# Eye Contact Reminder System for People with Autism

Xi Wang[1], Nicholas Desalvo[1], Xi Zhao[1], Tao Feng[1],
Katherine A. Loveland[2], Weidong Shi[1], Omprakash Gnawali[1]
Department of Computer Science, University of Houston[1]
The Autism Research Laboratory, The University of Texas Health Science Center[2]
Email: {xiwang, desalvo, xizhao1, tfeng3, larryshi, gnawali}@cs.uh.edu[1]
Katherine.A.Loveland@uth.tmc.edu[2]

***Abstract*—Avoiding eye contact behavior has been characteristic of individuals with autism. Such behavior prevents intrinsic development of social and communication skills. In this paper we present a directional eye contact reminder system which reminds people with autism to generally focus their eyes in the direction of a human speaker. This device detects a speaker's voice, calculates the sound direction, and directs their eyes by displaying a prompt on the computerized-eyewear which is a Head-Mounted Display, in the direction of the speaker. The experiments demonstrate the feasibility of our prototyping system.**

***Keywords*—Autism, Eye Contact Reminder, Wearable Device**

## I. Introduction

Autism spectrum disorder (ASD) is a set of developmental disabilities affecting how the brain processes information, causing delays and changes in how a socialization, communication, and overall behavior occurs [1]. The number of people diagnosed with autism has increased dramatically. According to the Centers for Disease Control (CDC), the rate of ASD in the United States has risen to its highest level in recent decades [2]. The CDC reports that about 1 in 68 children has been diagnosed with autism spectrum disorder [3]. ASD can affect children and adults, occurring in all races, ethnicities, and socioeconomic groups.

People identified with ASD 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 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.

A major therapy approach to treating the eye contact deficits in autism is Applied Behavior Analysis (ABA) which requires considerable human intervention. While ABA can be very effective in increasing desired behaviors, individuals with autism frequently have difficulty generalizing those behaviors to everyday situations. As a result, they may not make eye contact when it is called for, if they are stressed, busy or otherwise not attending to the need to do so. An inexpensive and easy-to-use solution for reminding individuals with autism to make eye contact when interacting could help them to improve their eye contact in daily life and lead to better social relating.

We propose 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 alerting the user to look at the speaker. Though traditional therapies have been fundamental in autism treatment, our system can be a novel supplement. The system can remind them to make eye contact and aid in building confidence in social interactions.

We have designed the system framework and provided solutions for each module, as detailed in section 3. We prove the effectiveness of those solutions in section 4. Lastly, we further describe how we will improve the current system design in the future.

## II. Related Work

Although there is no cure for autism, there are many therapies to improve symptoms. One important objective of these therapy programs is to help people with autism adapt to social situations including eye contact when in a conversation or otherwise appropriate social situation [4][5][6]. For example, in [7], individuals with autism were asked to look at a therapists face when verbally prompted (i.e., saying, Look at me). While an ABA therapist may physically prompt patients to attract their attention towards the eyes [8][9], people with autism cannot have a therapist next to them at all times reminding them to make eye contact in a real social settings.

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 [10][11]. For example, in [12], the researchers built a small creature-like robot, Keepon, which was carefully designed to engage children with and without autism in playful interactions. Our system, like other robotic therapies, offers

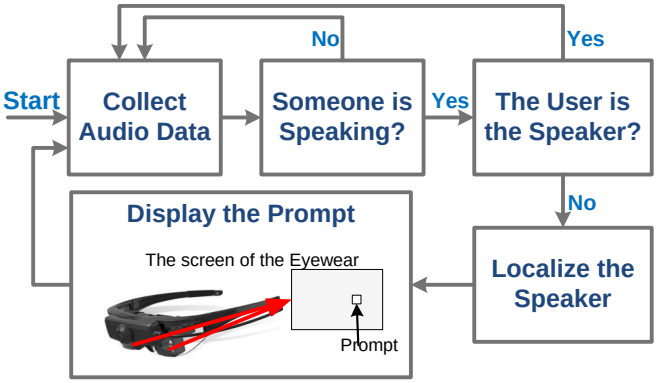



Fig. 1: The diagram of the eye contact reminder system

people with autism the ability to learn in daily life. But there are some advantages our system has that other therapies, including those with robotics lack. One is that the system can be wearable and extremely portable, compared with some other robotic devices. Secondly, our system 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.

