# SOCIAL INTERACTION AND DEVELOPMENT TOOLS FOR PEOPLE WITH AUTISM SPECTRUM DISORDER

---

A Thesis Presented to

the Faculty of the Department of Computer Science

University of Houston

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

By

Xi Wang

May 2016

# SOCIAL INTERACTION AND DEVELOPMENT TOOLS FOR PEOPLE WITH AUTISM SPECTRUM DISORDER

Xi Wang

APPROVED:

Weidong Shi, Chairman
Dept. of Computer Science

Shou-Hsuan Huang
Dept. of Computer Science

Xiaojing Yuan
Dept. of Engineering Technology

Dean, College of Natural Sciences and Mathematics

# ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Weidong Shi for the continuous support of my study and research, for his patience, motivation, enthusiasm, and deep knowledge. His guidance helped me in all the time for research and writing of my thesis. Without him, I would not have published all these papers. At the same time, I thank research scientist Yang Lu for encouraging me to stay in the lab and supporting me emotionally through the rough road to finish this thesis.

I thank Prof. Omprakash Gnawali, for continuously giving me new ideas. I have learnt from him how to organize and present ideas, and how to search interesting and meaningful topics for research.

I thank the experts who participated in my research projects: Prof. Katherine A. Loveland from University of Texas Health Science Center, and Prof. Dorothea C. Lerman from University of Houston Clear Lake Center for Autism and Developmental Disabilities. Without their expertise in Autism research and treatment, strong participation and input, my research and experiments could not have been able to be conducted.

I thank Xi Zhao, Lei Xu, Zhimin Gao, Tao Feng, Nicholas Desalvo, who as very good friends, were always willing to help and give their best suggestions. It would have been a lonely lab without them.

I thank the rest of my thesis committee: Prof. Shou-Hsuan Huang, and Prof. Xiaojing Yuan, for their encouragement, insightful comments, and hard questions.

Last but not the least, I would like to thank my family: my parents and my grandparents for supporting me spiritually throughout writing this thesis.

# SOCIAL INTERACTION AND DEVELOPMENT TOOLS FOR PEOPLE WITH AUTISM SPECTRUM DISORDER

---

An Abstract of a Thesis

Presented to

the Faculty of the Department of Computer Science

University of Houston

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

By

Xi Wang

May 2016

# Abstract

It is critical to promote social interaction and development in people with Autism. The deficit in social relationship building, including communication, may lead to decreased independent living or even severe mental health problems. Using eye contact is an important nonverbal communication behavior that most of us use automatically in social interactions. However, making eye contact with others can be demanding for people with Autism - adults as well as children. Other than the lack of eye contact, sensory issues are also very common for people with Autism. Some are hypersensitive. They have difficulty filtering out background sounds to focus on one particular voice source. Consequently, it can be very hard to join a conversation or remain in it upon establishing a connection with others via eye contact or brief verbal greeting.

The thesis proposes a series of novel tools to help people with Autism handle challenges during social interaction as mentioned above by leveraging the widely-recognized wearable technologies. The first tool can remind people with Autism to make eye contact by displaying a prompt on a head-mounted display. The second tool adopts virtual reality technology to train children to initiate and hold the eye contact via a fading prompt. The third tool allows people with Autism to focus on a single auditory stream (a persons voice) based on their preference of conversation participants by detecting the participants' angular position.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Autism-spectrum disorder (ASD) is a set of developmental disabilities affecting how the brain processes information, causing delays and changes in socialization, communication, and overall behavior [14].

The number of people diagnosed with ASD has increased dramatically. According to the Centers for Disease Control and Prevention (CDC), the rate of ASD in the United States has risen to its highest level in recent decades [42]. The CDC reports that about 1 in 68 children has been diagnosed with ASD [8]. ASD is very broad. It can affect children and adults, occurring in all races, ethnicities, and socioeconomic groups, from a brilliant scientist to a person who remains nonverbal with a severe disability [49]. Autism is the most common Autism-spectrum disorder.

Two of the most important are problems with social situations (e.g., poor eye contact), and sensory sensitivities [49]. Our goal is to build tools to help people with Autism overcome these problems.

People identified with Autism have been found to share similar symptoms including but not limited to poor eye contact. Eye contact is integral in effective communication because it allows a person to focus their eyes, ears, and mind on the sources of information. It also confirms to the speaker that the listener is attentive to what she/he is saying. This can give the speaker confidence in the message that she/he is delivering and facilitate further communication. People with Autism not making eye contact suffer from many social issues including an inability to communicate effectively, and solitude.

An inexpensive and easy-to-use solution for reminding people with Autism to make eye contact in real-life social scenarios could help them improve their eye contact in daily life and to lead to better social relating.

The first tool we have developed is a wearable eye-contact reminder system using computerized-eyewear which can make the user aware when further efforts are needed to establish eye contact. When a person other than the user is speaking, a prompt pops up on the screen of the eyewear indicating the general direction of the speaker and alerts the user to look at the speaker. Though traditional therapies have been fundamental in Autism treatment, our tool can be a useful supplement. It can remind the user to make eye contact, thus aid in building confidence in social interactions.

Researchers find that early intervention for children with Autism is highly effective [20] . The second tool is mainly intended to help children with Autism reinforce eye contact.

Previous studies suggest that diversity of prelinguistic pragmatic skills (e.g., eye contact and joint attention) act as a predictive element of subsequent vocabulary acquisition rates [29] which place these children at high risk of said effects. Also, it has even been suggested that poor eye contact can negatively affect previous educational

gains of children with Autism. This is due to the direct relationship between eye contact and the ability to perceive and carry out the teacher and instructional requests [24][37]. Since Autism profoundly affects eye-contact rates, it must be treated aggressively to limit the negative impact that it may have on other aspects of the child's life.

Current methods of teaching eye contact suggest prompts. There are two types of prompts: A gesture prompt such as signaling towards the eye or putting a piece of food that is of interest to the child [38]; and physical prompts such as guiding the child's head so that it is oriented towards the teacher. While useful in establishing eye contact there are some notable limitations in these approaches. One such limitation is that they are difficult to fade out or eliminate while continuing to hold the eye contact of the child. Besides, they are quite intrusive in that they interfere with natural social interactions. The problem in using a prompt without a way to fade out or unintrusively eliminate it, is that children with Autism tend to exhibit stimulus over-selectivity and inasmuch focus on the prompt itself rather than the teachers eyes [54][47]. In effect, once the prompt is physically removed there is a high probability that they follow the prompt and not focus on the eyes. Research suggests that when prompts are directly embedded in natural stimuli children perform better [50][58]. Hence, it is critical to increasing their attention to aspects of the environment that normally command the response of the child if not affected by Autism.

The second tool that we propose could overcome the intrusiveness of the prompt. The solution is an augmented reality system making use of a virtual reality headset with a stereo video feed. The child will wear the VR headset and see the natural world in the controlled treatment areas as to which they are accustomed to. The child and the teacher will interact with each other. When the child does not make eye contact,

and a prompt is needed, the teacher does not need to make unnatural actions such as a gesture towards the face or make use of food. The teacher need only press a key, and the prompt will appear to grab the attention of the child facilitating eye contact. Then, the prompt will gradually fade away without the possibility that the child's attention will follow the prompt as if it were to be removed manually. In essence, the eye contact will remain after the prompt is no longer visible. This approach essentially holds the advantages of a prompt driven system regarding facilitating eye contact but removes its disadvantages of being intrusive and the attention shift problem that the child will follow the prompt wherever it is moved.

Other than poor eye contact, another one of the most commonly reported challenges for people with Autism is sensory differences that can make them hypersensitive to stimulation in any or all sensory modalities [16][21][36][51][48]. Sensory hypersensitivities have been linked to distress and anxiety as well as difficulties with movement [51][48]. Difficulty processing and integrating sensory information from multiple sources (e.g., faces plus voices) can add to these problems. In particular, sensory differences may be a factor contributing to difficulty in social interactions, a primary impairment found in Autism. Recent work on the effects of sensory differences on the lives of persons with Autism supports the idea that it can be hard to hold conversations with other people in part because of the need to process simultaneous streams of information, as well as the need to focus selectively on the right information [46]. Thus, while conversing in a setting with several people present, a person with Autism could become confused and overwhelmed, unable to tune out extraneous sensory information (e.g., clocks ticking, other conversations) and unable to focus on the most relevant streams of information (the face and voice of the individual with whom one is speaking). Unfortunately, sensory problems are often overlooked, according to [49],

so we attempt to develop a tool to cope with stress caused by hypersensitivity.

The third tool is to isolate people with Autism from unattended speakers in the conversation via selectively canceling their voices. The soundproof earplug can isolate the sound from the environment while outputting the filtered sound from the tool. The tool detects the speakers in the conversation by localizing their sound source directions and passes/mutes all sounds from their directions according to the white/black speaker lists. These lists, which include the preferred speakers or the unattended speakers, are manually setup by the patient via the user interface. Compared to the aforementioned strategies, our tool (i) allows people with Autism to socialize with others, focus on important speaker and mute unwanted one instead of indiscriminately shutting down all sounds, (ii) is wearable and portable for daily life and fits for the busy lifestyles of most people instead of costing much physical human intervention.

The first and third tool are aimed at assisting people with Autism in everyday settings, and the second tool is for training. They are in the articles previously published in [56][55][57].

The thesis is organized as follows. Chapter 2 gives background information. Chapter 3 deliberates the design of the three tools and provides the experiments and user study to demonstrate the feasibility, usability, and effectiveness of these tools. Chapter 4 summarizes the thesis.

# Chapter 2

# Background

## 2.1 Wearable Technologies

Wearable technology is one of the today's hottest topics. Steve Mann, the wearable computer pioneer, has been designing and wearing the computer for decades, with the gear increasing markedly in sophistication over time. In [15], he introduced the wearable computers he has invented in his life.

Google Glass is a wearable computer with an optical head-mounted display (OHMD) that is being developed by Google [1]. Google Glass is smaller and slimmer than Mann's designs. Google Glass displays information in a smartphone-like hands-free format that can communicate with the Internet via natural language voice commands [1]. The Glass can be controlled by voice commands or a side touchpad. It also incorporates an externally facing camera to record pictures and video [25].

The STAR 1200XL, we utilize for eye contact reminder and speaker cancellation is the third generation of the see-through augmented reality eyewear system. It enables you to see the real world directly through its transparent widescreen video displays.

Computer content, such as text, images and video, are overlaid on the screen [7]. It is similar to Google Glass regarding assembly, functionality and operation.

We also use Oculus Rift [2], a virtual-reality headset, to create virtual prompts for eye contact training. A virtual-reality headset is a head-mounted device aimed to provide an immersive virtual-reality experience, for the purpose of computer games and 3D simulations. It consists of a stereoscopic head-mounted display (providing separate images for each eye) and head-motion-tracking sensors (which may include gyroscopes, accelerometers, structured light systems, etc.) [6]. The Oculus Rift is also arguably the best-supported virtual-reality headset [32].

There are other virtual reality devices on the market. Google Cardboard is the least-expensive virtual-reality experience we could purchase. Cardboard viewers are designed to work with nearly any phone [32]. YouTube has announced that it now supports virtual-reality videos, allowing anyone with an Android smartphone and a Google Cardboard headset to explore 360-degree virtual worlds in 3D. YouTube said the feature would come to the YouTube app for Apple's iPhone soon. There is already a selection of virtual reality videos available on YouTube [17]. Samsung's Gear VR is one of the most impressive and easy-to-use headsets available right now [32]. Gear VR is a product made by Samsung and Oculus to sell Samsung phones [32]. It has launched the official Netflix app for Gear VR which can stream video from within virtual reality [32].

There are a lot of applications of these devices for healthcare.

Engineers at Medopad noticed that the doctors had real challenges performing the essential tasks involving patient records, accessing scans, or having blood results available when they needed them [52]. Google Glass allows up to five clinicians to collaborate in real time, take pictures and share them, and access a patients records

simultaneously [9]. Rafael Grossmann a General and Trauma Surgeon has blogged about the value of Glasss ability to share a direct point of view perspective. This could be a useful source for telemedicine or education purposes. As an example, Dr. Grossman inserted a PEG (Percutaneous Endoscopic Gastrostomy) while wearing Glass and streamed the lives images to an iPad remotely [25]. Houston (UH) Graduate College of Social Works Virtual Reality Clinical Research Lab uses computer-generated virtual environments to help people with addictions, behavior and mental health. People are not just sitting in a traditional therapists office, pretending they're in that environment, the lab actually puts them or immerses in a virtual bar, or a virtual heroin-using situation. The therapist is in the environment with them and can teach them skills in real time, including skills to prevent relapse [10].

## 2.2 Applications for Autism

One important objective of the therapy programs is to help people with Autism adapt to social situations including eye contact when in a conversation or otherwise appropriate social situation [13][26].

An emerging type of therapy involves the use of robotics, which requires less human intervention. Different roles of robots include therapeutic playmates, social mediators, and model social agents. Robots have been used to study the therapeutic effects of social interactions between humans and robots [19][18]. For example, in [31], the researchers built a small creature-like robot, Keepon, which was carefully designed to engage children with and without Autism in playful interactions. My eye-contact-reminder tool, like other robotic therapies, offers people with Autism the ability to practice in daily life. But there are some advantages my tool has that other therapies,

including those with robotics lack. One is that the tool can be wearable and extremely portable, compared with some other robotic devices. Secondly, it encourages people with Autism to interact with people outside the home in real social situations, while the other robotic solutions require them to interact with robots at home instead of real people.

