A Real-Time Selective Speaker Cancellation System for Relieving Social Anxiety in Autistics

Xi Wang[1], Xi Zhao[2], Omprakash Gnawali[1], Katherine A. Loveland[3], Varun Prakash[4], Weidong Shi[1]
Department of Computer Science, University of Houston[1,2,4]
The Autism Research Laboratory,The University of Texas Health Science Center[3]
Email: {xiwang,larryshi,gnawali}@cs.uh.edu[1], xzhao21@central.uh.edu[2], vsprakash@uh.edu[4], Katherine.A.Loveland@uth.tmc.edu[3]

***Abstract*—Due to hypersensitivity to sound, patients with autism spectrum disorders (ASD) can feel frustrated and even profoundly fearful when talking with multiple speakers. This exacerbates their impairments in social interaction and communication. We propose a fully interactive system that allows ASD patient to focus on a single auditory stream (a person's voice) according to their preference during conversations. The system 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.**

***Index Terms*—Autism, auditory hypersensitivity, social anxiety, selective speaker cancellation**

I. Introduction

Autism spectrum disorders are a group of developmental disabilities affecting how the brain processes information, causing delays and changes in a person socialization, communication, and overall behavior [1], [2], [3]. The number of people diagnosed with autism has increased dramatically. According to the Centers for Disease Control (CDC), the rate of ASDs in the United States has risen to its highest level in recent decades [4]. The CDC reports that about 1 in 88 children has been diagnosed with an autism spectrum disorder [4]. ASDs can affect children and adults, occurring in all races, ethnicities, and socioeconomic groups.

One of the most commonly reported challenges for individuals with ASD is sensory differences that can make them hypersensitive to stimulation in any or all sensory modalities [5], [6], [7], [8], [9]. Sensory hypersensitivities have been linked to distress and anxiety as well as difficulties with movement [10], [9]. 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 spectrum disorders. 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 [11]. Thus, while conversing in a setting with several people present, a person with ASD 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).

Several strategies have been designed in response to the expressed needs of individuals with ASD who have described their difficulties in conversations where two or more other individuals are present. Unfortunately they cannot properly help autistic people manage auditory sensitivity. The most common strategy is sound isolation. Normally individuals with ASD wear earplugs or sound muffling headphones, or just curtail all social activities and isolate themselves from others. These sound-isolators shield not only unwanted sound but also important speech autistic people should regard. Besides, they stop autistic people from exposure to social environments and then opportunities to practice linguistic and social skills. So in long term this strategy can only exacerbate autistic people's deficits in linguistic communication and social development. Other strategies require autistic people to accommodate commitment to daily therapy programs. Adherence to these structured therapeutic intervention programs requires at least 25 hours per week [12], [13], [14], [15], [16]. For example, Koegel and his coworkers described a systematic treatment using systematic desensitization with several young children with ASDs in [17] and showed some positive outcomes. But their lengthy intensive treatments cost parents and caretakers much time and spending.

In this paper, we describe a novel system to isolate the patient from unattended speakers in the conversation via selectively canceling their voices. The soundproof earplug is able to isolate the sound from the environment while outputting the filtered sound from the system. The system 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 system's user interface. Compared to the aforementioned strategies, our system (i)allows people with ASD 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.

Our main contributions are: (i)explore and analyze the pathology of autism and needs from ASD patients, and propose a portable speaker cancellation system, which could improve life quality in people with ASD; (ii)design a hardware, which functions to capture audio and selectively isolate voice stream; (iii)design an software application based on two speech-processing algorithms.

The rest of this paper is organized as follows: Section 2 discusses the design of the speaker cancellation system, Section 3 details an evaluation of our system, and Section

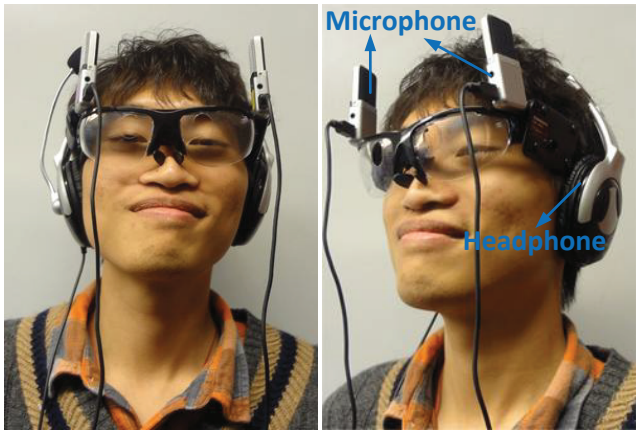


Fig. 1. Front and side views of a user wearing the headset.

