals. Additionally, the fusion approach provided the most robust

results since the best individual signal differed among different participants while the fusion model outperformed every individual channel in any of the two classifiers and any participant.

The potential impacts by endowing the tutoring system with the capability of real-time tracking of an autistic learner's engagement level could be various. For instance, when the engagement level has drifted low, the tutoring system could change its pedagogical strategies (using its instructional agents - PA and RP) or by introduction of interesting activities to retain the learner's attention. Additionally, system designers could optimize/redesign the whole pedagogical approach by analyzing the record of activities and the student's consequent engagement. The tutoring system, itself, could deliver a highly customized tutoring system by learning these patterns over time for each student, and automatically choose strategies that would result in the highest likelihood of the student's engagement and, thereby, improve learning gains.

In future studies, different fusion techniques such as decision-level fusion (as opposed to feature-level fusion) can be studied. In a decision-level fusion approach, one could train separate classifiers for each individual physiological signal, and subsequently fuse the models together according to some rule (e.g., voting or Bayesian fusion) to produce an ensemble of classifiers for the whole set of physiological signals. Chapter 7 adopted 20-second windows for the physiological signals. Since different signals may have different temporal dynamics, an alternative approach for sensory fusion could consider windows with

different lengths, or even windows with varying lengths for each physiological

channel. All these concepts warrant further investigation.

REFERENCES

1.   American Psychiatric Association, *Diagnostic and statistical manual of mental disorders (DSM-5®).* 2013: American Psychiatric Pub.

2.   Christensen, D. L., *Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012.* MMWR. Surveillance Summaries, 2016. 65.

3.   Rogers, S. J., et al., *Imitation performance in toddlers with autism and those with other developmental disorders.* Journal of child psychology and psychiatry, 2003. 44(5): p. 763-781.

4.   Dautenhahn, K. and I. Werry, *Towards interactive robots in autism therapy: Background, motivation and challenges.* Pragmatics & Cognition, 2004. 12(1): p. 1-35.

5.   Moore, D., *Computers and people with autism.* Asperger Syndrome, 1998: p. 20-21.

6.   Robins, B. *Mediators versus Isolators-the effect of robots on children with Autism and other user-groups.* in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication.* 2007.

7.   Robins, B., et al., *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, 2005. 4(2): p. 105-120.

8.   Werry, I., K. Dautenhahn, and W. Harwin, *Investigating a robot as a therapy partner for children with autism.* 6th European Conference for the Advancement of Assistive Technology, 2001.

9.   Volkmar, F. R., et al., *Handbook of Autism and Pervasive Developmental Disorders, Diagnosis, Development, Neurobiology, and Behavior*. Vol. 1. 2005: John Wiley & Sons.

10.   Gillberg, C., *Asperger syndrome and high-functioning autism.* The British Journal of Psychiatry, 1998. 172(3): p. 200-209.

11.   Lord, C., M. Rutter, and A. Le Couteur, *Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders.* Journal of autism and developmental disorders, 1994. 24(5): p. 659-685.

12.   Prizant, B. M., *Language Acquisition and Communicative Behavior in AutismToward an Understanding of the Whole of It.* Journal of speech and hearing disorders, 1983. 48(3): p. 296-307.

13.   Tager-Flusberg, H., *The challenge of studying language development in children with autism.* Methods for studying language production, 2000: p. 313-332.

14.   Bosseler, A. and D.W. Massaro, *Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism.* Journal of autism and developmental disorders, 2003. 33(6): p. 653-672.

15.   Chen, S. S. A. and V. Bernard-Opitz, *Comparison of personal and computer-assisted instruction for children with autism.* Mental retardation, 1993. 31(6): p. 368.

16.   Coleman-Martin, M. B., et al., *Using computer-assisted instruction and the nonverbal reading approach to teach word identification.* Focus on Autism and other developmental disabilities, 2005. 20(2): p. 80-90.

17.   Heimann, M., et al., *Increasing reading and communication skills in children with autism through an interactive multimedia computer program.* Journal of autism and developmental disorders, 1995. 25(5): p. 459-480.

18.   Hetzroni, O. E. and U. Shalem, *From logos to orthographic symbols: A multilevel fading computer program for teaching nonverbal children with*

autism. Focus on Autism and Other Developmental Disabilities, 2005. 20(4): p. 201-212.

19. Moore, M. and S. Calvert, *Brief report: Vocabulary acquisition for children with autism: Teacher or computer instruction.* Journal of autism and developmental disorders, 2000. 30(4): p. 359-362.

20. Whitcomb, S. A., J. D. Bass, and J. K. Luiselli, *Effects of a computer-based early reading program (Headsprout®) on word list and text reading skills in a student with autism.* Journal of Developmental and Physical Disabilities, 2011. 23(6): p. 491-499.

21. Baylor, A., *Intelligent agents as cognitive tools for education.* Educational technology, 1999. 39(2): p. 36-40.

22. Johnson, W. L., J. W. Rickel, and J.C. Lester, *Animated pedagogical agents: Face-to-face interaction in interactive learning environments.* International Journal of Artificial intelligence in education, 2000. 11(1): p. 47-78.

23. Kearsley, G., *Intelligent agents and instructional systems: Implications of a new paradigm.* Journal of Interactive Learning Research, 1993. 4(4): p. 295.

24. Atkinson, R. K., *Optimizing learning from examples using animated pedagogical agents.* Journal of Educational Psychology, 2002. 94(2): p. 416.

25. Moreno, R., et al., *The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents?* Cognition and instruction, 2001. 19(2): p. 177-213.

26. Baylor, A. L., *Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents.* Educational Technology Research and Development, 2002. 50(2): p. 5-22.

27. Massaro, D. W. and Bosseler, A., *Read my lips The importance of the face in a computer-animated tutor for vocabulary learning by children with autism.* Autism, 2006. 10(5): p. 495-510.

28. Tartaro, A. and J. Cassell. *Playing with virtual peers: bootstrapping contingent discourse in children with autism.* in Proceedings of the 8th

international conference on International conference for the learning sciences-Volume 2. 2008. International Society of the Learning Sciences.

29. Stahmer, A. C., Collings, N. M., and Palinkas, L. A., *Early intervention practices for children with autism: Descriptions from community providers.* Focus on Autism and Other Developmental Disabilities, 2005. 20(2): p. 66-79.

30. Collins, B. C., et al., *Small group instruction: Guidelines for teachers of students with moderate to severe handicaps.* Education and Training in Mental Retardation, 1991: p. 18-32.

31. Koegel, R.L., D.C. Russo, and A. Rincover, *Assessing and training teachers in the generalized use of behavior modification with autistic children.* Journal of Applied Behavior Analysis, 1977. 10(2): p. 197-205.

32. Duker, P., R. Didden, and J. Sigafoos, *One-to-One Training: Instructional Procedures for Learners with Developmental Disabilities*. 2004: ERIC.

33. Hayward, D.W., C.M. Gale, and S. Eikeseth, *Intensive behavioural intervention for young children with autism: A research-based service model.* Research in Autism Spectrum Disorders, 2009. 3(3): p. 571-580.

34. Lovaas, O.I. and T. Smith, *Early and intensive behavioral intervention in autism.* 2003.

35. Downs, A., et al., *Using discrete trial teaching within a public preschool program to facilitate skill development in students with developmental disabilities.* Education and Treatment of Children, 2007: p. 1-27.

36. Smith, T., *Discrete trial training in the treatment of autism.* Focus on autism and other developmental disabilities, 2001. 16(2): p. 86-92.

37. Strain, P.S. and I. Schwartz, *ABA and the development of meaningful social relations for young children with autism.* Focus on Autism and Other Developmental Disabilities, 2001. 16(2): p. 120-128.