Virtual-Reality technology applications have recently proliferated in Autism therapy. The Cai research team implemented a Virtual Dolphinarium, which allows children with Autism to interact with the virtual dolphins and to learn communication through hand gestures [12]. Developed by the University of North Carolina, the game "Astrojumper" allows the children to use their physical movements to avoid virtual space-themed objects flying towards them. This game assists the children in developing dexterity and motor skills [23]. Researchers have also invented a VR social cognition mechanism aimed at training young adults with high-functioning Autism. This was a substantial work in that it significantly increases social cognitive measures, in theory, emotion recognition, as well as real life social and occupational functioning were found post-training [28]. While all of these works above contribute substantially to the treatment of Autism in children, adolescents, and adults, none address the lack of eye contact in children with Autism. Researchers from Vanderbilt University attempted to condition eye contact in children with Autism by creating a virtual storyteller to guide their focus on the communicator [34]. However, unlike my eye training tool, theirs did not adopt a fading prompt approach which suggests that there is still substantial gains to be made in children's conditioning of eye contact in which we have seized the opportunity to address [22][39][40][41].

Unfortunately, few wearable technologies are utilized to help people with Autism control auditory sensitivity. The most common strategy is sound isolation. People

with hypersensitivity isolate themselves from others. Some practice listening to a mildly irritating piece of a conversation. We have designed a fully-interactive tool that allows people with Autism to choose which person's speech they would like to focus on, thereby giving them a sense of control.

# Chapter 3

# System Design and Evaluation

This chapter presents the design and evaluation of each of the three tools.

## 3.1 Eye-Contact Reminder

Avoiding eye-contact behavior has been characteristic of people with Autism. Such behavior prevents intrinsic development of social and communication skills.

In this section, we introduce the first tool, a directional eye-contact-reminder system which reminds the user to generally focus her/his eyes in the direction of a human speaker. This tool detects a speaker's voice, calculates the sound direction, and directs the user's eyes by displaying a prompt on the eyewear (STAR 1200XL) [7] in the direction of the speaker.

### 3.1.1 Design

The tool consists of two microphones, the eyewear, and a computation unit (a laptop here). The microphones are mounted onto two sides of the eyewear. They collect

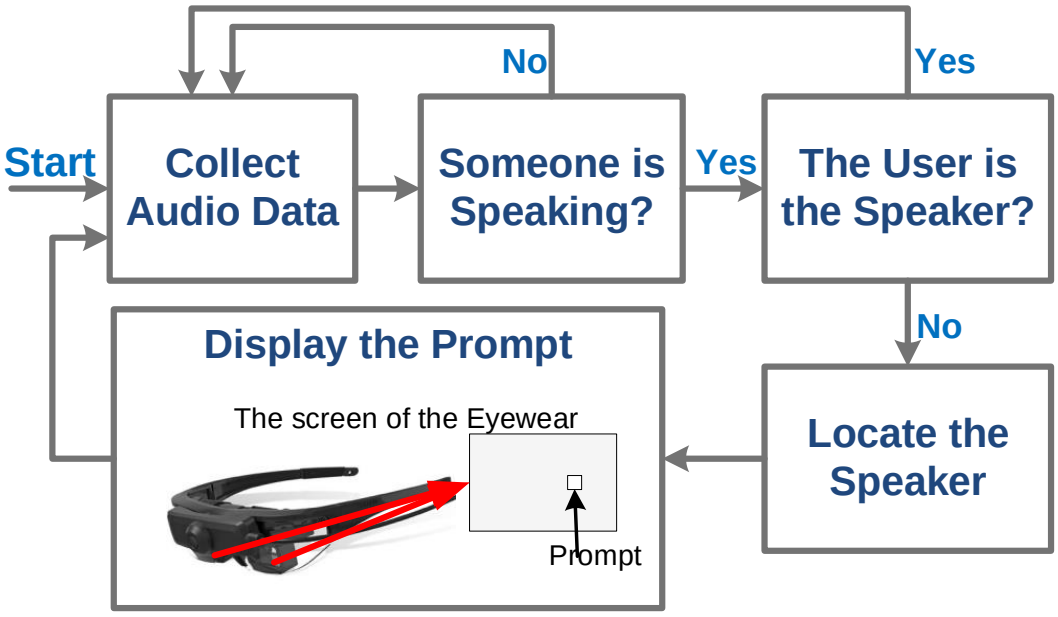



Figure 3.1: The design of the eye contact reminder tool

audio data and send it to the laptop which has the program to calculate source angle of voice. The laptop sends the calculated result back to the eyewear which then displays a prompt. The prompt may also be customized.

We use two audio processing algorithms. One is voice activity detection (VAD) based on Short Term Energy (STE) and Zero Crossing Rates (ZCR) [43][45]. This splits the signals into overlapping frames [44], extracts STE and ZCR features of framed signals, and compares the calculated thresholds to determine the onset and termination of speech boundaries. Another one is the sound localization algorithm named Jeffress Model[27]. It is a hypothetical model of how neurons in the brain make use of time differences. With these two algorithms, the tool can determine if someone is speaking as well as locate the speaker.

As depicted in Fig. 3.1, the first module is to collect audio data. Audio data is $X = [X_l, X_r]$, where the $X_l$ and $X_r$ are two arrays of the same length received by

the left and right microphone respectively. The length of $X$ (or $X_l$, $X_r$ ) is decided by frequency of sampling and the time of recording. The sampling frequency in our system is $44100/s$. Each time, the microphones collect 10000 samples of sound data X called an index, which is $10000/44100s$ long data.

In the second module, we feed the 10000 samples into the VAD. If no parts of speech are recognized as active, the tool determines that no one is speaking and doesn't display the prompt on the screen, so the user does not need to prepare to participate a conversation. $V$ consists of subsequences of $X$. In $V = \{V^1, V^2, \cdots V^n\}$, $V^i$ is an active part of a speech. All active parts are separated and picked using the VAD.

If VAD finds some parts of the speech are active, the tool requires the third module, which decides if the sound is made by the user or is extraneous. If this module is not employed and the speaker is the user, the prompt would be placed in the middle of the screen because the distances between the mouth of the user and the two microphones are the same. It is unnecessary and potentially confusing since no other person is speaking to the user at that point.

We use sound loudness $E$ to decide if the speaker is the user. Just like human ears which can detect the sound loudness levels, data collected by microphones can also reflect this. Loudness of the source sound decreases exponentially with distance: $E \sim \frac{1}{r^2}$, where $E$ is sound loudness and $r$ is the distance between the speaker and the microphone. The distance between the microphone and the user is around 5-8 cm, with the normal range for the distance between the speaker and the user being roughly at least one meter. Sound loudness calculated between the speaker and the microphone should be theoretically 156 to 400 times larger than sound power between the user and the microphone. We compare the sound loudness to a threshold and

| | $5cm$ | $10cm$ | $100cm$ |
|---|---|---|---|
| person1 | 768.7(l)<br>2046.8(m)<br>5097.1(h) | 327.3(l)<br>1921.0(m)<br>3895.2(h) | 9.0(l)<br>84.5(m)<br>281.8(h) |
| person2 | 2273.2(l)<br>4946.7(m)<br>8495.1(h) | 452.8(l)<br>2273.2(m)<br>4674.2(h) | 9.6(l)<br>93.9(m)<br>301.1(h) |
| person3 | 869.7(l)<br>3006.8(m)<br>5327.1(h) | 382.8(l)<br>2021.7(m)<br>3292.2(h) | 8.7(l)<br>88.5(m)<br>288.6(h) |

Table 3.1: Average sound loudness of each speaker with different volumes while standing at different distances from the microphones

estimate whether the speaker is the user.

Let $T_1, T_2, \cdots T_n$ represent the corresponding sample number of active parts. They can reflect the lasting times of active parts since $time \sim number\ of\ sample$, so we skip converting sound sample number into time like $s$ or $ms$. $V_{l1}, V_{l2}, \cdots V_{ln}$ or $V_{r1}, V_{r2}, \cdots V_{rn}$ is the data from left or right microphone. Normally $E$ would be almost the same no matter using $V_{l1}, V_{l2}, \cdots V_{ln}$ or $V_{r1}, V_{r2}, \cdots V_{rn}$. We choose $V_{r1}, V_{r2}, \cdots V_{rn}$. $E_1, E_2, \cdots E_n$ are the sound loudness of each active parts. Every $E_j$ is calculated like this: $E_j = \sum_{x \in V_{rj}} x^2$. We should not directly add all $E_j$ to show the magnitude of sound loudness for a whole 10000-sample-length speech. Because some speech contains more pauses than the others. The $E_j$ would be tremendously different though these speeches were spoken by the same person with consistent volume. So we need to average sound loudness: $\bar{E} = \frac{\sum E_j}{\sum T_j} \times 10^4$.

We calculate a proper threshold for sound loudness, which will be used to differentiate the user from the speaker. We collect speech samples from three participants. Each speaks with different volumes, from high to medium to low, while standing at different distances from the designated microphones (e.g. $5cm$, $10cm$ or $100cm$ away). Table 3.1.1 shows the three persons' average sound powers. We refer to $SD$ as the distance between speaker and microphone. The $\bar{E}$s with $SD = 100cm$

and $Volume = high$ are around 300 even less than the $\bar{E}$s with $SD = 10cm$ and $Volume = low$. In our experiment, low volume is normal in a social conversation. Generally, people remain low volume and switch to medium volume now and then. So according to Table 3.1.1, all $\bar{E}$s with $SD = 100cm$ and $Volume = low/medium$ are less than 100 while all $\bar{E}$s with $SD = 10cm$ and $Volume = low$ are more than 300. We choose 200 as the threshold between the user and speaker. This experiment provides a sense of the effectiveness of sound loudness discrimination ability to determine the user or speaker. In the future, we will investigate this further by recruiting more participants and taking into account their age, gender, and other factors.

The next module of our tool is to locate the speaker. We input $V_1, V_2, \cdots V_n$, the active parts of audio data into the Jeffress model. Without them, the result from Jeffress model would be incorrect because the data going into Jeffress model includes noise. Active parts contain noise as well, but human voices are loud enough to drown noise, so we consider these parts clean. The output is the angular position of sound source.

Lastly, the prompt is tagged on the transparent screen of the eyewear which alerts the user that someone is speaking and she/he should make eye contacts towards the speaker. The increase in eye contact reinforces visual and auditory coordination and supplements essential building blocks of learning and effective communication.

### 3.1.2 Evaluation

The tool includes two USB microphones (sample frequency is $44100/s$) and the eyewear worn by the user. Both microphones and eyewear are connected to a laptop via USB cables. Two microphones are placed $15cm$ apart near the left and right ends of the eyewear. A sound source, which is a human speaker, is placed $1m$ away from the

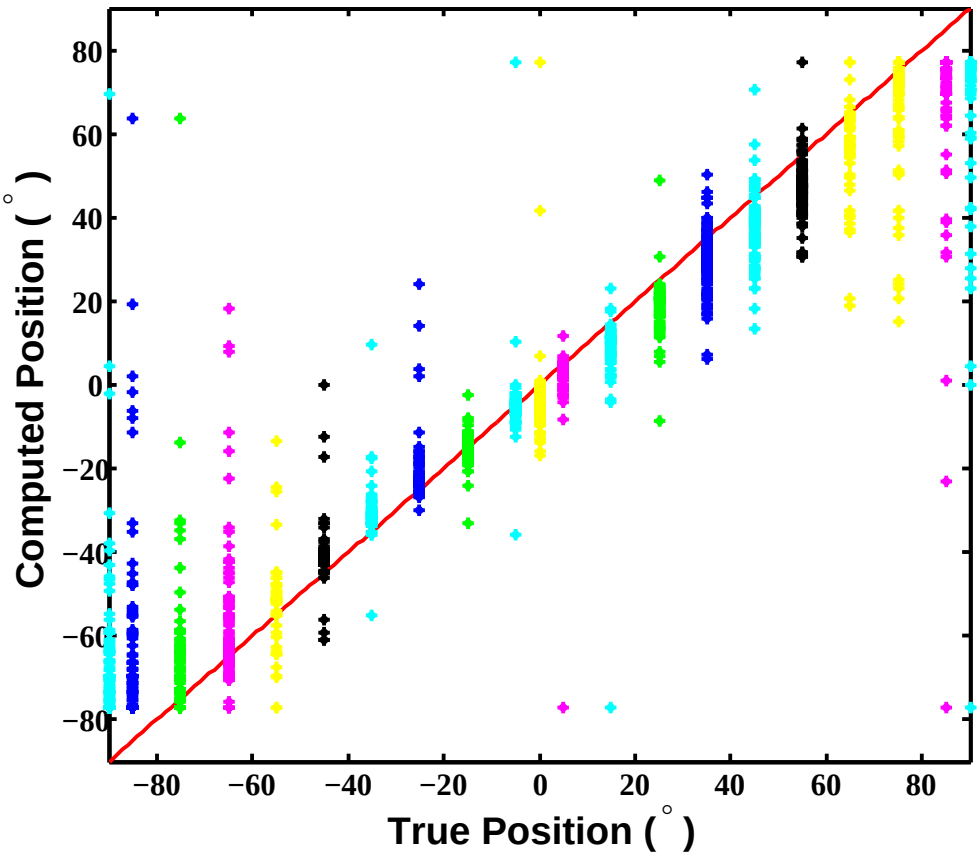
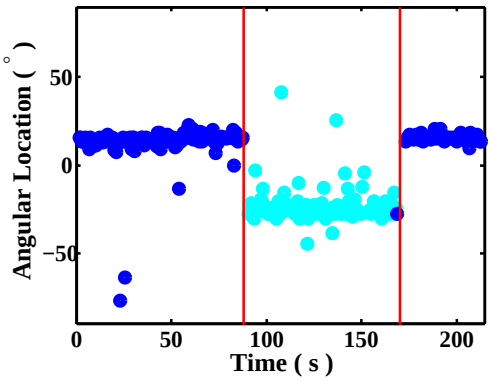



Figure 3.2: Performance of speaker localization. Each dot is computed position versus true position.