## III. System Design

The system consists of two microphones, computerized-eyewear (STAR 1200XL) [13], and a computation unit (a laptop for the prototype). The microphones are mounted onto two sides of the eyewear. They collect audio data and send it to the laptop which has the program to calculate source angle of a 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 technologies. One is voice activity detection (VAD) based on Short Term Energy (STE) and Zero Crossing Rates (ZCR) [14][15]. This splits the signals into overlapping frames [16], 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 voice localization Jeffress Model algorithm. This is a hypothetical model of how neurons in the brain make use of minute time differences. With these two technologies, the system can determine if someone is speaking as well as localize the speaker.

As depicted in Fig.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 linear arrays with the same length. $X_l$ and $X_r$ are 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 collects 40000 samples of sound data X, which is $40000/44100s$ long data.

In the second module, we feed the speech of 40000 samples into the VAD. If no parts of speech are recognized as active, the system 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 system 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 and potentially confusing to focus the users eyes since no other person is speaking to the user at that point.

We use sound power to decide if the speaker is the user. Just like human ears which are able to detect sound intensity and loudness levels, data collected by microphones can also reflect this. The sound power of the source voice decreases exponentially with distance: $E \sim \frac{1}{r^2}$, where $E$ is sound power 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 power 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 power to a threshold and estimate whether the speaker is the user of the system.

Let $T_1, T_2 \cdots T_n$ represent the corresponding sample number of active parts. $T_1, T_2 \cdots T_n$ 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 the power $E$ 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 power 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 power for a whole 40000-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 power: $\bar{E} = \frac{\sum E_j}{\sum T_j} \times 10^4$.

The next module of our system is to localize 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.

Lastly, the prompt is tagged on the transparent screen within the eyewear which alerts the user that someone is speaking and they should make eye contacts in the speaker's direction. The increase in eye contact reinforces visual and auditory coordination and supplements basic building blocks of learning and effective communication.

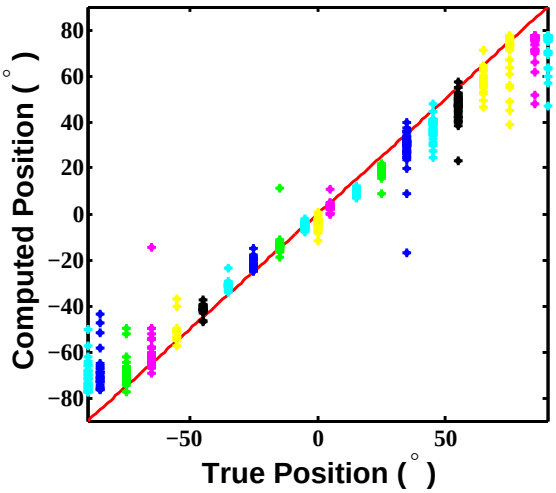



Fig. 2: Speaker Localization

## IV. Evaluation

The system includes two usb microphones (sample frequency is $44100/s$) and the eyewear worn by the user. Both microphones and eyewear are connected to a computer by 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 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 both 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 $0°$, $\pm5°$, $\pm15°$, $\pm25°$, $\pm35°$, $\pm45°$, $\pm55°$, $\pm65°$, $\pm75°$, $\pm85°$ and $\pm90°$ respectively. Each speech instance is around $10s$ long, which indicates $10 \times 44100 = 441000$ samples. In Fig.2, dots are displayed closely around the red diagonal. It indicates that the computed positions are mainly equal or very close to the true positions.

Then, we calculate a proper threshold for sound power, 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 I 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 experiments, low volume is normal in a social conversation. Generally people remain low volume and switch medium volume now and then. So according to Table I, 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 threshold between the user and speaker. This experiment provides a sense of effectiveness of sound power descimination ability in determining the user or speaker. In the future, we will investigate this deeper by recruiting more participants and taking into account their age, gender, and other factors.

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

TABLE I: Average sound power

Lastly, 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 validates that our system can follow and detect different stationary/roaming sound sources in common social settings in real time.

## V. Conclusion and Future Work

We have proposed a novel directional eye contact reminder system using wearable computerized-eyewear that displays a prompt to alert the user. This system used as a supplement to traditional therapy, allows individuals with or without autism who are likely not to make eye contact to easily progress in their development of social skills in real situations.