4 summarizes our findings and concludes the paper.

II. System Design

In this section, we present our real-time speaker cancellation system.

A. Hardware

The system is composed of two noise canceling 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 speaker detection and localization. The headphones are used to muffle extraneous and noise signals and output only the signals pertaining to the desired speaker's voice.

B. Software

The app recognizes speakers by detecting and localizing them using the core speech processing algorithms. The app provides user with a list of all the recognized speakers and let user decide the white and black speaker lists. A user can choose one of three operation modes. In pass-through mode, user can hear all the sound recorded through the microphones to the headphones. In blacklisting mode, all the speakers are initially turned on. If a user does not like to hear a speaker, he/she would add the speaker into the blacklist. Whenever the app recognizes that speaker in blacklist is speaking, it mutes the speaker. In whitelisting mode, all the speakers are muted after the app generates the list of the recognized speakers. If a user likes to hear a speaker, he/she would add the speaker into the whitelist. Whenever the app 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 app is able to 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.

C. Speaker Detection and Localization Algorithms

we describe two technologies applied in the system: voice activity detector (VAD) [18] utilized to detect speaker, and the Jeffress model [19] utilized to localize the direction of speaker.

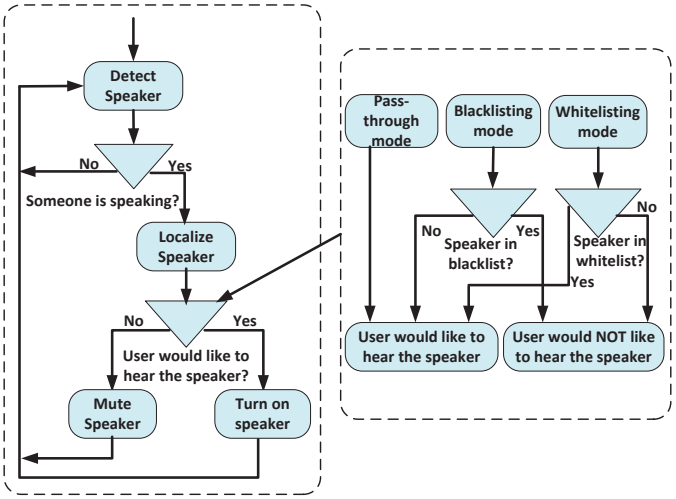


Fig. 2. The left figure depicts the process of speaker cancellation; the right figure depicts how the app determines whether user would like to hear the speaker.

1) Speaker Detection Algorithm: We use VAD algorithm to determine if someone is speaking. The VAD algorithm detects voice by using short term energy (STE) [20] and zero crossing rates (ZCR) [21], which are principal temporal features in speech analysis. VAD splits signals up into overlapping frames [22], extracts features of framed signals, such as STE and ZCR, and compares them to the calculated thresholds to determine the onset and termination of speech boundaries. In this approach, if short speech is found to be non-vocal and the ZCR and STE reach certain limits, we consider it as human speech. VAD can facilitate our system because it can be used to deactivate process during non-speech section of an audio session.

2) Speaker Localization Algorithm: Interaural time differences are by far the easiest technical implementation of sound source localization. The only parameter needed here is the distance between the microphones. We use Jeffress model to localize direction of sound source. Jeffress model is a hypothetical model of how neurons in the brain make use of small time differences to localize sound source. A detector neuron fires if both of its inputs are excited simultaneously. Every detector neuron represents a degree of position. In the example in Fig.3, the sound source is mounted closer to the left ear. Action potentials originating from the auditory nerve closer to the sound source will be able to travel farther along the lower axon than it would take action potentials to arrive from the opposite auditory nerve. Simultaneously the action potentials from both sides are exciting a coincidence detector neuron, which is related to the interaural time difference.