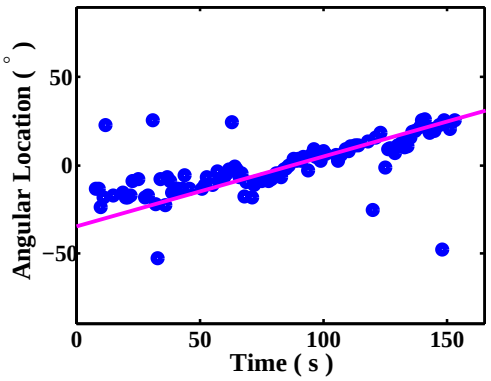
midpoint of the two microphones (also the midpoint of the eyewear). The line that is orthogonal to the plane supporting the two microphones is at 0°.

We test the validity and accuracy of the VAD and Jeffress model. We disregard any other utilized technologies to focus on the VAD and Jeffress model's performance. Speech is recorded from speakers placed at 0°, ±5°, ±15°, ±25°, ±35°, ±45°, ±55°, ±65°, ±75°, ±85°, and ±90° respectively. Each speech instance is around $20s$ long, which indicates $20 \times 44100 = 441000$ samples. We use every 10000 sample to compute position of speaker which is a dot in Fig. 3.2, so gain a set of dots. The closer the dot is to the red diagonal, the more accurate the speaker localization method is. In Fig. 3.2, dots are displayed closely around the red diagonal. It indicates that the computed positions are mainly equal to or very close to the true positions. If the absolute difference of true and computed position is not more than 10°, we consider computed position as accurate. 1308 out of 1598 dots fall within the range, so the accuracy rate is 81.85%.

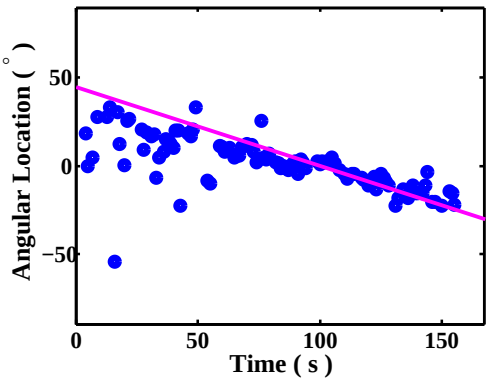
We test the entire mechanism. One participant wearing the eyewear is put in three scenarios: 1) two people speaking alternately in front of the user on left and right side respectively. 2) one person who moves slowly from left to right while speaking; 3) one person who moves slowly from right to left while speaking. Figure 3.3 suggests that our system can follow and detect different stationary/roaming sound sources in common social settings in real time.




(a) Two persons, represented by cyan (light) and blue (dark), speak alternately. Two vertical lines indicate when the speaker is changed.




(b) A person moves from left to right while speaking. The line represents the true movement direction of the speaker.




(c) A person moves from right to left while speaking. The line represents the true movement direction of the speaker.

Figure 3.3: Three scenarios

## 3.2 Eye Contact Training in Children

Children with Autism may suffer from a natural aversion to dyadic (i.e., eye-to-eye) contact. Research has shown this aversion to be an early indicator of slower development of linguistic skills, a narrow vocabulary, as well as social issues later in life. In addition, this aversion may also result in the loss of already acquired abilities such as language and life skills. Consequently, manual prompt techniques have been adopted to address this issue. However, they are plagued with some inherent flaws: (i) the teacher must make unnatural movements when using a manual prompt such as gesturing towards the face; (ii) The child's attention will follow this prompt as it is removed from the face defeating the purpose as it detracts the child's attention from the teacher's eyes.

To tackle these issues we have developed a tool that can utilize effective prompt methodologies aimed at conditioning these children to establish and maintain dyadic contact. Our tool not only reduces, but eliminates shortcomings present in the current manual method. This is accomplished through the use of a stereo camera and virtual reality headset to augment the child's vision when eye contact is not being established. The prompt is displayed in the child's vision over the eyes of the teacher to attract their attention. Once eye contact has been ascertained, the prompt is gradually fading leaving the child only to focus on the eyes of the teacher as is needed.

### 3.2.1 Design

We present the hardware and software components of our system, as explained in Fig. 3.4. The stereo camera is connected to the computation unit through a USB 2.0 interface. This data, once processing and prompt placement (an apple in this case)

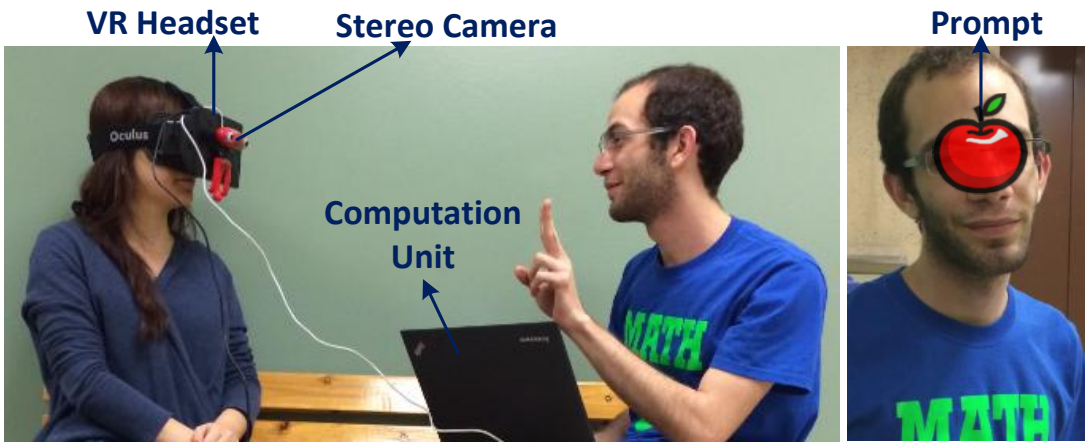



Figure 3.4: Eye contact-training tool overview: The camera captures 3D image being sent to the laptop. The laptop mounts a prompt on eye area of the teacher. The user would see the prompt fade gradually through VR headset.

have been applied, is then transferred to the VR Headset via an HDMI interface. Concurrently, required data, yet irrelevant in this context for VR Headset use is transferred back to the computation unit via another USB 2.0 interface. Note the VR Headset is connected bidirectionally to the computation unit.

#### 3.2.1.1 Hardware

The proposed real-time tool is only constrained by computational power. It has been decided the best device to use would be a desktop system in that it will create a much more fluid experience for the user. The only required constraint inherent in this tool is that the VR headset must not discomfort the user as to create a more immersive and beneficial experience where user acclimation time is minimized. The hardware system is comprised of three major entities: the computer, VR display headset, and stereo camera. Because both the stereo camera and the computer are standard they will not be addressed in much detail within this thesis.

The VR headset utilized is an Oculus Rift [2]. It is minimized in terms of weight

and also contains foam in parts that come in contact with the users face. The headset technology, as seen in Fig. 3.5, is comprised of an LCD (liquid crystal display) unit that is placed behind a lens (lens explained in depth in next paragraph) in front of the users eyes. This LCD measures seven inches diagonally, 1280x800 pixel resolution (640x800 pixel resolution for each eye since the display is split between both eyes), and 32-bit color depth [2].

The reason for hardware choices are as follows: The screen must be large enough to accommodate for minor visual overlap between the images viewed in each eye as well as the peripheral vision of each separate eye. The cone of vision for humans is not the same for the right and left eyes. We must compensate for this in the hardware. The FOV (field of view) related to the right eye extends further to the right than the left eye and vice versa for the left. In essence, images for the right and left eye must be different but overlap slightly for the brain to stitch correctly and render the images in 3D [33]. Even more importantly without significant user eyestrain. This amount of overlap between the right and left visual input is controlled by the respective subjects interpupillary distance (see Fig. 3.5). The users interpupillary distance and the amount of overlap are inversely proportional so that a larger interpupillary distance would create a smaller overlap and vice versa. When the images are not correctly overlapped, besides causing eye strain and discomfort for the user it will also detract from the benefits the user can achieve from using this device making efficacy quite less, so this is a significant issue that must be addressed correctly [35][30].

As you can see in Fig. 3.5, distance between pupils is the interpupulary distance. Inside the device is both the LCD and Lenses. While attached on the outside of the device is the stereo camera. Please also note the FOV is the total cone of vision that the user can see which is expressed in degrees.

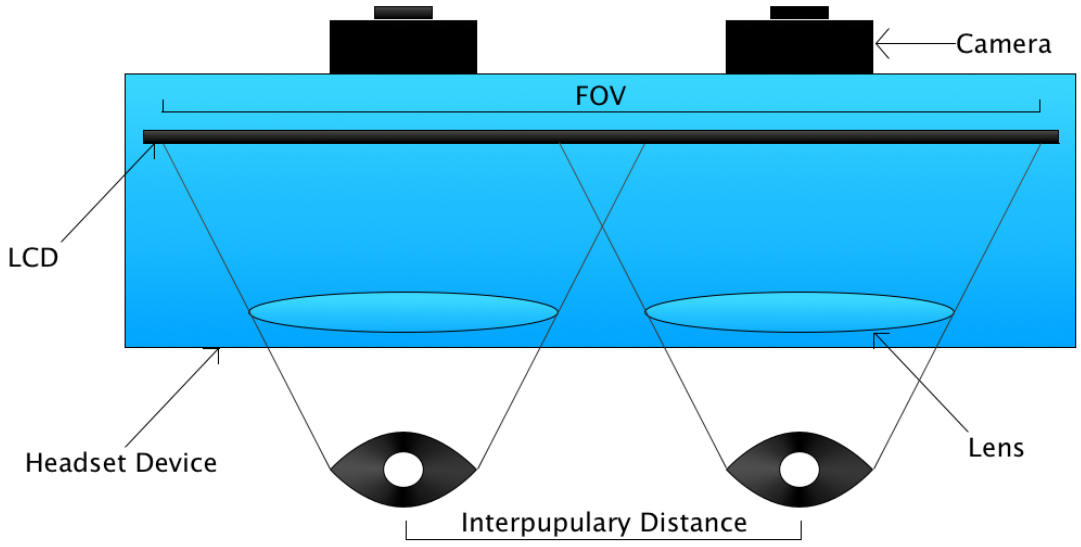
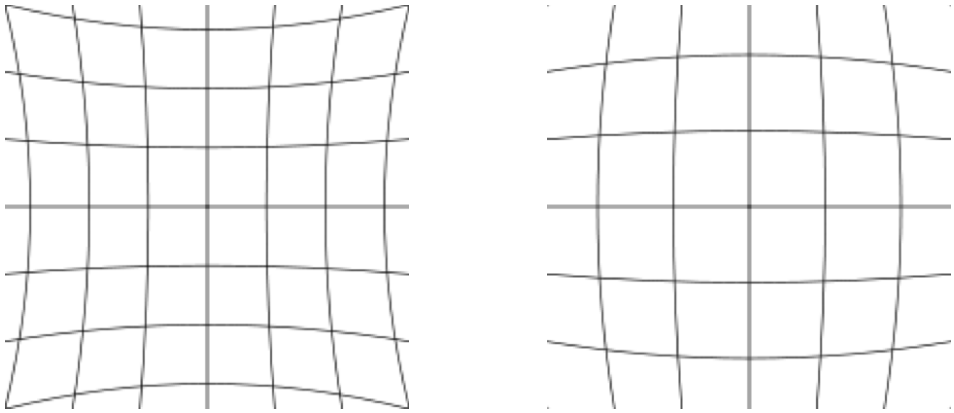



Figure 3.5: Inside, top-down view of the VR headset.

Now that this has been accomplished, a distortion by both lenses in front of the eyes as well as the video feed is imbued with a distortion to create a sense of depth perception for the user. The lenses in front of the users eyes create a pincushion distortion as seen in Fig. 3.6 (a), while the video feed placed in front of the user on the LCD has a barrel distortion applied as seen in Fig. 3.6 (b). When these two distortions are used in conjunction with each other they will effectively cancel each other out in the perception of the user. However what the user does not notice is that the pincushion distortion creates a wider FOV so that for example when the user looks 40° the light is bent such that they see 30° to the left of the LCD panel. This is how the VR headset creates a more realistic experience.

OpenGL [5] and OpenCV [3] are both used in this software (greater scrutiny will be provided within the software section following). To offload some computations and speed up the process overall, it has been determined that the barrel distortion will be done in OpenGL in Preprocessing. This will migrate these actions from the CPU on



(a) (b)

Figure 3.6: (a): Pincushion distortion; (b) Barrel distortion

to the GPU and create a more fluid experience within present real-time constraints. This also allows for a less powerful system to be used in settings that may otherwise be inhibited by this variable.