38. Bitterman, A., et al., *A national sample of preschoolers with autism spectrum disorders: Special education services and parent satisfaction.* Journal of autism and developmental disorders, 2008. 38(8): p. 1509-1517.

39.   Kamps, D., et al., *A comparison of one-to-one instruction by peers, one-to-one instruction by adults, and small group instruction with children with autism.* Education and Treatment of Children, 1990. 12: p. 197-215.

40.   Brown, L., J. Nietupski, and S. Hamre-Nietupski, *Criterion of ultimate functioning.* Hey, don't forget about me, 1976: p. 2-15.

41.   Favell, J.E., J.E. Favell, and J.F. McGimsey, *Relative effectiveness and efficiency of group vs. individual training of severely retarded persons.* American Journal of Mental Deficiency, 1978.

42.   Koegel, R., A. Egel, and G. Dunlap, *Learning characteristics of autistic children.* Methods of instruction for severely handicapped students, 1980: p. 259-301.

43.   Rogers, S.J., *Interventions that facilitate socialization in children with autism.* Journal of autism and developmental disorders, 2000. 30(5): p. 399-409.

44.   Strain, P.S., F.W. Kohler, and H. Goldstein, *Learning experiences... an alternative program: Peer-mediated interventions for young children with autism.* 1996.

45.   Brown, F., et al., *The individualized curriculum sequencing model (III): Small group instruction.* Research and Practice for Persons with Severe Disabilities, 1980. 5(4): p. 352-367.

46.   Colvin, G., E.J. Kameenui, and G. Sugai, *Reconceptualizing behavior management and school-wide discipline in general education.* Education and treatment of children, 1993: p. 361-381.

47.   Lewis, T.J., *A comparative analysis of the effects of social skill training and teacher-directed contingencies on social behavior of preschool children with disabilities.* Journal of Behavioral Education, 1994. 4(3): p. 267-281.

48.   Polloway, E.A., M.E. Cronin, and J.R. Patton, *The efficacy of group versus one-to-one instruction: A review.* Remedial and Special Education, 1986. 7(1): p. 22-30.

49. Snell, M., *Implementing and monitoring the IEP: Intervention strategies.* Systematic instruction of the moderately and severely handicapped, 1983: p. 113-145.

50. Fink, W.T. and S.R. Sandall, *One-to-one vs group academic instruction with handicapped and nonhandicapped preschool children.* Mental Retardation, 1978.

51. Conway, R.N. and L. Gow, *Mainstreaming special class students with mild handicaps through group instruction.* Remedial and Special Education, 1988. 9(5): p. 34-40.

52. Westling, D.L., K. Ferrell, and K. Swenson, *Intraclassroom comparison of two arrangements for teaching profoundly mentally retarded children.* American journal of mental deficiency, 1982.

53. Alberto, P., et al., *A comparison of individual and group instruction across response tasks.* Research and Practice for Persons with Severe Disabilities, 1980. 5(3): p. 285-293.

54. Koegel, R.L. and A. Rincover, *Treatment of psychotic children in a classroom environment: I. Learning in a large group.* Journal of Applied Behavior Analysis, 1974. 7(1): p. 45-59.

55. Schoen, S.F. and S. Ogden, *Impact of time delay, observational learning, and attentional cuing upon word recognition during integrated small-group instruction.* Journal of Autism and Developmental Disorders, 1995. 25(5): p. 503-519.

56. Taubman, M., et al., *The effectiveness of a group discrete trial instructional approach for preschoolers with developmental disabilities.* Research in Developmental Disabilities, 2001. 22(3): p. 205-219.

57. Frankosky, R.J. and B. Sulzer-Azaroff, *Individual and group contingencies and collateral social behaviors.* Behavior Therapy, 1978. 9(3): p. 313-327.

58. Peck, C.A., *Increasing opportunities for social control by children with autism and severe handicaps: Effects on student behavior and perceived classroom climate.* Research and Practice for Persons with Severe Disabilities, 1985. 10(4): p. 183-193.

59. Johnson, J.L., et al., *Interactive individualized instruction with small groups of severely handicapped students.* Education and Training of the Mentally Retarded, 1980: p. 230-237.

60. Oliver, P.R. and T.L. Scott, *Group versus individual training in establishing generalization of language skills with severely handicapped individuals.* Mental Retardation, 1981. 19(6): p. 285.

61. Browder, D.M., S.F. Schoen, and F.E. Lentz, *Learning to learn through observation.* The Journal of Special Education, 1986. 20(4): p. 447-461.

62. Cicchetti, D. and L.A. Sroufe, *The relationship between affective and cognitive development in Down's syndrome infants.* Child development, 1976: p. 920-929.

63. Graesser, A. and S.K. D'Mello, *Theoretical perspectives on affect and deep learning*, in *New perspectives on affect and learning technologies*. 2011, Springer. p. 11-21.

64. Kort, B., R. Reilly, and R.W. Picard. *An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion*. in *icalt*. 2001.

65. Csikszentmihalyi, M., *Flow: The psychology of optimal experience*. 1990, New York: Harper and Row.

66. Miserandino, M., *Children who do well in school: Individual differences in perceived competence and autonomy in above-average children.* Journal of Educational Psychology, 1996. 88(2): p. 203.

67. Graesser, A.C. and B.A. Olde, *How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down.* Journal of Educational Psychology, 2003. 95(3): p. 524.

68. Patrick, B.C., E.A. Skinner, and J.P. Connell, *What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain.* Journal of Personality and social Psychology, 1993. 65(4): p. 781.

69.   Craig, S., et al., *Affect and learning: an exploratory look into the role of affect in learning with AutoTutor.* Journal of educational media, 2004. 29(3): p. 241-250.

70.   Arroyo, I., et al. *Affective Gendered Learning Companions*. in *AIED*. 2009.

71.   D'Mello, S.K. and A.C. Graesser. *Malleability of Students' Perceptions of an Affect-Sensitive Tutor and Its Influence on Learning*. in *FLAIRS Conference*. 2012.

72.   D'mello, S. and A. Graesser, *AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back.* ACM Transactions on Interactive Intelligent Systems (TiiS), 2012. 2(4): p. 23.

73.   Mao, X. and Z. Li, *Agent based affective tutoring systems: A pilot study.* Computers & Education, 2010. 55(1): p. 202-208.

74.   Mazefsky, C.A., K.A. Pelphrey, and R.E. Dahl, *The need for a broader approach to emotion regulation research in autism.* Child development perspectives, 2012. 6(1): p. 92-97.

75.   Bartlett, M.S., et al. *Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction*. in *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*. 2003. IEEE.

76.   Lee, C.M. and S.S. Narayanan, *Toward detecting emotions in spoken dialogs.* IEEE transactions on speech and audio processing, 2005. 13(2): p. 293-303.

77.   Camurri, A., et al., *Eyesweb: Toward gesture and affect recognition in interactive dance and music systems.* Computer Music Journal, 2000. 24(1): p. 57-69.

78.   Picard, R.W., E. Vyzas, and J. Healey, *Toward machine emotional intelligence: Analysis of affective physiological state.* IEEE transactions on pattern analysis and machine intelligence, 2001. 23(10): p. 1175-1191.

79.   Green, D., et al., *The severity and nature of motor impairment in Asperger's syndrome: a comparison with specific developmental disorder of motor*

function. Journal of child psychology and psychiatry, 2002. 43(5): p. 655-668.

80.   Schultz, R.T., *Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area.* International Journal of Developmental Neuroscience, 2005. 23(2): p. 125-141.

81.   Shalom, D.B., et al., *Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism.* Journal of autism and developmental disorders, 2006. 36(3): p. 395-400.