III. System Evaluation

In this section, we report and analyze experimental results of our prototyping system based on the approach detailed in the last section.

A. Dataset

The dataset consists of a set of audio sessions recorded by the two microphone. Sample frequency is $44100/s$. The two microphones are placed $20cm$ apart. A speaker is placed $1m$ away from the midpoint of the two microphones. Sound is

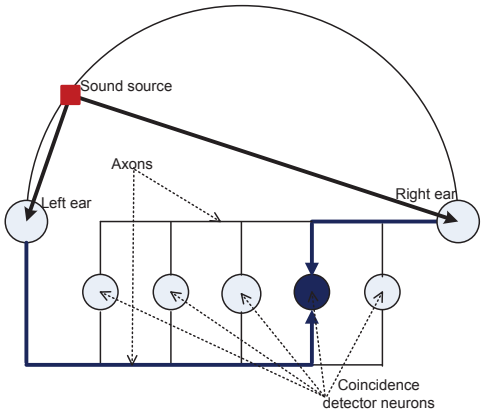


Fig. 3. Jeffress Model

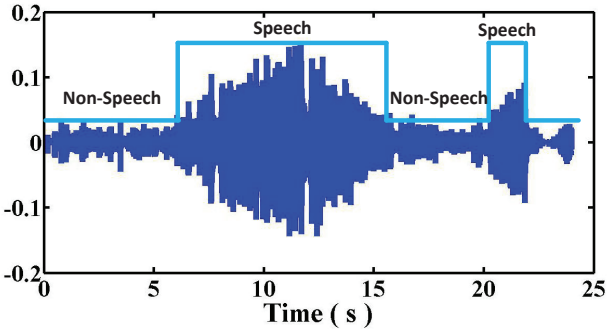


Fig. 4. Accuracy of speaker detection

recorded from speakers placed at $0°$, $\pm5°$, $\pm15°$, $\pm25°$, $\pm35°$, $\pm45°$, $\pm55°$, $\pm65°$, $\pm75°$, $\pm85°$ and $\pm90°$ respectively. Every session lasts from $20s$ to $40s$. To test the system's performance under real-world scenarios, we recorded a conversation containing three speakers, who speak one after another. The conversation lasts $85s$.

B. Experiments and Results

We evaluated the speaker detection performance by checking starting/end points of human speech. Fig.4 illustrates that the VAD algorithm is able to separate the presence and absence of speech in the recordings.

To test the performance of speaker localization, we compared computed position to the true position of the speaker. We let speaker stand at different angular positions: $0°$, $\pm5°$, $\pm15°$, $\pm25°$, $\pm35°$, $\pm45°$, $\pm55°$, $\pm65°$, $\pm75°$, $\pm85°$, $\pm90°$ and recorded sound. We used every 10000 sample to compute position of speaker in Jeffress model, so acquired a set of positional results based on every recording session. In Fig.5, dots are displayed closely around the red diagonal. It indicates that the computed positions are mainly equal or very close to the true positions.

Realistically people take turns when they have conversation in a group, as seen in Fig.6(a). Therefore canceling a particular voice stream will not affect the others. Fig.6(b) demonstrates that the system can recognize unwanted speakers by localizing direction of sound source, and mute their voice streams.

Furthermore, we evaluated the system delay. We call every 10000 audio samples a index, which is $10000 \div 44100 \approx 0.2268s$ long. We record an index of samples to detect and localize speaker. The mean delay for conducting speaker