#### 3.2.1.2 Software

As seen in Fig. 3.21, the software consists of two main components: (1) a classifier training and configuration settings in the offline phase, and (2) image processing including prompt overlay in the online phase. The two parts are explained as follows:

**3.2.1.2.1 Offline** The tool adopts an object detection algorithm using Haar feature-based cascade classifiers [53]. In order to increase eye detection accuracy, the tool detects the face from the image prior to detection of the eyes, which limits the search to the area detected as the face. The teacher can use default classifiers for both face and eye detection which are supplied by OpenCV, instead of training a new classifier.

The teacher may customize two settings which are the prompt image and its respective opacity. This allows the teacher to select one image out of the provided image bank that most attracts the interest and eye contact of the autistic child to

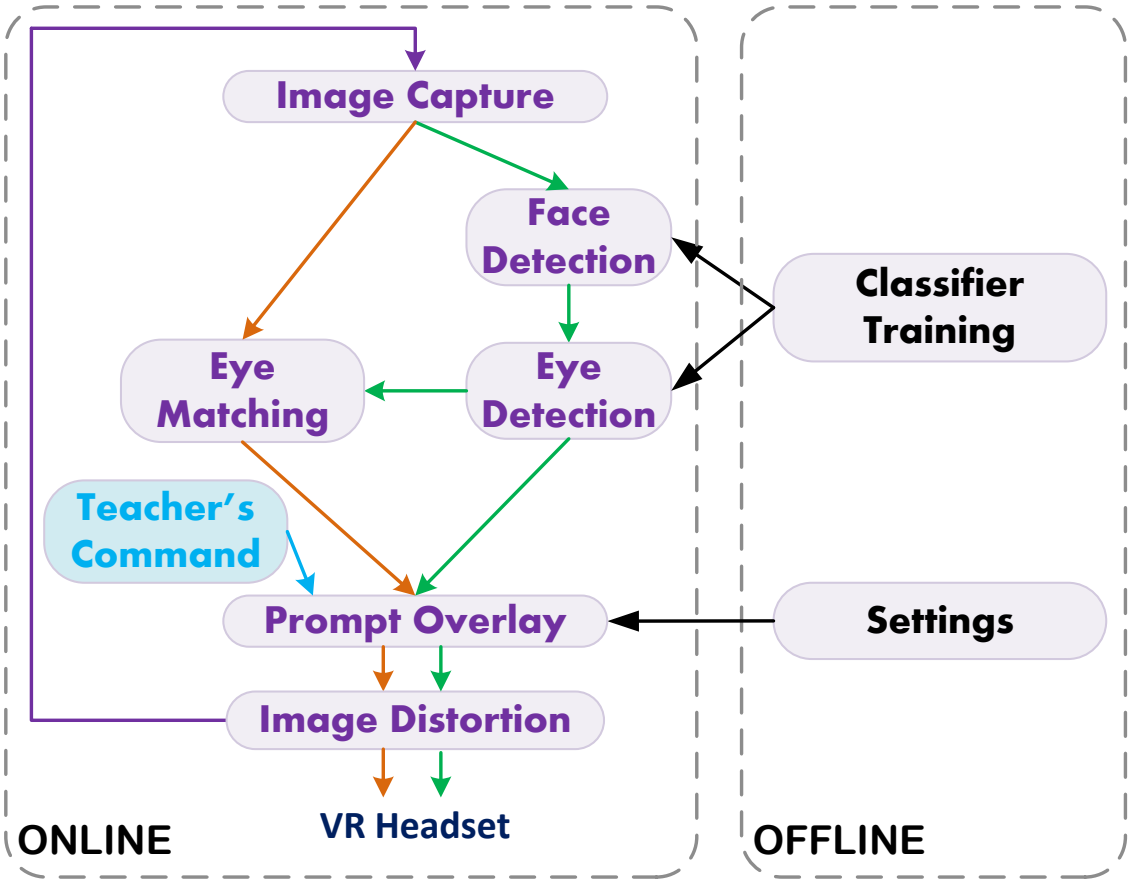



Figure 3.7: The software consists of two parts: online and offline. The online phase includes two threads, the command thread (acting as a user interface) and the image processing thread. The green and orange arrows represent processing of both left and right images from the stereo camera. The processed images are later fed into the VR headset.

be overlain on the video feed. The provided images include an apple, orange, and flower. The teacher may find that none of these images are to the liking of the child. This case would render the tool ineffective. To cope with this exception, the teacher may also upload an image not provided to be used as a prompt. The other adjustable setting is opacity, which translates to the transparency level of the overlain prompt. Opacity is expressed as a percentage in this context: 0% being completely transparent (invisible) while 100% is completely visible. The prompt gradually disappears at a constant rate. Every 100ms, the opacity level will decrease by X%, X being an integer value as defined by the teacher. While the system is running, the teacher may decide when the overlain prompt should be faded by clicking the fade button. This triggers the prompt to fade away at the predefined rate. To clarify, the teacher may also choose default settings for the two aforementioned variables. The default prompt is set to be a red apple and the default fading rate is two. This means the prompt will fade at a rate of 2% per 100ms rendering the prompt completely transparent in five seconds.

**3.2.1.2.2 Online** During the online phase, the tool is streamlined by utilizing two threads running in parallel (the command and image processing threads). The command thread waits for the teacher's prompt display command. If the teacher would like to start fading the prompt or overlay a 100% opaque prompt again, she/he may click the fade or reset buttons respectively. A button click will trigger the command thread to send the teacher's request to the image processing thread where the commands are executed. The two 2D cameras send images continuously to the image processing thread. In order to form a proper 3D image, the eye areas of each image must be synced as accurately as possible. The system does not conduct

autonomous eye detection inept of the other image fed at the same time (right and left images). It locates the eye area from one image and uses that data as template [11] to search the eye area of the other. Due to the small offset between the two images, when they are presented in front of a person, the brain renders the pictures and stitches them together accordingly in order to be perceived as a 3D image. The two prompts overlain on each image will then be merged into one prompt when the image is perceived by the user. In order to ensure highly accurate eye detection, the system detects the face first and then later detects eyes based upon the previously detected face area. Both face and eye detections need cascaded classifiers obtained from the offline phase. If no eyes are detected, the system continues detection for the following images. According to the teacher's requests, the system adjusts the opacity of the prompt, which could be from 0% to 100%, and then overlays the prompt on the two images next to the eyes. Before the two images are sent to the AR headset, a distortion is required to compensate for the VR headset. To accomplish this, OpenGL shaders are utilized. Because of constraints of the application as well as OpenGL itself, a vertex, geometric, and fragment shader are all required. All three of these shaders are loaded, compiled, and linked in one OpenGL program object. In this rendering process the program object is used on each image creating the distorted before being sent to the VR headset. The user will then physically see the image presented as two undistorted images, one in the right eye and one in the left eye. The user's brain then stitches these two 2D pictures together in order to perceive a 3D image.

### 3.2.2 Evaluation

We build a dataset consisting of ten participants' videos in order to evaluate the design of our prototyping tool. We ask each participant to stay directly in front

of Minoru 3D Webcam which records two videos, from its left and right cameras simultaneously for 40 seconds. The dataset contains 20 videos in total. In order to most closely mimic real training for autistic children, each participant was required to sit approximately 1-2 meters from the Minoru camera and to show their face without any veils or obstructions, in an adequately stable illuminated area. The computation unit in the experiments was a laptop equipped with an Intel Core I5 1.8GHZ CPU and 4GB RAM.

We evaluate eye detection performance by calculating the percentage of frames in which the eyes are successfully detected. The 10 videos recorded by the left camera were utilized. Fig. 3.8 depicts the eye detection accuracy of two methods. The first method was to locate and detect the face, then using this data detect eyes within the identified face area. The second method was to use eye detection with no other intermediary steps to locate the eyes directly. The average accuracies were 97.42% and 33.93%, respectively. The system tends to fail more in locating the eye areas if using eye detection only. Facial features are more distinct than ocular features. Hence, face detection is more reliable. Since the first method which confines eye detection to the facial areas instead of the whole image is more accurate, it has been selected as the means for eye detection.

Using the combination of face and eye detection, we found the eye areas of each video frame recorded by left camera. We used this as a template to find eye area of its peer frame recorded by the right camera. We evaluate the matching accuracies by calculating the percentage of pairs of frames whose square detection areas difference are zero. Fig. 3.9 depicts a high matching accuracy, which occurred 98.10% of the time on average. All ten trials achieve 95% accuracy or greater while three reach 100% accuracy. This results from a very minute distance between the two cameras

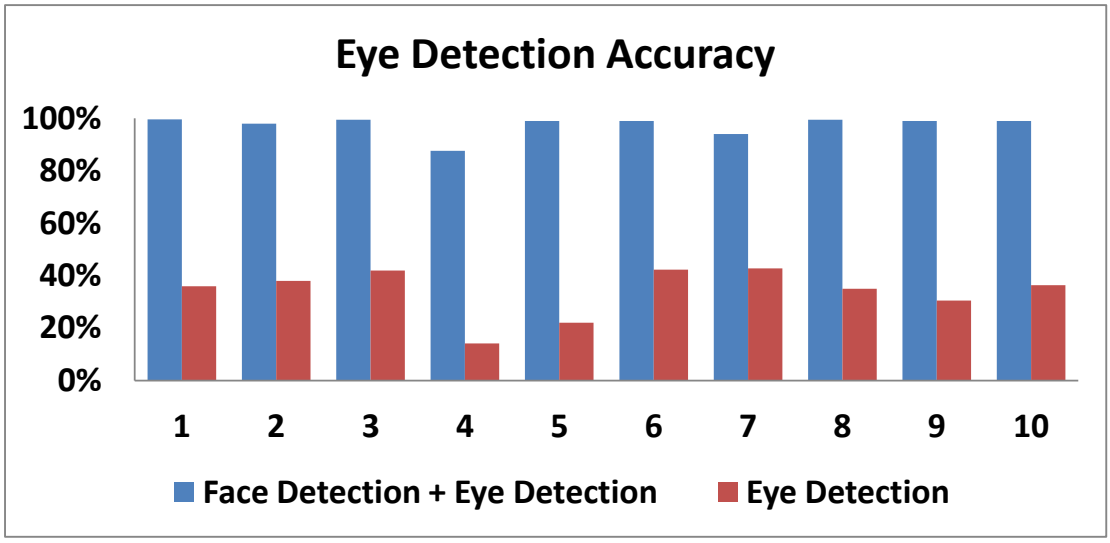



Figure 3.8: Eye detection accuracies using either a combination of face and eye detection or lone eye detection

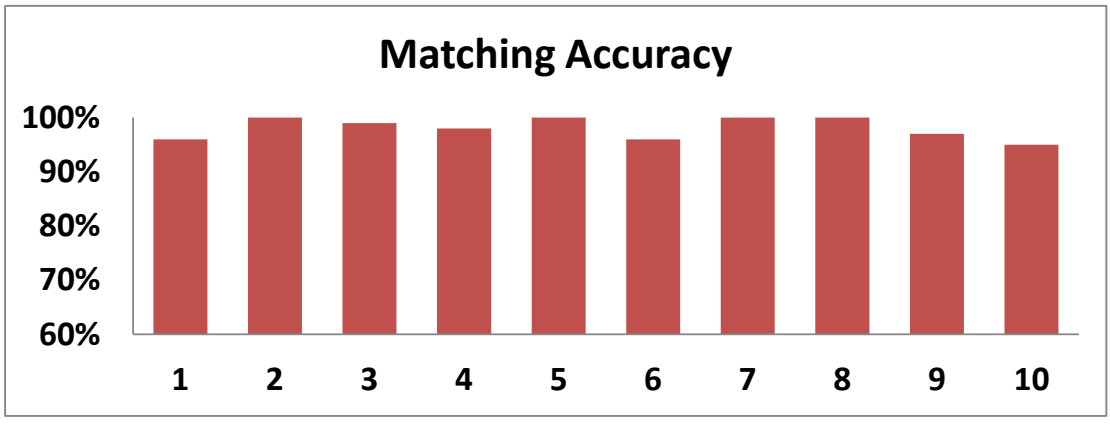



Figure 3.9: Template matching accuracy

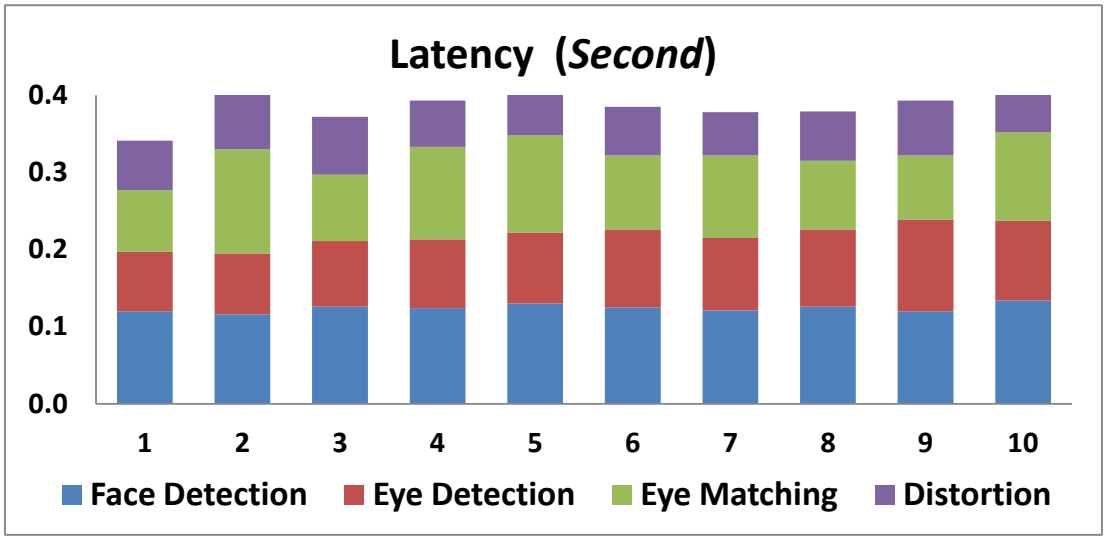



Figure 3.10: Latency

and ten participants sitting directly in front of the two cameras and showing their unobstructed face.