82.   Golan, O. and S. Baron-Cohen, *Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia.* Development and psychopathology, 2006. 18(02): p. 591-617.

83.   Golan, O., et al., *Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces.* Journal of autism and developmental disorders, 2010. 40(3): p. 269-279.

84.   Lacava, P.G., et al., *A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD.* Autism, 2010.

85.   Bernard-Opitz, V., N. Sriram, and S. Nakhoda-Sapuan, *Enhancing social problem solving in children with autism and normal children through computer-assisted instruction.* Journal of autism and developmental disorders, 2001. 31(4): p. 377-384.

86.   Bernard-Opitz, V., N. Sriram, and S. Sapuan, *Enhancing vocal imitations in children with autism using the IBM speech viewer.* Autism, 1999. 3(2): p. 131-147.

87.   Williams, C., et al., *Do children with autism learn to read more readily by computer assisted instruction or traditional book methods? A pilot study.* Autism, 2002. 6(1): p. 71-91.

88. Weir, S. and R. Emanuel, *Using LOGO to catalyse communication in an autistic child*. 1976: University of Edinburgh. Department of Artificial Intelligence.

89. Billard, A., K. Dautenhahn, and G. Hayes. *Experiments on human-robot communication with Robota, an imitative learning and communicating doll robot*. in *Proceedings of Socially Situated Intelligence Workshop held within the Fifth Conference on Simulation of Adaptive Behavior (SAB'98). Centre for Policy Modelling technical report series: No. CPM–98–38, Zurich, Switzerland*. 1998. Citeseer.

90. Kozima, H., C. Nakagawa, and Y. Yasuda. *Designing and observing human-robot interactions for the study of social development and its disorders*. in *2005 International Symposium on Computational Intelligence in Robotics and Automation*. 2005. IEEE.

91. Robins, B., P. Dickerson, and K. Dautenhahn. *Robots as embodied beings-interactionally sensitive body movements in interactions among autistic children and a robot*. in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. 2005. IEEE.

92. Kozima, H., 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*. 2005. IEEE.

93. Michaud, F., et al., *Perspectives on mobile robots as tools for child development and pediatric rehabilitation*. Assistive Technology, 2007. 19(1): p. 21-36.

94. Kozima, H. and H. Yano. *Designing a robot for contingency-detection game*. in *Workshop on Robotic and Virtual Interactive Systems in Autism Therapy*. 2001.

95. Dautenhahn, K. and I. Werry. *A quantitative technique for analysing robot-human interactions*. in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*. 2002. IEEE.

96.   Werry, I., et al., *Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy*, in *Cognitive technology: instruments of mind*. 2001, Springer. p. 57-74.

97.   Dautenhahn, K., et al., *KASPAR–a minimally expressive humanoid robot for human–robot interaction research.* Applied Bionics and Biomechanics, 2009. 6(3-4): p. 369-397.

98.   Iacono, I., et al. *Robots as social mediators for children with Autism-A preliminary analysis comparing two different robotic platforms*. in *2011 IEEE International Conference on Development and Learning (ICDL)*. 2011. IEEE.

99.   Lehmann, H., et al. *'Make it move': playing cause and effect games with a robot companion for children with cognitive disabilities*. in *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics*. 2011. ACM.

100.   Robins, B., K. Dautenhahn, and P. Dickerson. *From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot*. in *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*. 2009. IEEE.

101.   Wainer, J., et al. *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*. 2010. IEEE.

102.   Feil-Seifer, D. and M.J. Matarić. *Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders*. in *Experimental robotics*. 2009. Springer.

103.   Feil-Seifer, D. and M. Mataric. *Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features*. in *Proceedings of the 6th international conference on Human-robot interaction*. 2011. ACM.

104. Feil-Seifer, D. and M. Matarić. *Using proxemics to evaluate human-robot interaction*. in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. 2010. IEEE Press.

105. Bekele, E., et al. *Development of a novel robot-mediated adaptive response system for joint attention task for children with autism*. in *2011 RO-MAN*. 2011. IEEE.

106. Stanton, C.M., et al. *Robotic animals might aid in the social development of children with autism*. in *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*. 2008. IEEE.

107. Charman, T., *Why is joint attention a pivotal skill in autism?* Philosophical Transactions of the Royal Society B: Biological Sciences, 2003. 358(1430): p. 315-324.

108. Mundy, P., *Joint attention and social-emotional approach behavior in children with autism.* Development and Psychopathology, 1995. 7(01): p. 63-82.

109. McArthur, D. and L.B. Adamson, *Joint attention in preverbal children: Autism and developmental language disorder.* Journal of autism and developmental disorders, 1996. 26(5): p. 481-496.

110. Mundy, P. and M. Crowson, *Joint attention and early social communication: Implications for research on intervention with autism.* Journal of Autism and Developmental disorders, 1997. 27(6): p. 653-676.

111. Fasel, I., et al. *Combining embodied models and empirical research for understanding the development of shared attention*. in *Development and Learning, 2002. Proceedings. The 2nd International Conference on*. 2002. IEEE.

112. Kozima, H., C. Nakagawa, and Y. Yasuda, *Children–robot interaction: a pilot study in autism therapy.* Progress in Brain Research, 2007. 164: p. 385-400.

113. Robins, B., et al., *Robot-mediated joint attention in children with autism: A case study in robot-human interaction.* Interaction studies, 2004. 5(2): p. 161-198.

114. Robins, B. and K. Dautenhahn. *The role of the experimenter in hri research-a case study evaluation of children with autism interacting with a robotic toy*. in *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. 2006. IEEE.

115. Dautenhahn, K. and C.L. Nehaniv, *Imitation in animals and artifacts*. 2002: MIT Press Cambridge, MA.

116. Jordan, R., *Autistic spectrum disorders: an introductory handbook for practitioners*. 2013: Routledge.

117. Nadel, J., et al., *The evolving nature of imitation as a format for communication.* 1999.

118. Dawson, G. and A. Adams, *Imitation and social responsiveness in autistic children.* Journal of abnormal child psychology, 1984. 12(2): p. 209-226.

119. Stribling, P., J. Rae, and P. Dickerson, *Using conversation analysis to explore the recurrence of a topic in the talk of a boy with an autism spectrum disorder.* Clinical linguistics & phonetics, 2009. 23(8): p. 555-582.

120. Bird, G., et al., *Intact automatic imitation of human and robot actions in autism spectrum disorders.* Proceedings of the Royal Society of London B: Biological Sciences, 2007. 274(1628): p. 3027-3031.

121. Pierno, A.C., et al., *Robotic movement elicits visuomotor priming in children with autism.* Neuropsychologia, 2008. 46(2): p. 448-454.

122. Duquette, A., F. Michaud, and H. Mercier, *Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism.* Autonomous Robots, 2008. 24(2): p. 147-157.

123. Meltzoff, A.N. and M.K. Moore, *Persons and representation: Why infant imitation is important for theories of human development.* 1999.

124. Gupta, P. and B. MacWhinney, *Vocabulary acquisition and verbal short-term memory: Computational and neural bases.* Brain and language, 1997. 59(2): p. 267-333.

125. Neuman, S.B., C. Copple, and S. Bredekamp, *Learning to read and write: Developmentally appropriate practices for young children*. 2000: ERIC.

126.   Beirne-Smith, M., J.R. Patton, and R. Ittenbach, *Mental retardation*. 1994: Merrill.

127.   Vacca, J.S., *Autistic Children Can Be Taught to Read.* International Journal of Special Education, 2007. 22(3): p. 54-61.

128.   Randi, J., T. Newman, and E.L. Grigorenko, *Teaching children with autism to read for meaning: Challenges and possibilities.* Journal of autism and developmental disorders, 2010. 40(7): p. 890-902.