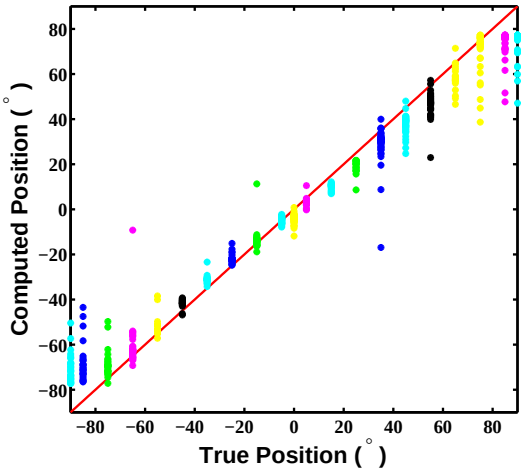


Fig. 5. Accuracy of speaker localization

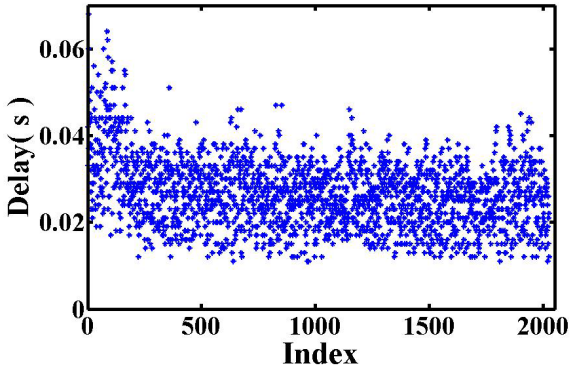


Fig. 7. Speaker detection and localization delay

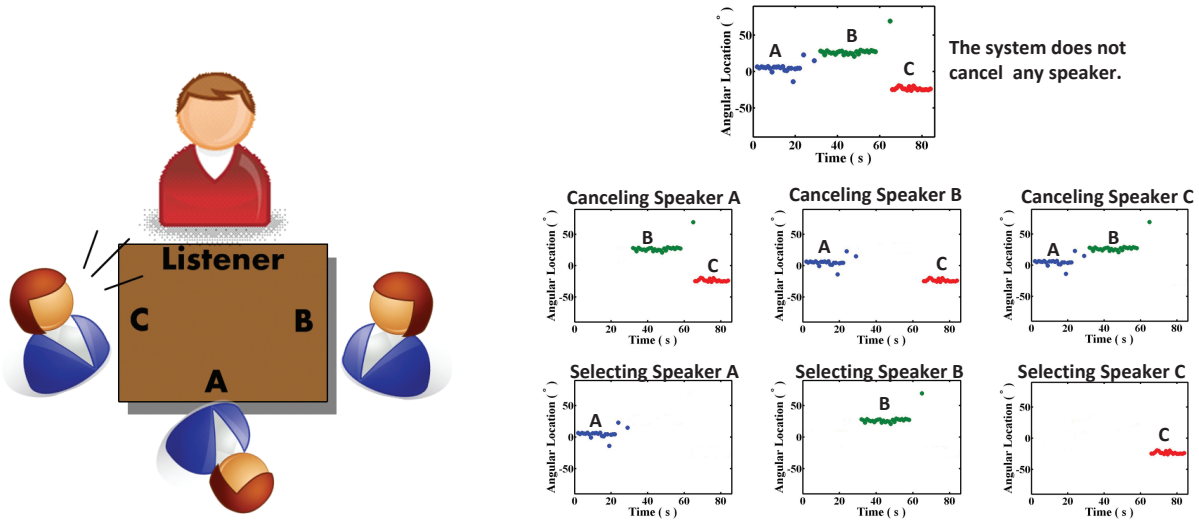
detection and speaker localization is around $0.0317s$, as seen in Fig.7. So the overall delay is around $0.2268 + 0.0317 = 0.258s$. This meets the average reaction time when a person hears a speaker in the real world scenarios [23], [24]. Thereby, our system can operate in real-time and mute undesirable voice before this voice causes discomfort to the ASD patient.

IV. Conclusions

This paper has proposed an assistive system to perform speaker cancellation in real-time for autistic patients. It would be useful for patient's social interaction via canceling speakers who cause stress on the patient. We tested our solution on a set of recording sessions. The experimental results have demonstrated the accuracy of the speaker detection and localization methods and that the proposed system is able to work reliably in social conversation with a few speakers. In the future, we will reduce obtrusiveness of the system by making two microphones built-in. We will also conduct assessments with autistic users.

References

[1] M. Rutter, "Diagnosis and definition of childhood autism," *Journal of autism and childhood schizophrenia*, vol. 8, no. 2, pp. 139–161, June 1978.



(a) In the scenario above, three people A, B, and C speak one after another.

(b) Canceling or selecting speakers.

Fig. 6. System performance when multiple people are present in a conversation

[2] I. Rapin and R. F. Tuchman, "Autism: Definition, neurobiology, screening, diagnosis," *Pediatric Clinics of North America*, vol. 55, no. 5, pp. 1–18, October 2008.