Moreover, we investigate the time latency from the time that two images are captured at the time that they are processed and sent to VR headset. Fig. 3.10 depicts respective latency of face detection, eye detection, eye matching, distortion, and the whole latency on 10 pairs of videos. The mean of the combined latencies is 0.387s, and the mean latency of each step is 0.124s, 0.939s, 0.103s, and 0.656s respectively. Due to the fact that the teacher should not move substantially, the latency won't cause strong discontinuities in the video feed the user sees. One other way to tackle this issue would be to, as suggested in the hardware section, increase the power of the computation unit depending upon the setting and constraints.

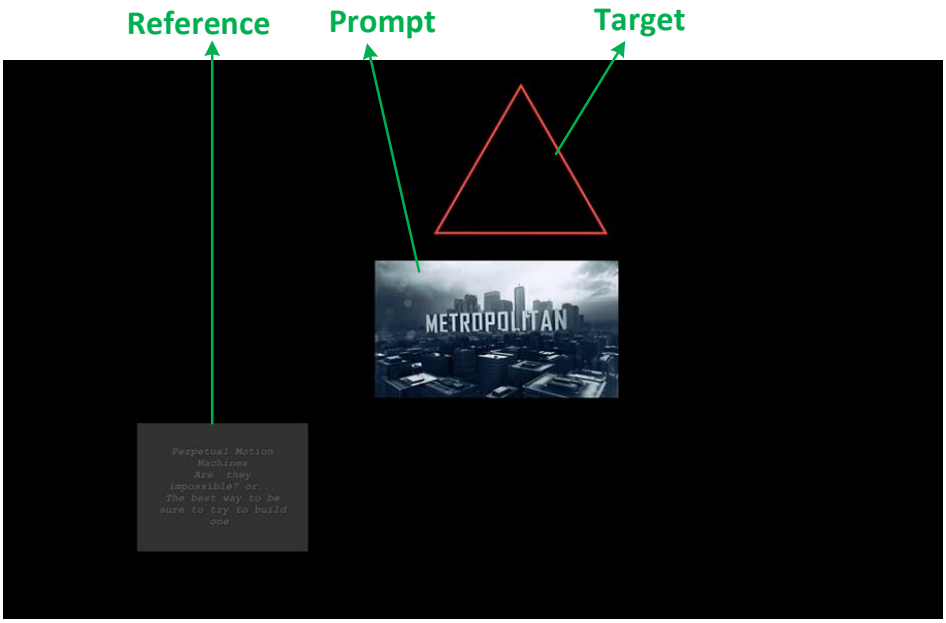


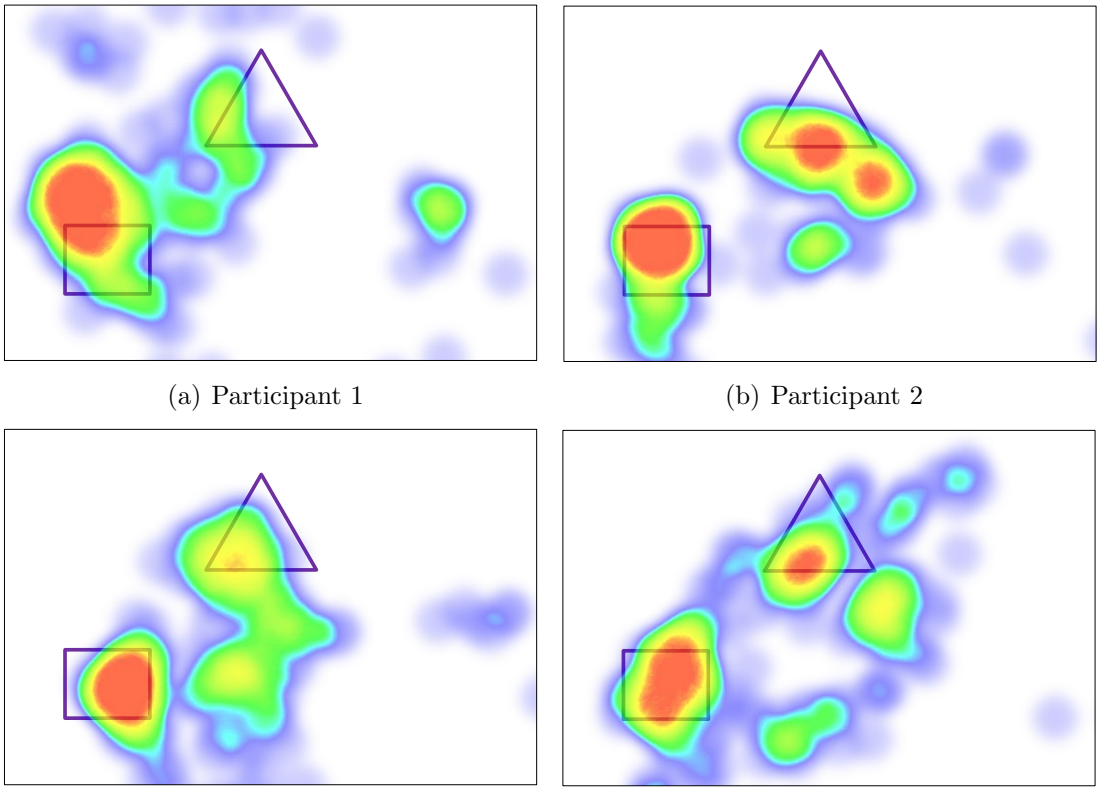

Figure 3.11: Each video includes a target, a reference, and a moving prompt.

### 3.2.3 User Study

#### 3.2.3.1 Setup

We make four videos with black backgrounds, each of which contains three objects as seen in Fig. 3.11: (1) A target which is a stationary, non-filled, red-bordered triangle in the upper middle section, (2) A reference which is a stationary video window on the bottom left displaying an instruction on mechanical assembly, and (3) A prompt which is a moving window displaying a movie trailer.

The target and reference stay visible throughout each video. In the first video, which lasts 10 seconds, the prompt appears in bottom right at six seconds and slowly moves towards the target. Once on the target, it ceases movement. In the rest of three movies, which all last 25 seconds, the prompt appears from bottom right corner, slowly moves towards the target, and at 21 seconds it disappears.



(a) Participant 1

(b) Participant 2

(c) Participant 3

(d) Participant 4

Figure 3.12: Heatmaps of four participants watching the first video. The rectangle and triangle represent the positions of the reference and the target respectively.

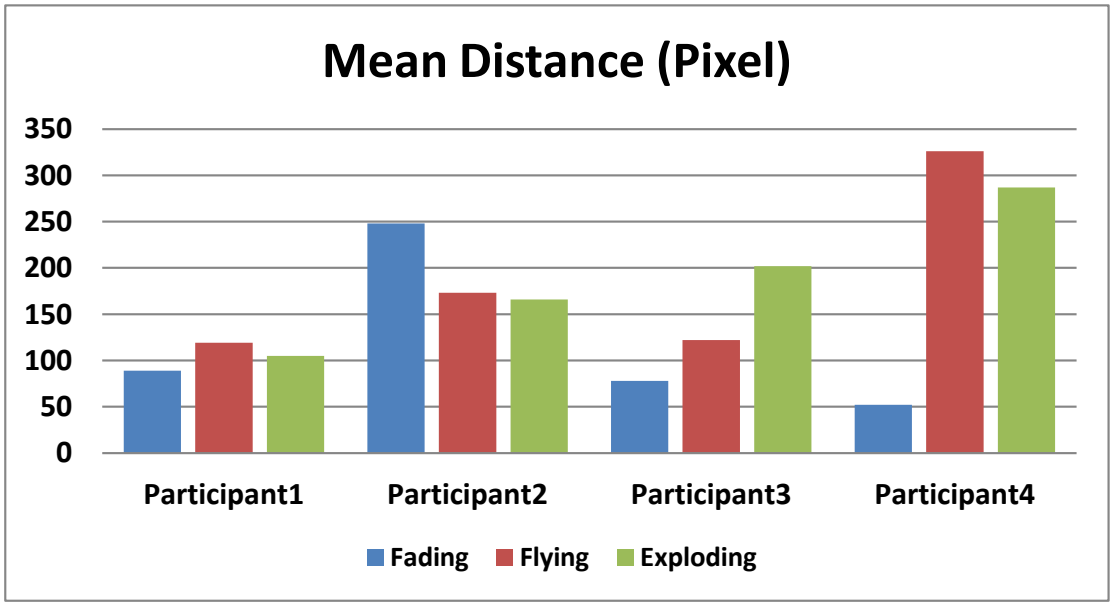



Figure 3.13: Mean distance between gaze and the target center

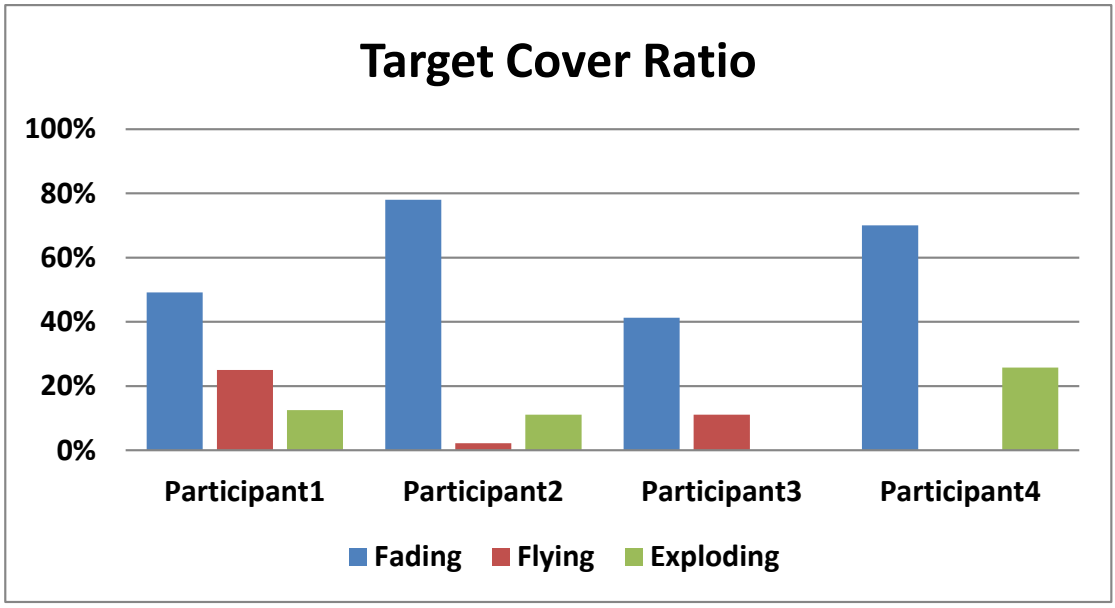



Figure 3.14: Ratio of gaze within target

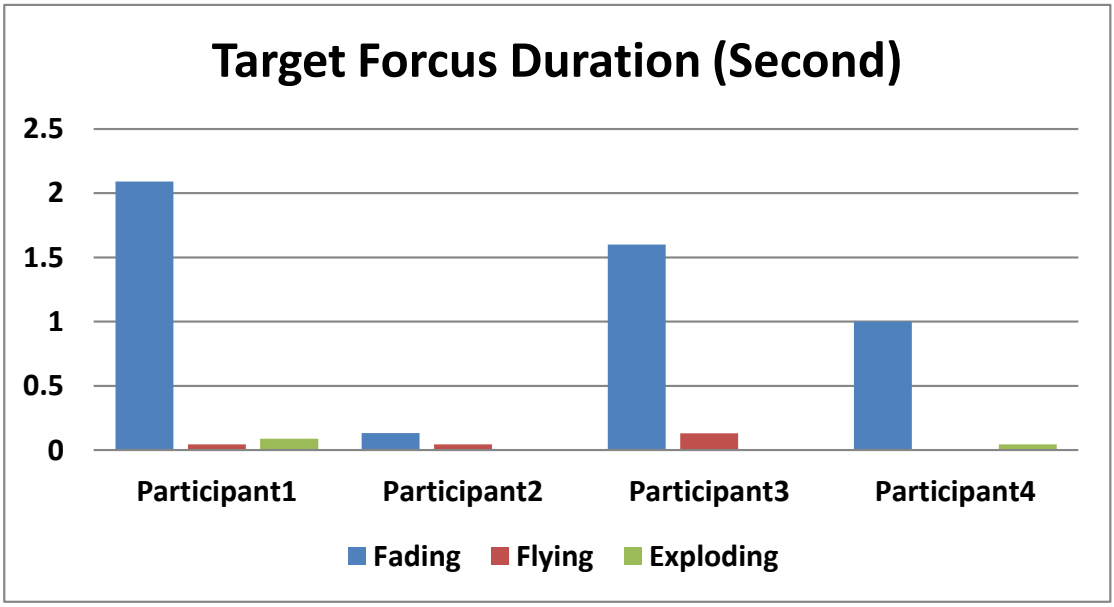



Figure 3.15: Duration of participants' focus on the target

The prompts in the three videos disappears in three different ways: (1) It stays on the target and fades out over time, (2) As is current conventions when used in eye contact training for autistic children, the prompt moves away from the target and out of the video at upper left corner, and (3) It stays on the target and explodes. After the prompt is no longer in the video, the reference and the target remain for five additional seconds.