129.   Light, J. and A.K. Smith, *Home literacy experiences of preschoolers who use AAC systems and of their nondisabled peers.* Augmentative and Alternative Communication, 1993. 9(1): p. 10-25.

130.   Stahl, S.A., M.C. McKenna, and J.R. Pagnucco, *The effects of whole-language instruction: An update and a reappraisal.* Educational Psychologist, 1994. 29(4): p. 175-185.

131.   Gabig, C.S., *Phonological awareness and word recognition in reading by children with autism.* Communication Disorders Quarterly, 2010. 31(2): p. 67-85.

132.   Browder, D.M. and J.S. Lalli, *Review of research on sight word instruction.* Research in developmental disabilities, 1991. 12(3): p. 203-228.

133.   Burns, M.K., *Comparison of opportunities to respond within a drill model when rehearsing sight words with a child with mental retardation.* School Psychology Quarterly, 2007. 22(2): p. 250.

134.   Burns, M.K. and H.E. Sterling-Turner, *Comparison of efficiency measures for academic interventions based on acquisition and maintenance.* Psychology in the Schools, 2010. 47(2): p. 126-134.

135.   Lerner, J., *Learning disabilities: Theories, diagnosis, and teaching practices.* 2003, Boston: Houghton Mifflin Company.

136.   Bos, C.S. and S. Vaughn, *Strategies for Teaching Students with Learning and Behavior Problems.* 1998: Allyn and Bacon.

137.   Browder, D.M., et al., *A treatment package for increasing sight word recognition for use in daily living skills.* Education and Training of the Mentally Retarded, 1984: p. 191-200.

138. Cuvo, A.J. and K.P. Klatt, *Effects of Community-based, videotape, and flash card instruction of community-referenced sight words on students with mental retardation.* Journal of Applied Behavior Analysis, 1992. 25(2): p. 499-512.

139. Schloss, P.J., et al., *Acquisition of functional sight words in community-based recreation settings.* The Journal of Special Education, 1995. 29(1): p. 84-96.

140. Bliss, S.L., C.H. Skinner, and R. Adams, *Enhancing an English language learning fifth-grade student's sight-word reading with a time-delay taped-words intervention.* School Psychology Review, 2006. 35(4): p. 663.

141. Nist, L. and L.M. Joseph, *Effectiveness and efficiency of flashcard drill instructional methods on urban first-graders' word recognition, acquisition, maintenance, and generalization.* School Psychology Review, 2008. 37(3): p. 294.

142. Yaw, J.S., et al., *Extending research on a computer-based sight-word reading intervention to a student with autism.* Journal of Behavioral Education, 2011. 20(1): p. 44-54.

143. Putnam, C. and L. Chong. *Software and technologies designed for people with autism: what do users want?* in *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility.* 2008. ACM.

144. Baron-Cohen, S., O. Golan, and E. Ashwin, *Can emotion recognition be taught to children with autism spectrum conditions?* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2009. 364(1535): p. 3567-3574.

145. Mechling, L.C., D.L. Gast, and K. Krupa, *Impact of SMART Board technology: An investigation of sight word reading and observational learning.* Journal of Autism and Developmental Disorders, 2007. 37(10): p. 1869-1882.

146. Coleman, M.B., K.J. Hurley, and D.F. Cihak, *Comparing teacher-directed and computer-assisted constant time delay for teaching functional sight*

words to students with moderate intellectual disability. Education and Training in Autism and Developmental Disabilities, 2012: p. 280-292.

147. Silver, M. and P. Oakes, *Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others.* Autism, 2001. 5(3): p. 299-316.

148. Raskind, M.H. and E.L. Higgins, *Speaking to read: The effects of speech recognition technology on the reading and spelling performance of children with learning disabilities.* Annals of Dyslexia, 1999. 49(1): p. 251-281.

149. Higgins, E.L. and M.H. Raskind, *Speaking to read: The effects of continuous vs. discrete speech recognition systems on the reading and spelling of children with learning disabilities.* Journal of Special Education Technology, 1999. 15(1): p. 19-30.

150. Maurice, C.E., G.E. Green, and S.C. Luce, *Behavioral intervention for young children with autism: A manual for parents and professionals.* 1996: Pro-ed.

151. Judge, S.L., *Computer applications in programs for young children with disabilities: Current status and future directions.* Journal of Special Education Technology, 2000. 16(1): p. 29-40.

152. Cassell, J., *Embodied conversational agents: representation and intelligence in user interfaces.* AI magazine, 2001. 22(4): p. 67.

153. Rajendran, G. and P. Mitchell, *Text chat as a tool for referential questioning in Asperger syndrome.* Journal of Speech, Language, and Hearing Research, 2006. 49(1): p. 102-112.

154. Schmidt, C. and M. Schmidt. *Three-dimensional virtual learning environments for mediating social skills acquisition among individuals with autism spectrum disorders.* in *Proceedings of the 7th international conference on Interaction design and children.* 2008. ACM.

155. Tartaro, A. and J. Cassell. *Authorable virtual peers for autism spectrum disorders.* in *Proceedings of the Combined workshop on Language-Enabled Educational Technology and Development and Evaluation for*

Robust Spoken Dialogue Systems at the 17th European Conference on Artificial Intellegence. 2006.

156. Milne, M., et al. *Development of a virtual agent based social tutor for children with autism spectrum disorders.* in *The 2010 International Joint Conference on Neural Networks (IJCNN).* 2010. IEEE.

157. Mower, E., et al. *Rachel: Design of an emotionally targeted interactive agent for children with autism.* in *2011 IEEE International Conference on Multimedia and Expo.* 2011. IEEE.

158. Walker, G., *Constant and progressive time delay procedures for teaching children with autism: A literature review.* Journal of Autism and Developmental Disorders, 2008. 38(2): p. 261-275.

159. Ault, M.J., et al., *Comparison of response prompting procedures in teaching numeral identification to autistic subjects.* Journal of Autism and Developmental Disorders, 1988. 18(4): p. 627-636.

160. Godby, S., D.L. Gast, and M. Wolery, *A comparison of time delay and system of least prompts in teaching object identification.* Research in Developmental Disabilities, 1987. 8(2): p. 283-305.

161. Heckaman, K.A., et al., *A comparison of least-to-most prompts and progressive time delay on the disruptive behavior of students with autism.* Journal of Behavioral Education, 1998. 8(2): p. 171-201.

162. Dipipi, C.M., A.K. Jitendra, and J.A. Miller, *Reducing repetitive speech: Effects of strategy instruction.* Preventing School Failure: Alternative Education for Children and Youth, 2001. 45(4): p. 177-181.

163. Liber, D.B., W.D. Frea, and J.B. Symon, *Using time-delay to improve social play skills with peers for children with autism.* Journal of Autism and Developmental Disorders, 2008. 38(2): p. 312-323.

164. Swain, R., J.D. Lane, and D.L. Gast, *Comparison of Constant Time Delay and Simultaneous Prompting Procedures: Teaching Functional Sight Words to Students with Intellectual Disabilities and Autism Spectrum Disorder.* Journal of Behavioral Education, 2015. 24(2): p. 210-229.

165. Ledford, J.R., et al., *Observational and incidental learning by children with autism during small group instruction.* Journal of Autism and Developmental Disorders, 2008. 38(1): p. 86-103.

166. Wilson, R., D. Majsterek, and D. Simmons, *The effects of computer-assisted versus teacher-directed instruction on the multiplication performance of elementary students with learning disabilities.* Journal of Learning Disabilities, 1996. 29(4): p. 382-390.