[3] M. Amitai, M. Peskin, D. Gothelf, and G. Zalsman, "Autism spectrum disorders: updates and new definitions," *Harefuah*, vol. 151, no. 3, pp. 167–188, March 2012.

[4] J. Baio, "Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network," *Centers for Disease Control and Prevention*, vol. 61, no. SS03, pp. 1–11, March 2012.

[5] C. R. Costa and L. Carolina, "Findings on sensory deficits in autism: Implications for understanding the disorder," *Psychology & Neuroscience*, vol. 5, no. 2, pp. 231–237, July–December 2012.

[6] M. Elwin, L. Ek, A. Schroder, and L. Kjellin, "Autobiographical accounts of sensing in asperger syndrome and high-functioning autism," *Archives of Psychiatric Nursing*, vol. 26, no. 5, pp. 420–429, October 2012.

[7] 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*, vol. 37, no. 5, pp. 894–910, May 2007.

[8] 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*, vol. 61, no. 2, pp. 190–200, March–April 2007.

[9] 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*, vol. 42, no. 5, pp. 718–725, May 2012.

[10] M. O. Mazurek, R. A. Vasa, L. G. Kalb, S. M. Kanne, D. Rosenberg, A. Keefer, D. S. Murray, B. Freedman, and L. A. Lowery, "Anxiety, sensory over-responsivity, and gastrointestinal problems in children with autism spectrum disorders," *Journal of Abnormal Child Psychology*, vol. 41, no. 1, pp. 165–176, January 2013.

[11] 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*, vol. 6, pp. 1–13, November 2012.

[12] C. Maurice, G. Green, and S. C. Luce, *Behavioral Intervention for Young Children with Autism: A Manual for Parents and Professionals.* Austin,TX: Pro Ed, May 1996.

[13] E. Schopler, N. Yirmiya, C. Shulman, and L. M. Marcus, *The Research Basis for Autism Intervention*. New York: Springer, August 2001.

[14] K. Lawton and C. Kasari, "Teacher-implemented joint attention intervention: Pilot randomized controlled study for preschoolers with autism," *Journal of Consulting and Clinical Psychology*, vol. 80, no. 4, pp. 687–693, 2012 August.

[15] H. E. Dingfelder and D. S. Mandell, "Bridging the research-to-practice gap in autism intervention: An application of diffusion of innovation theory," *Journal of Autism and Developmental Disorders*, vol. 41, no. 5, pp. 597–609, May 2011.

[16] J. Perrin, D. Coury, N. Jones, and C. Lajonchere, "The autism treatment network and autism intervention research network on physical health: Future directions," *Pediatrics*, vol. 130, no. 2, pp. 198–201, November 2012.

[17] R. L. Koegel, D. Openden, and L. K. Koegel, "A systematic desensitization paradigm to treat hypersensitivity to auditory stimuli in children with autism in family contexts," *Research & Practice for Persons with Severe Disabilities*, vol. 29, no. 2, pp. 122–134, June 2004.

[18] J. Kola, C. Espy-Wilson, and T. Pruthi, "Voice activity detection," *MERIT BIEN*, pp. 1–6, August 2011.

[19] L. A. Jeffress, "A place theory of sound localization," *Journal of Comparative & Physiological Psychology*, vol. 41, no. 1–2, pp. 35–39, February 1948.

[20] S. H. K. P, R.Padmanabhan, and H. A. Murthy, "Robust voice activity detection using group delay functions," December 2006, pp. 2603–2607.

[21] L. R. Rabiner and R. W. Schafer, "Introduction to digital speech processing," *Foundations and Trends in Signal Processing*, vol. 1, no. 1–2, pp. 1–194, 2007.

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

[23] J. Shelton and G. P. Kumar, "Comparison between auditory and visual simple reaction times," *Journal of Neuroscience & Medicine*, vol. 1, no. 1, pp. 30–32, September 2010.

[24] J. T. Eckner, J. S. Kutcher, and J. K. Richardson, "Effect of concussion on clinically measured reaction time in 9 NCAA division I collegiate athletes: a preliminary study." *Journal of injury function and rehabilitation*, vol. 3, no. 3, pp. 212–218, March 2011.