We design a user study to answer two fundamental questions: (1) Whether the virtual moving prompt draws participant's attention when he/she focuses on the reference and not the target, and (2) Which of the three prompts holds the participant's eye contact the longest.

We invite four participants (college students) and use Opengazer[4] to record participants' gaze positions in the video. In order to ensure the tests are not affected by extrinsic factors the participants were instructed to rate the movie trailers and were not informed of why calibrations were needed prior to the four videos as well as

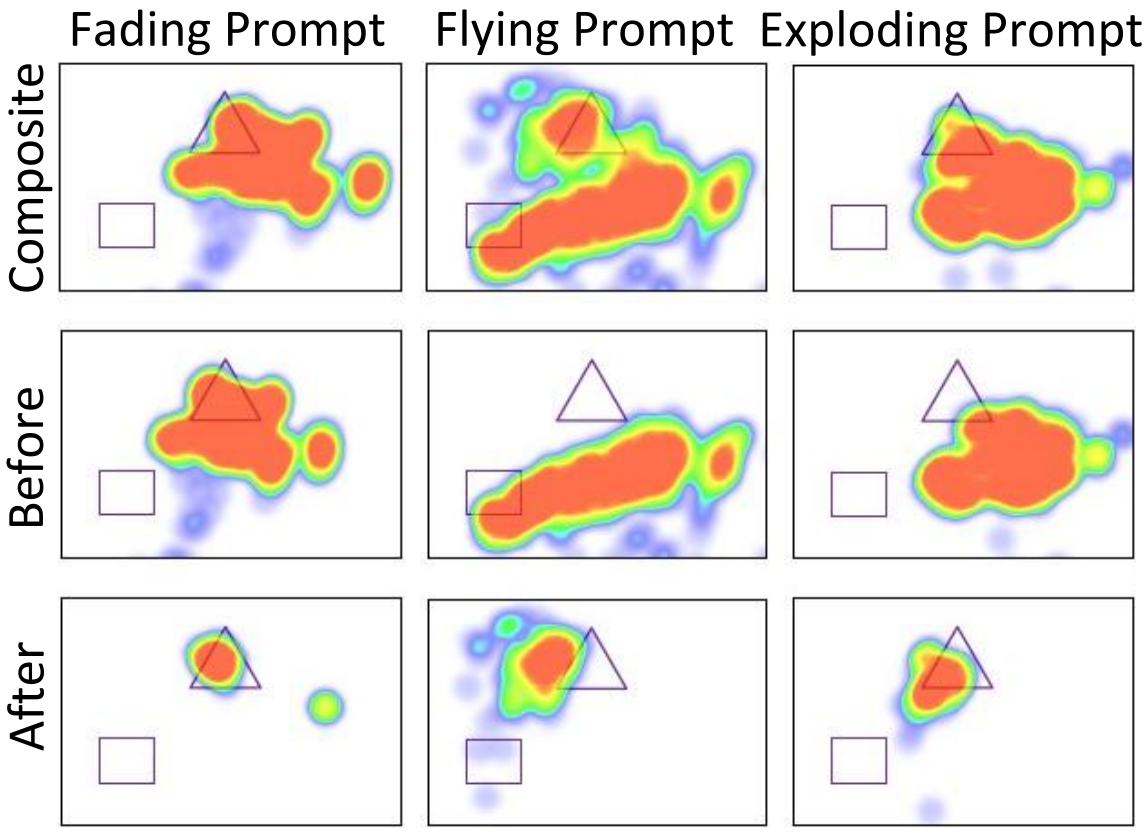



Figure 3.16: Heatmaps of participant 1

the fact that their eyes were tracked in the process of the experiment.

#### 3.2.3.2 Results

We use the first video to test whether the virtual prompt is able to divert the participants' attention from the reference. Respectively four participants start focusing on the prompt 2.2, 0.8, 1.3, 1.5 seconds after the prompt appeared in the video. Heatmaps show the cumulative intensity with which a participant views different locations on the video. As seen in Fig. 3.12, participants' focus is led to the target by the prompt.

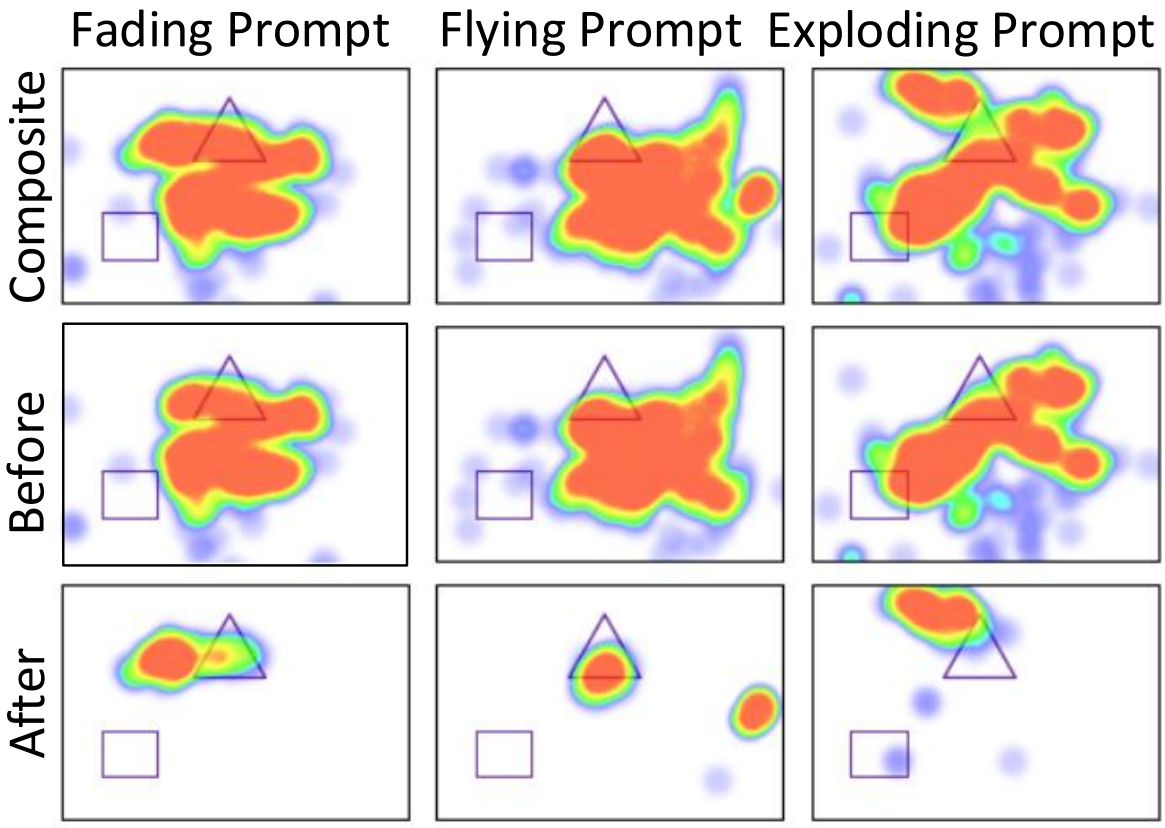



Figure 3.17: Heatmaps of participant 2

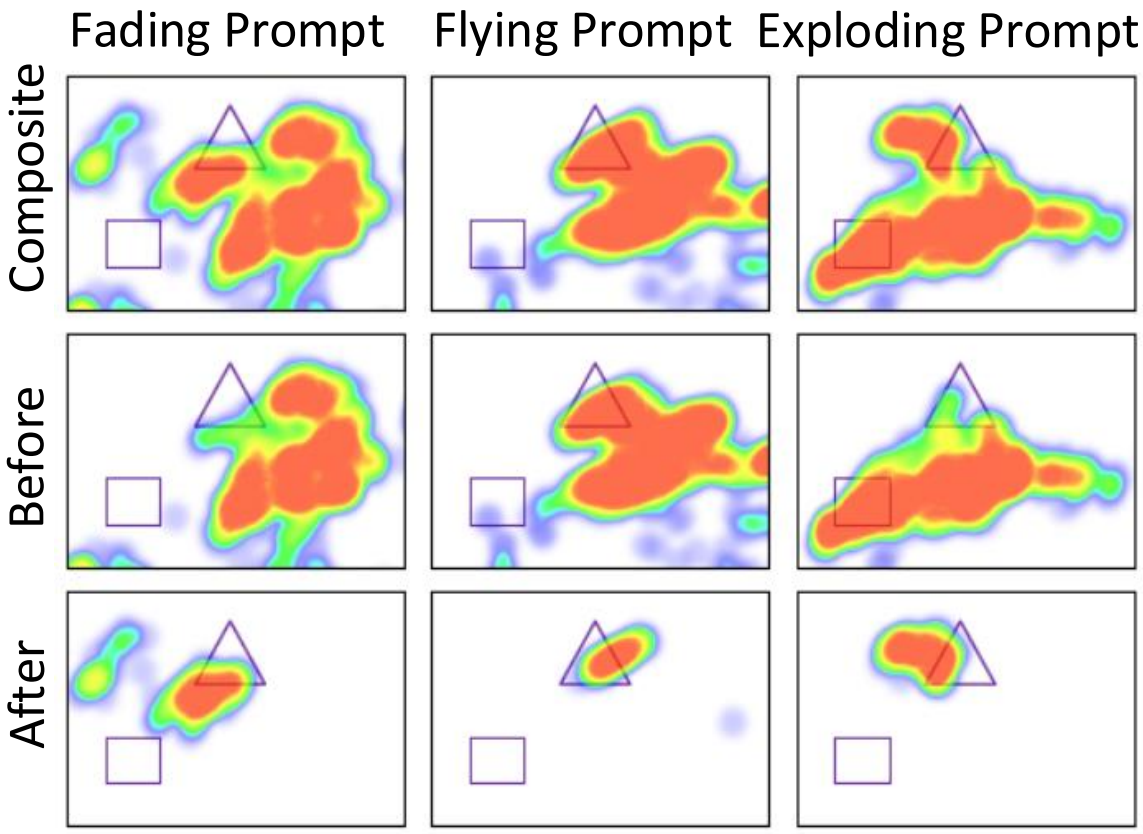



Figure 3.18: Heatmaps of participant 3

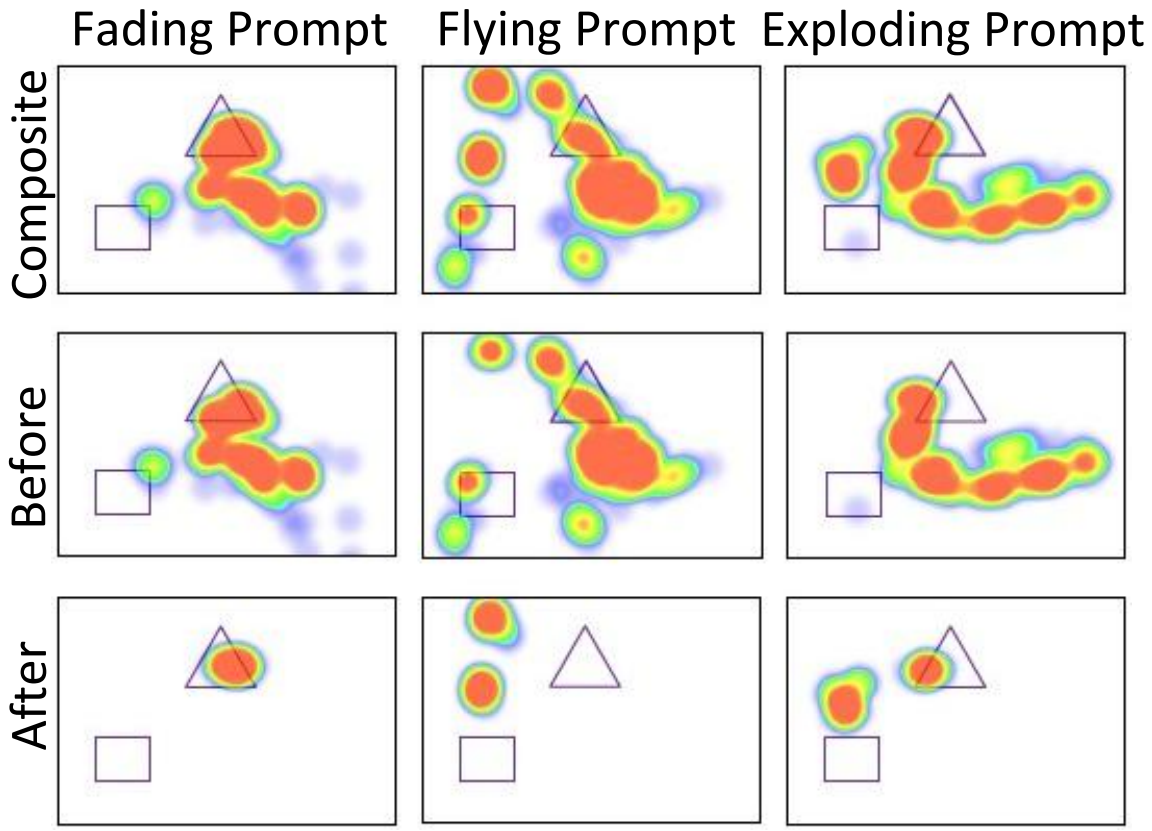



Figure 3.19: Heatmaps of participant 4

We use the other three videos to evaluate how long the prompt can hold participants' eye contact point (gaze) at the target after the prompt begins to disappear. Mean distance between the gaze and target center represents how close a participant's gaze is to the target. In the Fig. 3.13, three participants have smaller mean distances when using fading prompts as compared to others. The mean distance is on average 25.5% and 27.3% smaller than when using flying or exploding prompts. Fig. 3.14 demonstrates how many focal points fall within target. Fading prompts caught 16 and three times larger ratios over flying and exploding prompts. Fig. 3.15 shows the fading prompt is able to keep participants' attention 19 and 20 times longer than the other two.

