Project Proposal

Project Title: Machine Learning in Sound Source Localization for the Hearing Impaired

Author: Alexander C. Sun

Date: October 28th, 2019

Phrase 1:
People with hearing impairments are unable to sense their environment outside of their field of vision, which leaves them in danger of missing crucial sounds that could pose a danger in their everyday lives.

Phrase 2:
The overall aim of this project is to engineer a wearable device that allows the hearing-impaired to gain more awareness by identifying the direction of incoming sound with availability for all, low cost, short latency and minimal invasiveness.

Background:

People with hearing impairments are unable to sense their environment outside of their field of vision which often leaves them with the danger of missing crucial noises that could pose a danger to their lives in public, and creates heavy communication barriers between other people and themselves. Over five percent of the world's population faces some aspect of impaired hearing, and the cost of hearing aids and cochlear implants, the modern-day solutions to this problem, are unaffordable to the wide majority of the hearing impaired.

Currently, around 466 million people suffer from disabling hearing loss, which is around five percent of the total world population. The main impact of disabling hearing loss is the communication barrier. Often they are left isolated from others because they are socially ignored, and because the majority of the population fails to learn sign language which cuts off much of their communication. Feelings of isolation and loneliness surface as even when they are surrounded by others, they are ignored and treated as if they are invisible("Deafness and Hearing Loss," 2019). In families with hearing-impaired children, less than a third of the parents sign regularly, which creates a permanent social divide between even those closest to one (Correll, 2019).

American Sign Language is the main form of communication that the deaf can use. ASL has its strengths, but its viability to help all deaf people is lacking. There are an estimated between 100,000 and one million ASL interpreters in the US, which is far from the amount necessary for every deaf person to have someone available(Correll, 2019). Furthermore, with the presence of an interpreter all the time, a deaf person loses a sense of independence and freedom as all communication has to be relayed through another. In order to communicate with others at all, the interpreter would have to go everywhere with the hearing impaired (Correll, 2019). More often than not, interpreters are not available at all, especially in third world countries. Those without access to this help face not just social barriers, but economic barriers in the future when one needs employment ("How is Deafness Affecting your mental Health," 2016). Lip reading is another method that allows a deaf person to at least, understand what another is saying. Unfortunately, not all deaf people can master this, and even those who cannot always consistently get

accurate meaning (Correll, 2019). Therefore, current communication methods are lacking in function, and accessibility to all of the impaired.

Cochlear implants and hearing aids improve everywhere in terms of size, invasiveness, and effectiveness. On the other hand, their cost increases because of these improvements or stays the same at least. For most, these devices pose economic challenges, as hearing aids can range between $1000 and $4000, and cochlear implants costing around $400,000. The adoption rate of hearing aids is approximately only one third in the US, with the main cause being cited as price(Valente & Amlani, 2017). The production of hearing aids also fails to meet the demand as global production of hearing aids meets less than 10% of global need. Often availability lacks because of an inability to set up fitting appointments, maintenance, and lack of batteries in third world countries or for those in poverty("Deafness and Hearing Loss", 2019). These devices are not accessible to all either. Cochlear implants are far more expensive and require surgical installation in order to function. The cost and invasiveness of the procedure dissuades most deaf people from even considering the option. Furthermore, for older individuals that get the operation, they can still face trouble interpreting the sound and new capabilities that they gain. Corrective devices can also create issues and problems. For example, often they over amplify background noise which can blur the words of others("Impact of hearing loss on daily life and workplace," 1970). The transition is very difficult for many, causing them to remain as a good solution for the very young only("Pros and Cons of Cochlear Implants," n.d).

Need still remains for an affordable device that is available to all people with a hearing disability. Many solutions exist, but all fit situationally and lose their effectiveness in anything but an ideal environment.

Currently, researchers are creating machines that are able to interpret sound and identify its location. Sound source localization(SSL) describes different methods for doing this task in robotics. The most common methods involve microphone arrays consisting of multiple spaced microphones and applying a time difference of arrival algorithm(TDOA) and then applying generalized cross-correlation(GCC) to algorithmically determine the correct direction the sound comes from. Algorithms like MUSIC or PHAT exist to deal and process data from different microphone setups. Many drawbacks come with this as the algorithms are very computationally heavy and require large hardware to process the data otherwise the time delay becomes immense and it loses effectivity (Mandlik, Nemec & Dolecek, 2012). This method lacks viability in benefiting hearing impaired people as implementing this algorithm would require heavy computers to achieve solid response time, which makes any system more invasive.