167. Neitzel, J. and M. Wolery, *Steps for implementation: Time delay.* Chapel Hill: The National Professional Development Center on Autism Spectrum Disorders, Frank Porter Graham Child Development Institute, The University of North Carolina, 2009.

168. www.worldviz.com.

169. Alcorn, A., et al. *Social communication between virtual characters and children with autism.* in *International Conference on Artificial Intelligence in Education.* 2011. Springer.

170. Argyle, M. and J. Dean, *Eye-contact, distance and affiliation.* Sociometry, 1965: p. 289-304.

171. Bancroft, W.J., *Research in Nonverbal Communication and Its Relationship to Pedagogy and Suggestopedia.* 1995.

172. Hall, E.T., *The anthropology of manners.* Scientific American, 1955. 192: p. 84-91.

173. Welch, K.C., et al., *An approach to the design of socially acceptable robots for children with autism spectrum disorders.* International journal of social robotics, 2010. 2(4): p. 391-403.

174. Gillott, A., F. Furniss, and A. Walter, *Anxiety in high-functioning children with autism.* Autism, 2001. 5(3): p. 277-286.

175. www.ald.softbankrobotics.com.

176. www.nuance.com.

177. www.sitepal.com.

178. Dolch, E.W., *Problems in reading.* 1948: Garrard Press.

179. Brigance, A.H., *BRIGANCE (TM) Diagnostic Inventory of Early Development Including Developmental Record Book.* 1978.

180. Baer, D.M., M.M. Wolf, and T.R. Risley, *Some current dimensions of applied behavior analysis1.* Journal of applied behavior analysis, 1968. 1(1): p. 91-97.

181. Billingsley, F., O.R. White, and R. Munson, *Procedural reliability: A rationale and an example.* Behavioral assessment, 1980. 2(2): p. 2.

182. Koegel, R. and A. Rincover, *RESEARCH ON THE DIFFERENCE BETWEEN GENERALIZATION AND MAINTENANCE IN EXTRA-THERAPY RESPONDING1.* Journal of Applied Behavior Analysis, 1977. 10(1): p. 1-12.

183. Handleman, J.S., *Generalization by autistic-type children of verbal responses across settings.* Journal of Applied Behavior Analysis, 1979. 12(2): p. 273-282.

184. Mirenda, P., *Supporting individuals with challenging behavior through functional communication training and AAC: Research review.* Augmentative and Alternative Communication, 1997. 13(4): p. 207-225.

185. Picard, R.W. and R. Picard, *Affective computing.* Vol. 252. 1997: MIT press Cambridge.

186. Sidney, K.D., et al. *Integrating affect sensors in an intelligent tutoring system.* in *Affective Interactions: The Computer in the Affective Loop Workshop at.* 2005.

187. Liu, C., et al., *Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback.* International Journal of Human-Computer Interaction, 2009. 25(6): p. 506-529.

188. Breazeal, C., *Emotion and sociable humanoid robots.* International Journal of Human-Computer Studies, 2003. 59(1): p. 119-155.

189. Liu, C., et al., *Online affect detection and robot behavior adaptation for intervention of children with autism.* IEEE Transactions on Robotics, 2008. 24(4): p. 883-896.

190. Lisetti, C. and C. LeRouge. *Affective computing in tele-home health*. in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*. 2004. IEEE.

191. Pantic, M. and L.J.M. Rothkrantz, *Automatic analysis of facial expressions: The state of the art*. IEEE Transactions on pattern analysis and machine intelligence, 2000. 22(12): p. 1424-1445.

192. El Ayadi, M., M.S. Kamel, and F. Karray, *Survey on speech emotion recognition: Features, classification schemes, and databases*. Pattern Recognition, 2011. 44(3): p. 572-587.

193. Jerritta, S., et al. *Physiological signals based human emotion recognition: a review*. in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. 2011. IEEE.

194. Kim, J., *Bimodal emotion recognition using speech and physiological changes*. 2007: Citeseer.

195. www.biopac.com.

196. Picard, R.W. and J. Scheirer. *The galvactivator: A glove that senses and communicates skin conductivity*. in *Proceedings 9th Int. Conf. on HCI*. 2001.

197. Strauss, M., et al. *The handwave bluetooth skin conductance sensor*. in *International Conference on Affective Computing and Intelligent Interaction*. 2005. Springer.

198. www.brainquiry.com.

199. Axisa, F., et al. *Wrist ambulatory monitoring system and smart glove for real time emotional, sensorial and physiological analysis*. in *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*. 2004. IEEE.

200. Peter, C., et al. *EREC-II in use–Studies on usability and suitability of a sensor system for affect detection and human performance monitoring*. in *International Conference on Human-Computer Interaction*. 2007. Springer.

201. Fletcher, R.R., et al., *iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity*. IEEE

Transactions on Information Technology in Biomedicine, 2010. 14(2): p. 215-223.

202. Poh, M.-Z., N.C. Swenson, and R.W. Picard, *A wearable sensor for unobtrusive, long-term assessment of electrodermal activity.* IEEE transactions on Biomedical engineering, 2010. 57(5): p. 1243-1252.

203. www.empatica.com.

204. Heydon, R., *Bluetooth low energy.* 2013: Prentice Hall.

205. Roth, W.T., M.E. Dawson, and D.L. Filion, *Publication recommendations for electrodermal measurements.* Psychophysiology, 2012. 49: p. 1017-1034.

206. Khalfa, S., et al., *Event-related skin conductance responses to musical emotions in humans.* Neuroscience letters, 2002. 328(2): p. 145-149.

207. Healey, J. and R. Picard. *SmartCar: detecting driver stress.* in *Pattern Recognition, 2000. Proceedings. 15th International Conference on.* 2000. IEEE.

208. Randall, W.C., *Quantitation and regional distribution of sweat glands in man.* Journal of Clinical Investigation, 1946. 25(5): p. 761.

209. Saga, K., *Structure and function of human sweat glands studied with histochemistry and cytochemistry.* Progress in histochemistry and cytochemistry, 2002. 37(4): p. 323-386.

210. Frewin, D.B. and J.A. Downey, *Sweating--physiology and pathophysiology.* Australasian Journal of Dermatology, 1976. 17(3): p. 82-86.

211. Tronstad, C., et al., *Electrical measurement of sweat activity.* Physiological measurement, 2008. 29(6): p. S407.

212. van Dooren, M. and J.H. Janssen, *Emotional sweating across the body: Comparing 16 different skin conductance measurement locations.* Physiology & behavior, 2012. 106(2): p. 298-304.

213. Takahashi, T., et al., *Changes in EEG and autonomic nervous activity during meditation and their association with personality traits.* International Journal of Psychophysiology, 2005. 55(2): p. 199-207.

214. Mendelson, Y. and C. Pujary. *Measurement site and photodetector size considerations in optimizing power consumption of a wearable reflectance*

pulse oximeter. in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE.* 2003. IEEE.

215. Ahmed, M.U., et al., *A multi-module case-based biofeedback system for stress treatment.* Artificial Intelligence in Medicine, 2011. 51(2): p. 107-115.

216. Yang, C.-Y., et al., *The Effect of Music Therapy on Hospitalized Psychiatric Patients' Anxiety, Finger Temperature, and Electroencephalography A Randomized Clinical Trial.* Biological research for nursing, 2012. 14(2): p. 197-206.

217. Harioka, T., et al., *"Deep-forehead" temperature correlates well with blood temperature.* Canadian Journal of Anesthesia, 2000. 47(10): p. 980-983.

218. Hebbar, K., et al., *Comparison of temporal artery thermometer to standard temperature measurements in pediatric intensive care unit patients.* Pediatric Critical Care Medicine, 2005. 6(5): p. 557-561.

219. *www.bluegiga.com.*

220. www.ti.com.