Fig. 3.16 , 3.17 , 3.18 , and 3.19 depict, after the prompt begins to disappear, the participants' gaze concentrated on the target more when using a fading prompt, and the participants follow flying prompts as they were removed from target. Out of all prompts, fading prompts performs stronger among the three prompt styles.

## 3.3 Selective Speaker Cancellation

Due to hypersensitivity to sound, people with Autism can feel frustrated and even profoundly fearful when talking with multiple speakers. This exacerbates their impairments in social interaction and communication.

In this section, we introduce the third tool, a fully interactive system that allows people with Autism to focus on a single auditory stream (a person's voice) according to their preference during conversations. The tool has the capacity to filter out other speakers' voices based on distinguishing their locations. The experimental results have demonstrated our prototyping system works reliably in regular conversations.

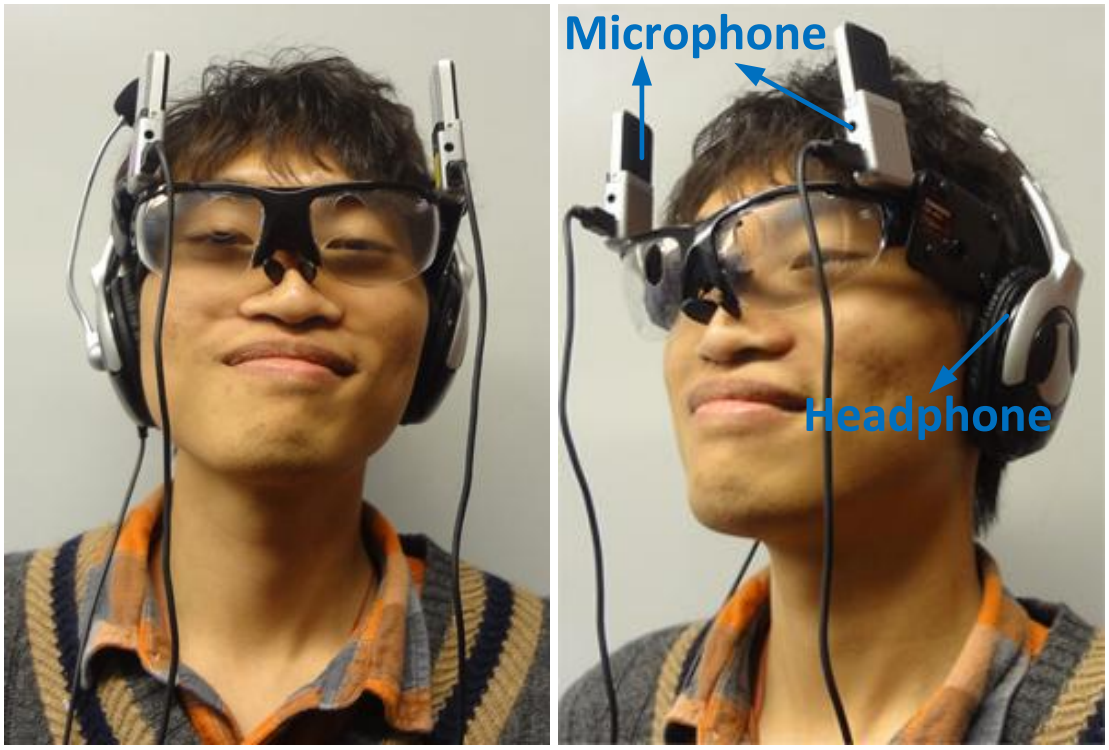



Figure 3.20: Front and side views of a user wearing the headset.

### 3.3.1 Design

As seen in Fig. 3.20, the tool is composed of two microphones mounted on the two sides of wearable glasses and conventional headphones. The audio streams are captured from two microphones from the environment and converted into digital audio data. This converted data is then forwarded to the algorithms running on the portable devices for sound detection and localization the same way as described in section 3.1.1. The headphones are used to muffle extraneous and noise signals and output only the signals pertaining to the desired speaker's voice.

The tool provides a user with a list of all the recognized speakers and lets her/him decide the white and black speaker lists. The user can choose one of three operation modes. In pass-through mode, the user can hear all the sound recorded through

the microphones to the headphones. In the blacklisting mode, all the speakers are initially on. If the user does not like to hear a speaker, she/he would add the speaker into the blacklist. Whenever the tool recognizes that speaker in blacklist is speaking, it mutes the speaker. In the whitelisting mode, all the speakers are muted after the tool generates the list of the recognized speakers. If the user likes to hear a speaker, she/he would add the speaker into the whitelist. Whenever the tool detects that a speaker in the whitelist is speaking, it outputs the voice from the speaker through the earphone. The default mode is whitelisting mode. According to the whitelist or blacklist, the tool can find out whether the user likes to hear the speaker or not, and then it can automatically output or mute the speaker during the whole conversation. In addition, the user can also enhance or muffle the volume of speakers' voices.

In Fig. 3.21, we can see that the left figure depicts the process of speaker cancellation; the right figure illustrates how the tool determines whether the user would like to hear the speaker.

### 3.3.2 Evaluation

Realistically people take turns when they have a conversation in a group, as seen in Fig. 3.22. Therefore canceling a particular voice stream will not affect the others. Fig. 3.23 demonstrates that the tool can recognize unwanted speakers by localizing direction of the sound source, and mute their voice streams.

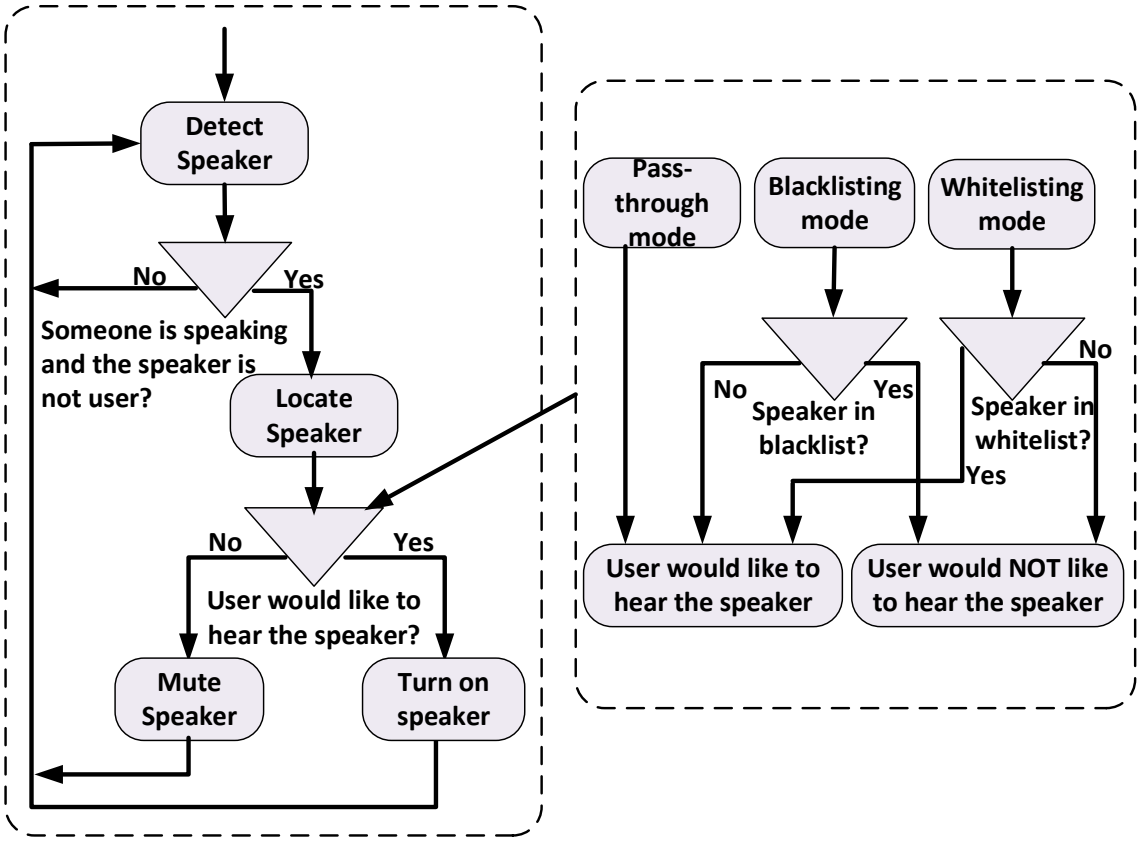



Figure 3.21: The diagram of speaker cancellation tool

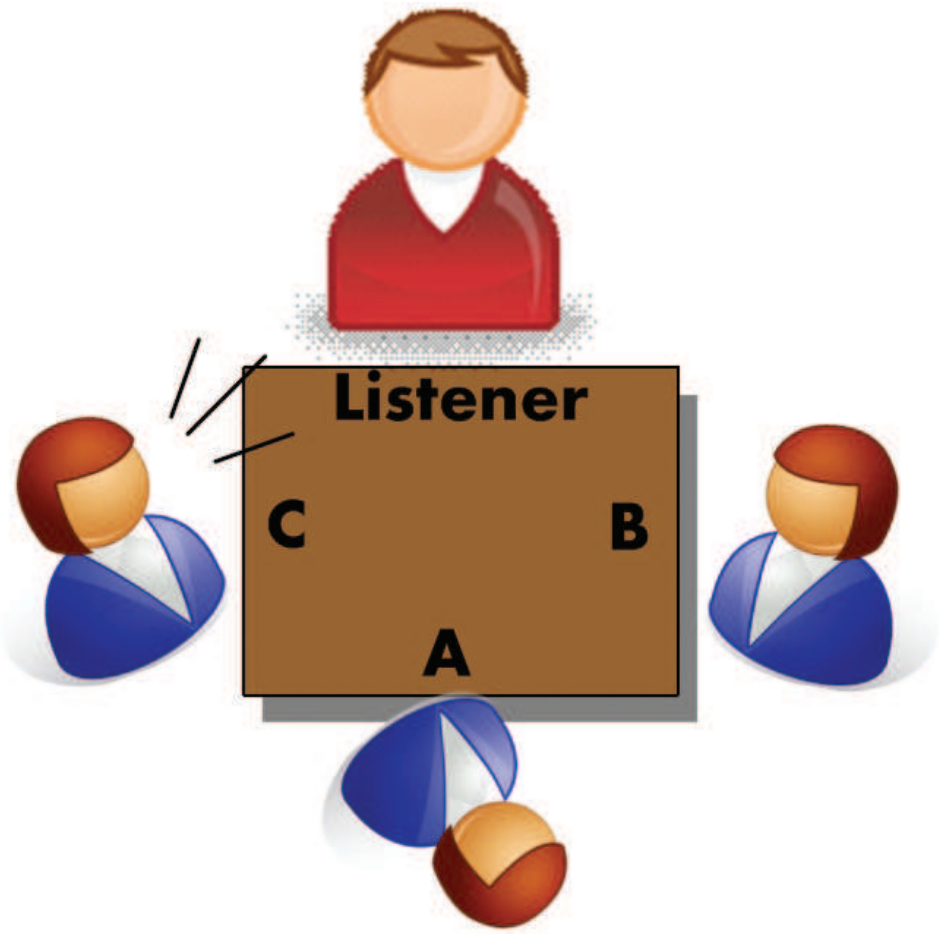



Figure 3.22: In the scenario above, three people A, B, and C speak one after another.