Project Definition:

The goal of this project is to create a device that allows a hearing-impaired person to be more aware of their environment by identifying the direction of incoming sound within a reasonable delay, by being wearable, lightweight, and minimally invasive and by being cost-effective and available to all hearing-impaired people with minimal restriction.

The device will sense loud incoming sounds, and notify the user of the direction that the sound comes from. This will allow a deaf person to gain more independence and confidence going out in public without fear of missing sounds or being completed isolated in their environment. The device needs to be able to identify the correct direction the sound comes from in order to notify the user accurately of where their attention needs to be brought.

This project will make use of ç neural networks to accomplish sound source localization. Supervised machine learning will be used to train a model by playing sounds from different azimuth angles. If this method is successful it will create a significantly less computationally heavy method of determining the correct sound location, allowing it to be viable for use on an embedded or wearable

system. This method would cover all the criteria as no heavy and expensive computation devices would be necessary for lowering cost and making the device less invasive. The delay would be kept minimal as well. The processing could be done on a raspberry pi and easily fit in the pocket of a backpack.

The device also has to notify the user when it finds the sound source. This will be done through a wearable belt-like strap with vibration motors at a different location. Haptic feedback will indicate the direction the sound comes from allowing the user to react accordingly.

This device is available to all and fits most environments easily. There would be no cost barrier and very little invasiveness with an easy learning curve. Therefore, this method was chosen as the one to be pursued. The device is expected to perform within an accuracy of 45 degrees on the azimuth angle direction that the sound source lays, with a hopefully much higher precision that is possible to be achieved.

Experimental Design/Research Plan Goals:

The convolutional neural network(CNN) marks the center of the project. The entire device is based on upon feeding inputs into it and allowing it to predict the correct location. Most of the testing involves acquiring training data and training the CNN to achieve a high accuracy upon the validation set, and eventually the real testing scenario.

In order to solve this problem, the microphone array needs to be built, and an automatic testing rig needs to be created in order to gather enough training data to accurately determine the sound source location. In order to create the microphone array, microphone and amplifiers will be needed, as well as arduino and Raspberry Pi controllers to process and store the testing data. The holder can be created through CAD and the use of 3D printing. Next a separate wire connected component will be connected that uses vibrating motors that correspond to the predicted sound location. This device will be able to take in ambient sound information, process the information and notify the user using these components and materials. No hazardous materials will be necessary.

The testing rig will be able to rotate the microphone array in reference to the speaker, so that multiple azimuth angles can be tested. The rig will be controlled by an arduino which will drive a stepper motor. This allows the arduino to pinpoint rotate to angles that can then be used to create the data set. The testing rig will then be setup in my house in order to collect data. Furthermore, this process will be automated so that thousands of data points may be recorded without needing my presence to be available.

Next, the training set created by the testing rig, will then be used for training of the convolutional neural network model. The model will be trained using Tensorflow Keras, and the finished model's accuracy will be tested through the creation of a validation set. The testing set will be a division of the original training set. Each tested model will improve and once the training is sufficient the model will be flashed onto the Raspberry Pi. Then the device will be tested using the testing rig and the accuracy will be measured over the test set with real life reverberation and conditions.

The experiment consists of moving the azimuth angle between the device and the speaker (Independent Variable) and reading the output of the created device (Dependent Variable). The testing rig will screw onto the bottom of the microphone array and will be attached to the ground by tape. The speaker will then be placed three meters away from the center of the microphone array. The testing rig arduino controller will rotate the stepper motor a certain amount of degrees and send via serial to the controller on the microphone array. The testing rig controller then makes the speaker play a sound at a set frequency, and the microphone array controller records the sound as recorded from the five microphones. The information from the five microphones is used as the input for the convolutional neural network. The model processes the inputs and predicts the position the sound is in reference to the microphone array. The prediction and the actual position is recorded and then analysed to conclude on the accuracy of the device. The accuracy will then lead to revision and different ideas for improvement. First trials and

iterations will only classify in which eighth the sound source is located in, while later trials may aim to increase accuracy of the model.