We plan to work on many improvements of the system. First, 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. Second, the system can handle the situation where speakers speak alternately. We will research a mechanism to help people with autism when multiple people are speaking concurrently.

## References

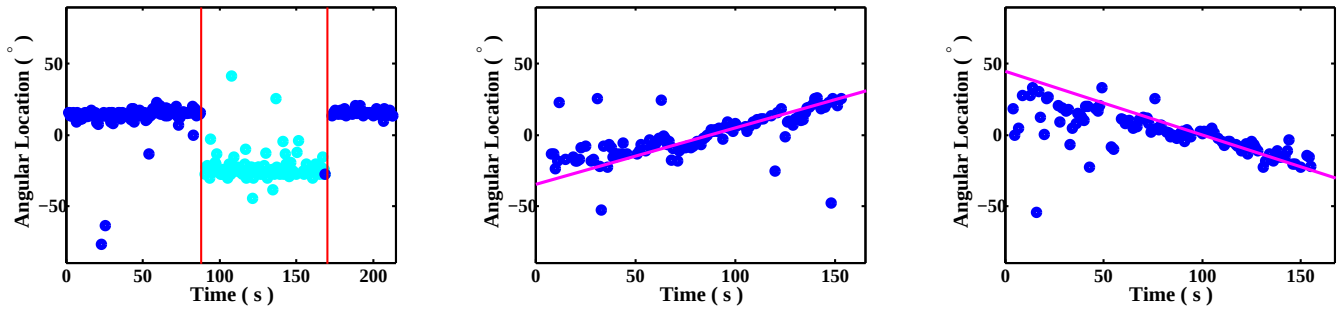
[1] Community Report from the Autism and Developmental Disabilities Monitoring (ADDM) Network, Centers for Disease Control and Prevention, 2010.

[2] An Introduction to Autism. Psych Central. Retrieved on September 12, 2014.

[3] J. Baio: Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, pp.1–21, 2010.

[4] Vincent J. Carbone, Leigh OBrien, Emily J. Sweeney-Kerwin and Kristin M. Albert: Teaching Eye Contact to Children with Autism: A Conceptual Analysis and Single Case Study. Education and Treatment of Children, vol. 36, 2013.

[5] P. Holth: An operant analysis of joint attention skills. Journal of Early and Intensive Behavioral Intervention, 2, pp.160-175, 2005.




(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.

Fig. 3: Three scenarios


[6] P. Holth: Joint attention in behavior analysis. In E. A. Mayville and J. A. Mulick (Eds.), Behavioral Foundations of Effective Autism Treatment. pp. 73-89, Cornwall-on-Hudson, NY: Sloan Publishing, 2011.

[7] R. M. Foxx: Attention training: the use of overcorrection avoidance to increase the eye contact of autistic and retarded children. Journal of Applied Behavior Analysis. pp. 489–499, 1977.

[8] C. N. Macrae, B. M. Hood, A. B. Milne, A. C. Rowe, M. F. Mason: Are you looking at me? Eye gaze and person perception.Psychological Science, pp. 460–464, 2002.

[9] 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, pp. 1–10, 2014.

[10] I. K. Dautenhahn: Towards interactive robots in autism therapy: Background, motivation and challenges, John Benjamins Publishing Company, pp.1–35, 2004.

[11] K. Dautenhahn: Roles and functions of robots in human society: implications from research in autism therapy. Robotica. pp. 443–452, 2003.

[12] H. Kozima, C. Nakagawa, Y. Yasuda:Interactive Robots for Communication-Care: A Case-Study in Autism Therapy. IEEE International Workshop on Robots and Human Interactive Communication, pp.342–346, 2005.

[13] Wearable glasses technology, http://www.vuzix.com/augmented-reality/products star1200xl.html.

[14] S. Hari Krishnan P, R. Padmanabhan, H.A. Murthy: Robust voice activity detection using group delay functions. IEEE International Conference on Industrial Technology, pp. 2603-2607, 2006.

[15] L. R. Rabiner, R. W. Schafer: Introduction to digital speech processing. Foundations and Trends in Signal Processing, pp. 1–2, 2007.

[16] L. R. Rabiner, and R. W. Schafer: Digital Processing of Speech Signals (Prentice-Hall Series in Signal Processing). Prentice Hall, Sept. 1978.