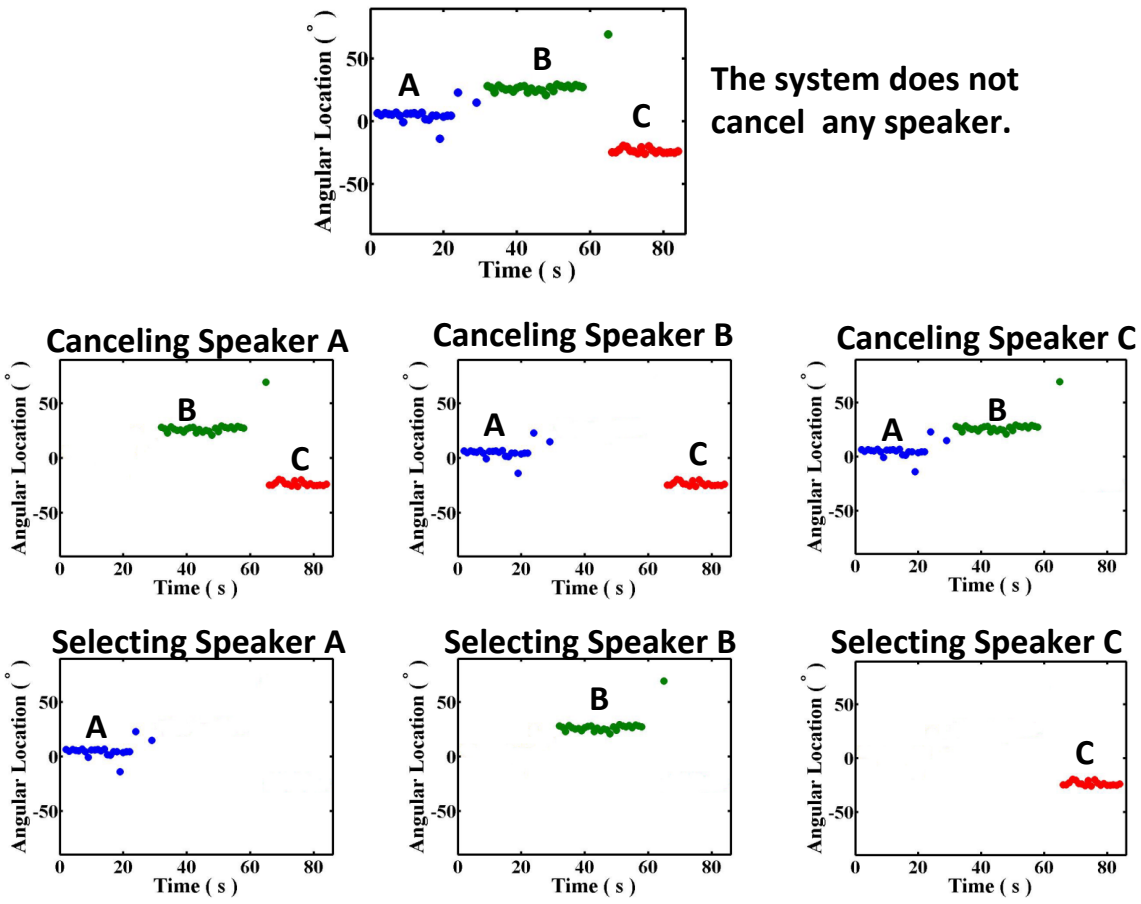



Figure 3.23: Multiple people are present in a conversation

# Chapter 4

# Conclusion

## 4.1 Summary of Contributions

We have presented a set of tools to help social communication for people with autism. The first tool is a directional eye contact reminder tool using wearable computerized-eyewear that displays a prompt to alert the user. It is used as a supplement to traditional therapy. It allows individuals with or without Autism who are likely not to make eye contact to progress easily in their development of social skills in real situations. The second tool adopts VR technology to train children with autism to establish eye contact via a fading prompt approach. We conduct a study to demonstrate that a virtual prompt can draw user's attention to the target and the fading prompt is more effective than the traditional flying prompt and exploding prompt. The third tool performs speaker cancellation. It could be beneficial to social interaction via canceling speakers who cause stress in people with Autism.

## 4.2 Future Work

For the first and third tool, firstly we will investigate sound power more or provide a new method. We will invite more participants to test sound power and design an automatic mechanism to adjust the threshold according to gender, age, and other factors. Secondly, the tools would handle the situation where speakers speak alternately. We will research a mechanism to help people with Autism when multiple people are speaking concurrently. Future work for the second tool includes expanding the current tool so that it is tolerant to scenarios in which multiple people are present in the child's view. The tool would be able to identify the individual the child is interacting with and display a prompt autonomously as is necessary.

# Bibliography


[1] Google glasses. http://en.wikipedia.org/wiki/Google_Glass.

[2] Oculus Rift. http://www.oculusvr.com/.

[3] OpenCV (Open Source Computer Vision). http://www.opencv.org/.

[4] Opengazer. http://www.inference.phy.cam.ac.uk/opengazer/.

[5] OpenGL (Open Graphics Library). http://www.opengl.org/.

[6] Virtual Reality. https://en.wikipedia.org/wiki/Virtual_reality.

[7] Wearable glasses technology. http://www.vuzix.com/augmented-reality/productsstar1200xl.html.

[8] J. Baio. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States. *MMWR Surveillance Summaries*, 63(2):1–21, 2010.

[9] C. Baraniuk. Doctors with iPads could transform hospital care. https://www.newscientist.com/article/mg22229734-700-doctors-with-ipads-could-transform-hospital-care/, June 2014.

[10] P. Bordnick. UH Moment: Unique Virtual Reality Lab Expands, Tackles Heroin Addiction. http://www.uh.edu/news-events/stories/2014/July/07282014Unique%20Virtual%20Reality%20Lab%20Expands,%20Tackles%20Heroin%20Addiction.php, July 2014.

[11] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley, 2009.

[12] Y. Cai, N. K. H. Chia, D. Thalmann, K. N. N. Kee, J. Zheng, and N. M. Thalmann. Design and Development of a Virtual Dolphinarium for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2):208–217, January 2013.

[13] V. J. Carbone, L. OBrien, E. J. Sweeney-Kerwin, and K. M. Albert. Teaching Eye Contact to Children with Autism: A Conceptual Analysis and Single Case Study. *Education and Treatment of Children*, 36(2):139–159, May 2013.

[14] Centers for Disease Control and Prevention. Community Report from the Autism and Developmental Disabilities Monitoring (ADDM) Network. Technical report, National Center on Birth Defects and Developmental Disabilities, 2010.

[15] S. Cherry. Steve Mann: My Augmediated Life. `http://spectrum.ieee.org/podcast/geek-life/profiles/steve-manns-better-version-of-reality`, March 2013.

[16] C. R. Costa and L. Carolina. Findings on sensory deficits in autism: Implications for understanding the disorder. *Psychology & Neuroscience*, 5(2):231–237, July–December 2012.

[17] S. Curtis. YouTube makes all of its videos viewable on Cardboard virtual reality headset. `http://www.telegraph.co.uk/technology/google/11979229/YouTube-makes-all-of-its-videos-viewable-on-Cardboard-VR-headset.html`, November 2015.

[18] K. Dautenhahn. Roles and functions of robots in human society: implications from research in autism therapy. *Robotica*, 21(4):443–452, August 2003.

[19] K. Dautenhahn and I. P. Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *PRAGMATICS & COGNITION*, 12(1):1–35, December 2003.

[20] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, and J. Varley. Randomized, controlled trial of an intervention for toddlers with autism: The early start denver model. *Pediatrics*, 125(1):17–23, January 2010.

[21] M. Elwin, L. Ek, A. Schroder, and L. Kjellin. Autobiographical accounts of sensing in Asperger syndrome and high-functioning autism. *Archives of Psychiatric Nursing*, 26(5):420–429, October 2012.

[22] C. B. Ferster, J. I. Nurnberger, and E. B. Levitt. The control of eating. *Obesity Research & Clinical Practice*, 4(4):401–410, 1962.

[23] S. L. Finkelstein, A. Nickel, T. Barnes, and E. A. Suma. Astrojumper: Designing a virtual reality exergame to motivate children with autism to exercise. In *Virtual Reality Conference (VR)*, pages 267 –268, USA, March 2010.

[24] R. D. Greer and D. E. Ross. *Verbal Behavior Analysis: Inducing and Expanding New Verbal Capabilities in Children with Language Delays*. Pearson, 2007.

[25] P. Haslam and S. Mafeld. Google Glass: Finding True Clinical Value. http://www.whichmedicaldevice.com/news/article/391/google-glass-finding-true-clinical-value, October 2013.

[26] P. Holth. An Operant Analysis of Joint Attention Skills. *Journal of Early and Intensive Behavior Intervention*, 2(3):160–175, September 2005.

[27] L. A. Jeffress. A place theory of sound localization. *Journal of Comparative & Physiological Psychology*, 41(1–2):35–39, February 1948.

[28] M. R. Kandalaft, N. Didehbani, D. C. Krawczyk, T. T. Allen, and S. B. Chapman. Virtual Reality Social Cognition Training for Young Adults with High-Functioning Autism. *Journal of Autism and Developmental Disorders*, 43(1):34–44, 2013.

[29] C. L. Kleinke. Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1):78–100, 1986.

[30] F. L. Kooi and A. Toet. Visual comfort of binocular and 3d displays. *Displays*, 25(2):99–108, 2004.

[31] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. In *IEEE International Workshop on Robots and Human Interactive Communication*, pages 342–346, Japan, 2005.

[32] B. Kuchera. The complete guide to virtual reality in 2016 so far (Update: February 2016). http://www.polygon.com/2016/1/15/10772026/virtual-reality-guide-oculus-google-cardboard-gear-vr, Febulary 2016.

[33] D. L. Lacrama and D. Fera. Virtual reality. *Anale. Computer Science*, 25(1):99–108, 2009.

[34] U. Lahiri, Z. Warren, and N. Sarkar. Design of a Gaze-Sensitive Virtual Social Interactive System for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(4):443–452, May 2011.

[35] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx. Visual discomfort in stereoscopic displays: A review. In *Proceedings of SPIE - The International Society for Optical Engineering*, pages 1–13, Netherlands, January 2007.

[36] S. R. Leekam, C. Nieto, S. J. Libby, L. Wing, and J. Gould. Describing the sensory abnormalities of children and adults with autism. *Journal of Autism and Developmental Disorders*, 37(5):894–910, May 2007.

[37] O. I. Lovaas. *The autistic child: Language development through behavior modification*. John Wiley & Sons Inc, 1977.

[38] O. I. Lovaas. *Teaching Developmentally Disabled Children: The Me Book*. Univ Park Pr, 1981.

[39] O. I. Lovaas, J. P. Berberich, B. F. Perloff, and B. Schaeffer. Acquisition of imitative speech by schizophrenic children. *Science*, 151(3711):705–707, 1966.

[40] O. I. Lovaas, J. P. Berberich, B. F. Perloff, and B. Schaeffer. The establishment of imitation and its use for the development of complex behavior in schizophrenic children. *Behaviour Research and Therapy*, 5(3):171–181, 1967.

[41] J. R. Metz. Conditioning Generalized Imitation in Autistic Children. *Journal of Experimental Child Psychology*, 2(4):389–399, 1965.

[42] National Institute of Mental Health. An Introduction to Autism. Technical report, Psych Central, September 2014.

[43] S. H. K. P, R.Padmanabhan, and H. A. Murthy. Robust voice activity detection using group delay functions. In *IEEE International Workshop on Robots and Human Interactive Communication*, pages 2603–2607, Mumbai, 2006.

[44] L. Rabiner and R. Schafer. *Digital Processing of Speech Signals (Prentice-Hall Series in Signal Processing)*. Prentice Hall, September 1978.

[45] L. R. Rabiner and R. W. Schafer. Introduction to Digital Speech Processing. *Foundations and Trends in Signal Processing*, 1(1):1–194, 2007.

[46] J. Robledo, A. M. Donnellan, and K. Strandt-Conroy. An exploration of sensory and movement differences from the perspective of individuals with autism. *Frontiers in Integrative Neuroscience*, 6(1):1–13, November 2012.

[47] L. Schreibman, R. L. Koegel, and M. S. Craig. Stimulus overselectivity of autistic children in a two-stimulus situation. *Journal of Abnormal Child Psychology*, 5(4):425–436, 1977.

[48] P. Siaperas, H. A. Ring, C. J. McAllister, S. Henderson, A. Barnett, P. Watson, and A. J. Holland. Atypical movement performance and sensory integration in Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 42(5):718–725, May 2012.

[49] C. Sicile-Kira. *Autism Spectrum Disorder: The Complete Guide to Understanding Autism*. TarcherPerigee, January 2014.

[50] P. M. Smeets, G. E. Lancioni, and S. Striefel. Discrimination training through time delay of multistimulus prompts: The shapes and locations of the prompts. *The Psychological Record*, 37(4):507–521, 1987.

[51] S. D. Tomchek and W. Dunn. Sensory Processing in Children With and Without Autism: A Comparative Study Using the Short Sensory Profile. *American Journal of Occupational Therapy*, 61(2):190–200, March–April 2007.

[52] D. Vahdat. Google Glass and Medopad: Rich and Dan tell their story. `http://westminster.impacthub.net/2014/03/31/google-glass-medopad-rich-and-dan-tell-their-story/`, March 2014.

[53] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 511–518, USA, January 2001.

[54] C. W. Walpole and E. M. Roscoe. Reduction of stimulus selectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis*, 40(4):707–712, 2007.

[55] X. Wang, N. Desalvo, Z. Gao, X. Zhao, D. C. Lerman, O. Gnawali, and W. Shi. Eye Contact Conditioning in Autistic Children Using Virtual Reality Technology. In *Proceedings of the 4th International Symposium on Pervasive Computing Paradigms for Mental Health*, pages 79–89, Japan, May 2014.

[56] X. Wang, N. Desalvo, X. Zhao, T. Fang, K. A. Loveland, W. Shi, and O. Gnawali. Eye Contact Reminder System for People with Autism. In *Proceedings of the Sixth International Conference on Mobile Computing, Applications and Services*, pages 160–163, USA, Nov. 2014.

[57] X. Wang, X. Zhao, O. Gnawali, K. Loveland, V. Prakash, and W. Shi. Real-Time Selective Speaker Cancellation System for Relieving Social Anxiety in Autistics. In *Proceedings of the Third International Workshop on Pervasive Computing Paradigms for Mental Health*, Italy, May 2013.

[58] D. M. Zygmont, R. Lazar, W. V. Dube, and W. Mcilvane. Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals: a methodological note. *Journal of the Experimental Analysis of Behavior*, 57(1):109–117, 1992.