Timeline:

Three main phases define this project. The first phase involves preparation and setup of the testing environment and the testing materials. This involves 3D printing and assembling the actual devices, writing the data collection software, and setting them up properly in the testing space. This is set to span the course of two to three weeks.

The second phase marks the data collection, training, and testing of the device. This phase means running the training set data collection, and training the machine learning model. It also involves the testing of the devices accuracy to meet the standards set above in the experimental design. This is the majority of the project and is set to span for around six to eight weeks.

The third phase marks the revision, improvement and finalization portion. The model will be improved to obtain higher accuracy than the minimal viable portion. The vibration output devices will be added to the device as a method of notifying the user. This marks the completion of all that is planned for the project, as well as multiple layers of revision before the finish. Data analysis and statistics will be calculated and used as a baseline for improving the device. This phase should run for around three to four weeks.

Potential Roadblocks:

The largest potential roadblock is obtaining and training the data for the neural network. The use of a custom microphone array means that all input training data for the machine learning algorithm will have to be obtained through testing and experimenting with the device. Proper setup of the device in a room and consistently taking readings of sound at the same noise level from the same distance will be time-consuming and difficult to do accurately. Obtaining the training set could take up to 40 hours because over a thousand data points if not more is necessary to train a strong machine learning model. Then, training the model itself once the data is acquired is also a potential roadblock. One knowledge gap is the use of tensor flow to train models, and it is a difficult tool to learn. Furthermore, models will have to be trained and retrained to find the best and most accurate one which could take more than months to do. Both of these are the largest potential roadblocks as they are time-consuming and difficult to accomplish well.

References

Correll, R. (2019, June 30). Challenges That Still Exist for the Deaf Community. Retrieved October 21, 2019, from https://www.verywellhealth.com/what-challenges-still-exist-for-the-deaf-community-4153447.

Deafness and hearing loss. (2019, March 20). Retrieved October 21, 2019, from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

How is Deafness Affecting Your Mental Health? (2016, May 18). Retrieved October 21, 2019, from https://deafunity.org/article-interview/deafness-and-mental-health/

Mandlik, M., Nemec, Z., & Dolecek, R. (2012). Real-Time sound source localization. Retrieved October 21, 2019, from https://ieeexplore.ieee.org/document/6233370

National Research Council (US) Committee on Disability Determination for Individuals with Hearing Impairments. (1970, January 1). Impact of Hearing Loss on Daily Life and the Workplace. Retrieved October 21, 2019, from https://www.ncbi.nlm.nih.gov/books/NBK207836/.

Pros and Cons of Cochlear Implants. (n.d.). Retrieved October 21, 2019, from

https://www.babyhearing.org/devices/cochlear-implant-pros-cons.
Valente, M. (2017, July 1). Cost as a Barrier for Hearing Aid Adoption. Retrieved October 21, 2019, from https://jamanetwork.com/journals/jamaotolaryngology/fullarticle/2627924

Background Knowledge Goals:

Date	Topic	Completed Date
10/8/19	Neural network structure and function	10/20
10/8/19	Training Convolutional Neural Networks	10/27
10/16/19	Arduino with Microphones	10/15

Want to use	Why?	Assumptions made with this idea	How can these assumptions be changed
Spatial detection models on-chip (miniaturization) models on the cloud	It provides a possible method for locating the sound. If the models could be properly then this idea is feasible given better input.	Model inputs would match outputs that I can provide for the computer.	Research it and browse through current models to see what they require.
Microphone array with ML model interpretations	This is the current plan. It's the most precise option and is not based on visual input which can be misleading. It is also quick and lightweight so its the best option for the user.	The processing power can be turned down, and that I can train a ML model that is actually able to interpret properly.	Start learning to train ML models.

Don't want to use	Why?
Emotion detection ML models	It doesn't relate to the idea.
Risk of big brother	It's a warning not an idea.
Echolocation	Ultrasonic sensors already use this, and has no correlation to my actual idea.