221. Davis, M., *Neurochemical modulation of sensory-motor reactivity: acoustic and tactile startle reflexes.* Neuroscience & Biobehavioral Reviews, 1980. 4(2): p. 241-263.

222. Jensen, A.R. and W.D. Rohwer, *The Stroop color-word test: a review.* Acta psychologica, 1966. 25: p. 36-93.

223. Vinkers, C.H., et al., *The effect of stress on core and peripheral body temperature in humans.* Stress, 2013. 16(5): p. 520-530.

224. Carr, E.G. and M. Darcy, *Setting generality of peer modeling in children with autism.* Journal of Autism and Developmental Disorders, 1990. 20(1): p. 45-59.

225. Ganz, J.B., et al., *Implementing visually cued imitation training with children with autism spectrum disorders and developmental delays.* Journal of Positive Behavior Interventions, 2008. 10(1): p. 56-66.

226. Garfinkle, A.N. and I.S. Schwartz, *Peer imitation increasing social interactions in children with autism and other developmental disabilities in*

inclusive preschool classrooms. Topics in Early Childhood Special Education, 2002. 22(1): p. 26-38.

227. Baer, D.M., R.F. Peterson, and J.A. Sherman, *THE DEVELOPMENT OF IMITATION BY REINFORCING BEHAVIORAL SIMILARITY TO A MODEL1.* Journal of the Experimental analysis of Behavior, 1967. 10(5): p. 405-416.

228. Cooper, J.O.H., et al., *Applied behavior analysis.* 2007.

229. Rogers, S.J., *An examination of the imitation deficit in autism.* 1999.

230. Williams, J.H., A. Whiten, and T. Singh, *A systematic review of action imitation in autistic spectrum disorder.* Journal of autism and developmental disorders, 2004. 34(3): p. 285-299.

231. Catania, A., *Learning (interim 4th ed.).* Cornwall-on-Hudson, NY: Sloan, 2007.

232. Gautreaux, G., *The effects of monitoring training on the acquisition of an observational learning repertoire under peer tutoring conditions, generalization and collateral effects.* 2005, [Sl: sn].

233. Taylor, B.A. and J.A. DeQuinzio, *Observational learning and children with autism.* Behavior Modification, 2012. 36: p. 341-360.

234. Bandura, A., *Social learning through imitation.* 1962.

235. Bandura, A., *Vicarious processes: A case of no-trial learning.* Advances in experimental social psychology, 1965. 2: p. 1-55.

236. Bandura, A., *Influence of models' reinforcement contingencies on the acquisition of imitative responses.* Journal of personality and social psychology, 1965. 1(6): p. 589.

237. Bandura, A. and R.H. Walters, *Social learning and personality development.* Vol. 14. 1963: JSTOR.

238. Bandura, A. and R.H. Walters, *Social learning theory.* 1977.

239. Douglas Greer, R., J. Dudek-Singer, and G. Gautreaux, *Observational learning.* International Journal of Psychology, 2006. 41(6): p. 486-499.

240. Farmer, J.A., et al., *Small group instruction for students with severe handicaps: A study of observational learning.* Education and Training in Mental Retardation, 1991: p. 190-201.

241. Shelton, B.S., et al., *The role of small group instruction in facilitating observational and incidental learning.* Language, Speech, and Hearing Services in Schools, 1991. 22(3): p. 123-133.

242. Orelove, F.P., *Acquisition of incidental learning in moderately and severely handicapped adults.* Education and Training of the Mentally Retarded, 1982: p. 131-136.

243. Stinson, D.M., et al., *Acquisition of nontargeted information during small-group instruction.* Exceptionality: A Special Education Journal, 1991. 2(2): p. 65-80.

244. Leaf, J.B., et al., *A COMPARISON OF DISCRETE TRIAL TEACHING IMPLEMENTED IN A ONE-TO-ONE INSTRUCTIONAL FORMAT AND IN A GROUP INSTRUCTIONAL FORMAT.* Behavioral Interventions, 2013. 28(1): p. 82-106.

245. Kroeger, K., J.R. Schultz, and C. Newsom, *A comparison of two group-delivered social skills programs for young children with autism.* Journal of autism and developmental disorders, 2007. 37(5): p. 808-817.

246. Wilson, K.P., *Teaching social-communication skills to preschoolers with autism: Efficacy of video versus in vivo modeling in the classroom.* Journal of autism and developmental disorders, 2013. 43(8): p. 1819-1831.

247. Colozzi, G.A., L.W. Ward, and K.E. Crotty, *Comparison of simultaneous prompting procedure in 1: 1 and small group instruction to teach play skills to preschool students with pervasive developmental disorder and developmental disabilities.* Education and Training in Developmental Disabilities, 2008: p. 226-248.

248. Tryon, A.S. and S.P. Keane, *Promoting imitative play through generalized observational learning in autisticlike children.* Journal of Abnormal Child Psychology, 1986. 14(4): p. 537-549.

249. Rehfeldt, R.A., D. Latimore, and R. Stromer, *Observational learning and the formation of classes of reading skills by individuals with autism and other developmental disabilities.* Research in Developmental Disabilities, 2003. 24(5): p. 333-358.

250. Tekin-Iftar, E. and B. Birkan, *Small Group Instruction for Students With Autism General Case Training and Observational Learning.* The Journal of Special Education, 2010. 44(1): p. 50-63.

251. Venn, M.L., M. Wolery, and M. Greco, *Effects of every-day and every-other-day instruction.* Focus on Autism and Other Developmental Disabilities, 1996. 11(1): p. 15-28.

252. Patten, E. and L.R. Watson, *Interventions targeting attention in young children with autism.* American Journal of Speech-Language Pathology, 2011. 20(1): p. 60-69.

253. Lovaas, O.I., R.L. Koegel, and L. Schreibman, *Stimulus overselectivity in autism: a review of research.* Psychological bulletin, 1979. 86(6): p. 1236.

254. Spiker, M.A., et al., *Restricted interests and anxiety in children with autism.* Autism, 2011: p. 1362361311401763.

255. Lovaas, O.I. and T. Smith, *A comprehensive behavioral theory of autistic children: Paradigm for research and treatment.* Journal of behavior therapy and experimental psychiatry, 1989. 20(1): p. 17-29.

256. Thurm, A., et al., *Predictors of language acquisition in preschool children with autism spectrum disorders.* Journal of autism and developmental disorders, 2007. 37(9): p. 1721-1734.

257. Hume, K., J. Plavnick, and S.L. Odom, *Promoting task accuracy and independence in students with autism across educational setting through the use of individual work systems.* Journal of autism and developmental disorders, 2012. 42(10): p. 2084-2099.

258. Nikopoulos, C.K. and M. Keenan, *Promoting social initiation in children with autism using video modeling.* Behavioral interventions, 2003. 18(2): p. 87-108.

259. Koegel, R.L. and L. Schreibman, *Teaching autistic children to respond to simultaneous multiple cues.* Journal of Experimental Child Psychology, 1977. 24(2): p. 299-311.

260. Pierce, K. and L. Schreibman, *Using Peer Trainers to Promote Social Behavior in Autism Are They Effective at Enhancing Multiple Social Modalities?* Focus on Autism and Other Developmental Disabilities, 1997. 12(4): p. 207-218.

261. Camargo, S.P.H., et al., *A review of the quality of behaviorally-based intervention research to improve social interaction skills of children with ASD in inclusive settings.* Journal of autism and developmental disorders, 2014. 44(9): p. 2096-2116.

262. Tawney, J. and D. Gast, *Single participant research in special education.* Columbus, OH: Charles Merill, 1984.

263. Senju, A., et al., *Eye contact does not facilitate detection in children with autism.* Cognition, 2003. 89(1): p. B43-B51.

264. Deguchi, H., *Observational learning from a radical-behavioristic viewpoint.* The Behavior Analyst, 1984. 7(2): p. 83.

265. Masia, C.L. and P.N. Chase, *Vicarious learning revisited: A contemporary behavior analytic interpretation.* Journal of behavior therapy and experimental psychiatry, 1997. 28(1): p. 41-51.

266. Smith, I.M. and S.E. Bryson, *Imitation and action in autism: a critical review.* Psychological bulletin, 1994. 116(2): p. 259.

267. Donnelly, J.L., P.D. Luyben, and C.S. Zan, *Increasing eye contact toward learning materials in a toddler with autism.* Journal of prevention & intervention in the community, 2009. 37(3): p. 170-176.

268. Leekam, S.R., E. Hunnisett, and C. Moore, *Targets and Cues: Gaze-following in Children with Autism.* Journal of Child Psychology and Psychiatry, 1998. 39(7): p. 951-962.

269. Sallows, G.O., T.D. Graupner, and W.E. MacLean Jr, *Intensive behavioral treatment for children with autism: Four-year outcome and predictors.* American Journal on Mental Retardation, 2005. 110(6): p. 417-438.

270. Stone, W. L. and P. J. Yoder, *Predicting spoken language level in children with autism spectrum disorders.* Autism, 2001. 5(4): p. 341-361.

271. Varni, J. W., et al., *An analysis of observational learning in autistic and normal children.* Journal of Abnormal Child Psychology, 1979. 7(1): p. 31-43.

272. Weiss, M. J. and S. L. Harris, *Teaching social skills to people with autism.* Behavior modification, 2001. 25(5): p. 785-802.

273. Garcia, E. E., *The development and generalization of delayed imitation.* Journal of applied behavior analysis, 1976. 9(4): p. 499-499.

274. Delgado, J.A.P. and R.D. Greer, *The effects of peer monitoring training on the emergence of the capability to learn from observing instruction received by peers.* The Psychological Record, 2009. 59(3): p. 407.

275. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders DSM-IV-TR fourth edition (text revision).* 2000.

276. Plaisted, K., M. O'Riordan, and S. Baron-Cohen, *Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task.* Journal of Child Psychology and Psychiatry, 1998. 39(5): p. 765-775.

277. Schreibman, L., *Effects of within-stimulus and extra-stimulus prompting on discrimination learning in autistic children.* Journal of Applied Behavior Analysis, 1975. 8(1): p. 91-112.

278. Hermelin, B., *Recent psychological research.* Early Childhood Autism: Clinical, Educational, and Social Aspects. London: Pergamon Press, Inc, 1966.

279. Koegel, R.L. and H. Wilhelm, *Selective responding to the components of multiple visual cues by autistic children.* Journal of Experimental Child Psychology, 1973. 15(3): p. 442-453.

280. Lovaas, O.I. and L. Schreibman, *Stimulus overselectiv1ty of autistic children in a two stimulus situation.* Behaviour research and therapy, 1971. 9(4): p. 305-310.

281. Lovaas, O.I., et al., *Selective responding by autistic children to multiple sensory input.* Journal of abnormal psychology, 1971. 77(3): p. 211.

282. Sailor, W. and T. Taman, *Stimulus factors in the training of prepositional usage in three autistic children.* Journal of Applied Behavior Analysis, 1972. 5(2): p. 183-190.

283. Stokes, T.F. and D.M. Baer, *An implicit technology of generalization1.* Journal of applied behavior analysis, 1977. 10(2): p. 349-367.

284. *National Autistic Society.* Available from: www.nas.org.uk.

285. Groden, J., et al., *Assessing cardiovascular responses to stressors in individuals with autism spectrum disorders.* Focus on Autism and Other Developmental Disabilities, 2005. 20(4): p. 244-252.

286. Toichi, M. and Y. Kamio, *Paradoxical autonomic response to mental tasks in autism.* Journal of autism and developmental disorders, 2003. 33(4): p. 417-426.

287. Goodwin, M.S., et al., *Cardiovascular arousal in individuals with autism.* Focus on autism and other developmental disabilities, 2006. 21(2): p. 100-123.

288. Hirstein, W., P. Iversen, and V.S. Ramachandran, *Autonomic responses of autistic children to people and objects.* Proceedings of the Royal Society of London B: Biological Sciences, 2001. 268(1479): p. 1883-1888.

289. Liu, C., et al., *Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder.* International journal of human-computer studies, 2008. 66(9): p. 662-677.

290. De Lemos, J., et al. *Measuring emotions using eye tracking.* in *Proceedings of measuring behavior.* 2008.

291. Afzal, S. and P. Robinson, *Natural affect data: Collection and annotation*, in *New perspectives on affect and learning technologies.* 2011, Springer. p. 55-70.

292. Ruble, L.A. and D.M. Robson, *Individual and environmental determinants of engagement in autism.* Journal of autism and developmental disorders, 2007. 37(8): p. 1457-1468.

293. Matsubara, Y. and M. Nagamachi. *Motivation system and human model for intelligent tutoring*. in *International Conference on Intelligent Tutoring Systems*. 1996. Springer.

294. De Vicente, A. and H. Pain. *Motivation diagnosis in intelligent tutoring systems*. in *International Conference on Intelligent Tutoring Systems*. 1998. Springer.

295. Hill, E., S. Berthoz, and U. Frith, *Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives.* Journal of autism and developmental disorders, 2004. 34(2): p. 229-235.

296. Wedel, M. and R. Pieters, *Eye fixations on advertisements and memory for brands: A model and findings.* Marketing science, 2000. 19(4): p. 297-312.

297. Voit, M. and R. Stiefelhagen. *Deducing the visual focus of attention from head pose estimation in dynamic multi-view meeting scenarios*. in *Proceedings of the 10th international conference on Multimodal interfaces*. 2008. ACM.

298. Zhang, Z., et al. *Head pose estimation in seminar room using multi view face detectors*. in *International Evaluation Workshop on Classification of Events, Activities and Relationships*. 2006. Springer.

299. Stern, J.A., L.C. Walrath, and R. Goldstein, *The endogenous eyeblink.* Psychophysiology, 1984. 21(1): p. 22-33.

300. Fogarty, C. and J.A. Stern, *Eye movements and blinks: their relationship to higher cognitive processes.* International Journal of Psychophysiology, 1989. 8(1): p. 35-42.

301. Palomba, D., et al., *Cardiac responses associated with affective processing of unpleasant film stimuli.* International Journal of Psychophysiology, 2000. 36(1): p. 45-57.

302. Jones, W., K. Carr, and A. Klin, *Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder.* Archives of general psychiatry, 2008. 65(8): p. 946-954.

303. Dautenhahn, K., et al., *Robotic Playmates*, in *Socially intelligent agents*. 2002, Springer. p. 117-124.

304. Tardif, C., et al., *Micro-analysis of social interactions between autistic children and normal adults in semi-structured play situations.* International Journal of Behavioral Development, 1995. 18(4): p. 727-747.

305. Wagner, J., J. Kim, and E. André. *From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification.* in *IEEE International Conference on Multimedia and Expo.* 2005.

306. Benedek, M. and C. Kaernbach, *Decomposition of skin conductance data by means of nonnegative deconvolution.* Psychophysiology, 2010. 47(4): p. 647-658.

307. Goldberger, J. J., et al., *Relationship of heart rate variability to parasympathetic effect.* Circulation, 2001. 103(15): p. 1977-1983.

308. Lacey, J. I. and B. C. Lacey, *Verification and extension of the principle of autonomic response-stereotypy.* The American journal of psychology, 1958. 71(1): p. 50-73.

309. Lang, P. and M. M. Bradley, *The International Affective Picture System (IAPS) in the study of emotion and attention.* Handbook of emotion elicitation and assessment, 2007. 29.

310. Scherer, K. R., H. G. Wallbott, and A. B. Summerfield, *Experiencing emotion: A cross-cultural study.* 1986: Cambridge University Press.

311. Asendorpf, J. and K. Scherer, *The discrepant repressor: Differentiation of coping style by autonomic-facial-verbal behavior patterns.* Journal of Personality and Social Psychology, 1983. 45: p. 1334-1346.

312. Cover, T. and P. Hart, *Nearest neighbor pattern classification.* IEEE transactions on information theory, 1967. 13(1): p. 21-27.

313. Pudil, P., J. Novovičová, and J. Kittler, *Floating search methods in feature selection.* Pattern recognition letters, 1994. 15(11): p. 1119-1125.

314. Lord, C. and J.P. McGee, *Educating children with autism.* 2001: National Academic Press Washington DC.

APPENDIX A

AC   Affective Computing

ADC   Analog-to-digital Converter

ADOS   Autism Diagnostic Observation Schedule

AI   Application Program Interface

ANOVA   Analysis of Variance

API   Application Programming Interface

ASD   Autism Spectrum Disorders

ASDI   ASD Individual

ASR   Automatic Speech Recognition

AuBT   Augsburg Biosignal Toolbox

BLE   Bluetooth Low Energy

CAI   Computer-assisted Instruction

CBI   Computer-based Instruction

CTD   Constant Time Delay

DD   Developmental Disabilities

DP   Development Profile

ECG   Electrocardiography

EDA   Electrodermal Activity

EL   Experiential Learning

EVL   Experiential-plus-vicarious Learning

FMD	Functional Mental Disability
GSR	Galvanic Skin Response
GUI	Graphical User Interface
HRV	Heart Rate Variability
IBI	Inter-beat Interval
ID	Intellectual Disability
IM	Imitation
IQ	Intelligence Quotient
IRB	Institutional Review Board
JA	Joint Attention
k-NN	k-Nearest Neighbors
LED	Light-emitting Diode
MD	Mental Disability
NB	Naïve Bayes
OL	Observational Learning
PA	Pedagogical Agent
PCB	Printed Circuit Board
PD	Photo Diode
PGA	Programmable-gain amplifier
PPG	Photoplethysmography
PSG	PPG+SKT+GSR
RP	Robotic Peer
SC	Skin Conductance

SCL   Skin Conductivity Level

SCR   Skin Conductivity Response

SFFS   Sequential Forward Feature Selection

SGI   Small-group Instruction

SKT   Skin Temperature

SWR   Sight Word Reading

TI   Texas Instruments

TT   Turn Taking

TTS   Text to Speech

VL   Vicarious Learning

VR   Virtual Reality

WOZ   Wizard of Oz

WPPSI   Wechsler Preschool and Primary Scale of Intelligence

CURRICULUM VITAE

Mohammad Nasser Saadatzi

Machine Learning and Interactions Lab

Department of Electrical and Computer Engineering

University of Louisville, Louisville, KY, 40292

Email: mn.saadatzi@louisville.edu

Education

- Ph.D. in Electrical and Computer Engineering   Jan. 2012 – Dec. 2016

University of Louisville, KY, USA

GPA: 4.0

- Master of Science in Electrical Engineering   Sept. 2007 – July 2010

Iran University of Science and Technology, Tehran, Iran

GPA: 4.0

- Bachelor of Science in Electrical Engineering   Sept. 2000 – July 2004

Yazd University, Yazd, Iran

GPA: 3.62

Selected Publications

- Robert Pennington, Mohammad Nasser Saadatzi, Karla Welch, and Renee Scott. "Using Robot-Assisted Instruction to Teach Students with Intellectual Disabilities to Use Personal Narrative in Texts Messages," Journal of Special Education Technology, vol. 29, no. 4, 2014.

- Mohammad Nasser Saadatzi and Javad Poshtan, "Novel system identification method and multi-objective-optimal multivariable disturbance observer for electric wheelchair," ISA Transactions, vol. 52, no. 1, pp. 129-139, 2013

- Mohammad Nasser Saadatzi and Javad Poshtan, "Application of MIMO Disturbance Observer to Control of an Electric Wheelchair Using NSGA-II," Journal of Medical Signals and Sensors, pp. 122-129, May 2011.

- Mohammad Nasser Saadatzi, Karla Conn Welch, Faezeh Tafazzoli, and James Graham, "EmotiGO: Bluetooth-Enabled Eyewear for Unobtrusive Physiology-based Emotion Recognition," in 12th IEEE Conference on Automation Science and Engineering, Texas, August 2016.

- Mohammad Nasser Saadatzi, Karla Conn Welch, Robert Pennington, and James Graham, "Towards an Affective Computing Feedback System to Benefit Underserved Individuals: An Example Teaching Social Media Skills," in International Conference on Human-Computer Interaction, Las Vegas, Nevada, July 2013.

- Mohammad Nasser Saadatzi, Karla Conn Welch, and James Graham, "Robotic assistance for individuals with Autism: A survey of current technology and insight into future directions," in ISCA International Conference on Computer Applications in Industry and Engineering, New Orleans, Louisiana, Nov. 2012.

- Robert Pennington, Mohammad Nasser Saadatzi, and Karla Conn Welch, "Using Robots to Teach Texting to Students with Intellectual Disabilities," in Ohio Center for Autism and Low Incidence National Conference (OCALICON

2013- national conference and exhibition on autism, assistive technology, and low-incidence disabilities), Columbus, Ohio, Nov. 2013.

- Mohammad Nasser Saadatzi, Javad Poshtan, and Mohammad Hossein Saadatzi, "Optimal Multivariable Two-Degree-of-Freedom Controller for Electric Wheelchair Using Non-dominated Sorting Algorithm-II," in Proc. IEEE Conference on Control Applications, Colorado, USA, Sept. 28-30, 2011.

- Mohammad Nasser Saadatzi and Javad Poshtan, "Multivariable Robust Controller for Linear and Angular Velocities of Electrically-Powered Wheelchair Using Characteristic Loci Method," in Proc. IEEE Conference on Control Applications, Yokohama, Japan, Sept. 8-10, 2010.

Professional Experience

- University of Louisville, Electrical and Computer Engineering Department (August 2015 – December 2016)
  - o Position: Teaching Assistant
  - o Courses: Digital Signal Processing, Signals and Linear Systems, Electronics, Logic Design

- University of Louisville, Electrical and Computer Engineering Department (August 2013 – July 2015)
  - o Position: Research Assistant
  - o Project: Novel visualization and emotion-sensitive digital storybook systems for enhancing reading comprehension and literacy skills of children with Autism

- University of Louisville, Electrical and Computer Engineering Department (January 2012-July 2013)

    o Position: Research Assistant

    o Project: Robotic Monitoring and Early Treatment for Children with Autism

Honors and Awards

- National Science Foundation (NSF) Doctoral Consortium Award (July 2016)

    o Issuer: NSF - Divisions of Computer & Information Science & Engineering (CISE) and Information and Intelligent Systems (IIS)

- National Science Foundation (NSF) Innovation Corps Award (February 2016)

    o Issuer: University of Louisville-NSF I-Corps

- Diebold Scholarship Award (March 2016)

    o Issuer: J.B. Speed School of Engineering and the Community Foundation of Louisville

- Engineering Exposition Graduate Award (March 2016)

    o Issuer: University of Louisville

- Speed Up Entrepreneurial Competition Award (August 2015)

    o Issuer: University of Louisville