Project Notes:

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

Note Well: There are NO SHORT-cuts to reading journal articles and taking notes from them. Comprehension is paramount. You will most likely need to read it several times so set aside enough time in your schedule.

Contents:

Knowledge Gaps:

This list provides a brief overview of the major knowledge gaps for this project, how they were resolved and where to find the information.

Knowledge Gap	Resolved By	Information is located	Date resolved
Neural network structure and function			
Training Convolutional Neural Networks			
Arduino with Microphones			

Literature Search Parameters:

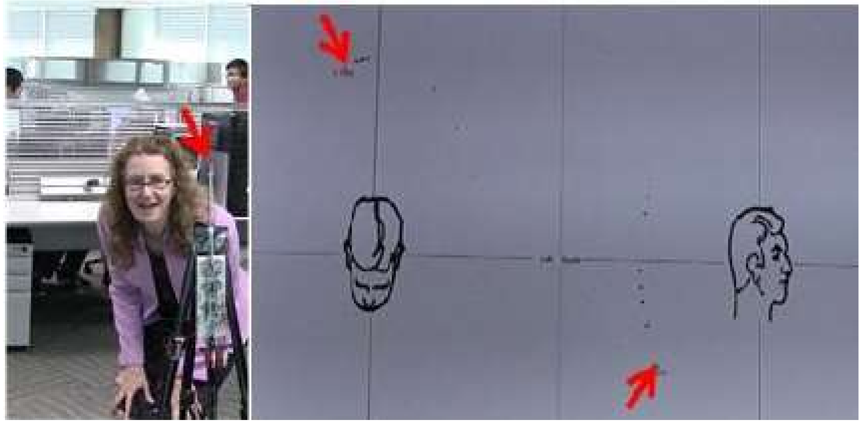
These searches were performed between (Start Date of reading) and XX/XX/2019.
List of keywords and databases used during this project.

Database/search engine	Keywords	Summary of search
IEEE	Sound source Localization, Real Time,	Received a lot of information and different articles of different implementations of sound source localization. Common were algorithmic implementations like MUSIC or PHAT.
IEEE	Neural Network sound source localization	Learned of Neural Network approaches to sound localization and certain benefits computationally.
Google	Microphone Array Patent	Saw patents and the intention behind different designs of microphone arrays.

Article #1 Notes: A Real-Time 3D Sound Localization System with Miniature Microphone Array for Virtual Reality

Article notes should be on separate sheets

Source Title	A Real-Time 3D Sound Localization System with Miniature Microphone Array for Virtual Reality
Source Author	Shengkui Zhao, Saima Ahmed, Yun Liang, Kyle Rupnow, Deming Chen, and Douglas Jones
Source citation	Zhao, Shengkui., Ahmed, Saima., Liang, Yun., Rupnow, K., Chen, D., Jones, D. 2012. A Real-Time 3D Sound Localization System with Miniature Microphone Array for Virtual Reality. Retrieved from https://www.researchgate.net/publication/258423607_A_Real-Time_3D_Sound_Localization_System_with_Miniature_Microphone_Array_for_Virtual_Reality
Original URL	https://www.researchgate.net/publication/258423607_A_Real-Time_3D_Sound_Localization_System_with_Miniature_Microphone_Array_for_Virtual_Reality
Source type	Online Publication
Keywords	3D Sound Localization
Summary of key points	Purpose of paper is to show how they used a combination of omnidirectional and bidirectional microphones to achieve the smallest microphone array for the purpose of sound location.
Important Figures	

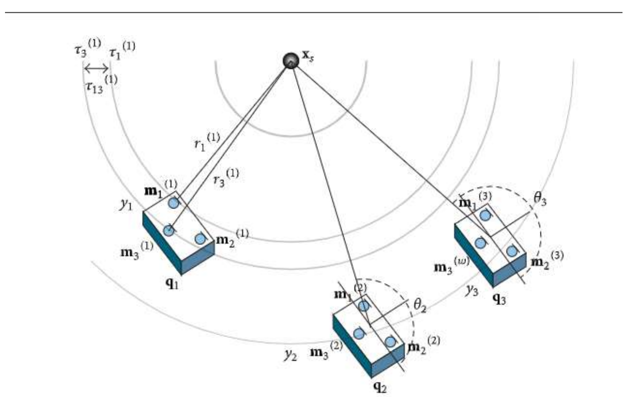
	
Reason for interest	3D sound localization
Notes	● Current microphone arrays large ● used in many applications like virtual reality and human-computer interactions ● used 4 collocated microphones vs. spaced out microphones ● used 3 bidirectional mics, one in each x,y,z plane, and used one omnidirectional ● omnidirectional finds the direction of sound with equal magnitudes ● Use amplitude differences vs. time differences ● used mainly pressure gradient microphones vs normal mics ● Multiple Signal Classification algorithm ● Successful with main advantage of being useful based on its size
Follow up Questions	How does the MUSIC algorithm work? How did they collect enough data to be able to create an algorithm that will do this accurately? What are the advantages vs. disadvantages of time vs. amplitude of arrival?

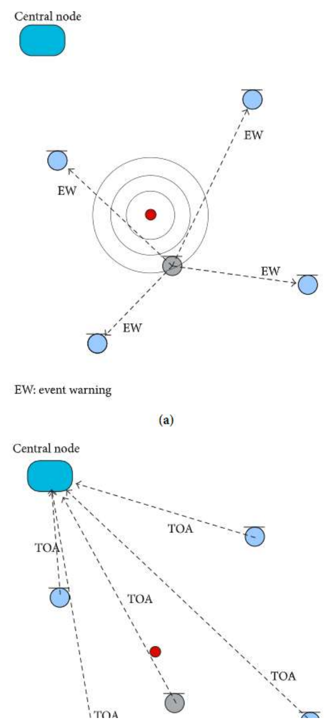
Abstract Summary:
3D sound localization systems are generally created using only omnidirectional microphones, but the researchers use a combination of omnidirectional and bidirectional microphones which allows the array's size to be shrunk. They claim to have created the world's smallest microphone array arrangement for this purpose. Furthermore, they demonstrate a real-time demonstration of the technology and its applications in virtual reality.

Introductions of articles were read and notes were written in the notes section and figures added.

Article #2 Notes: A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks

Article notes should be on separate sheets

Source Title	A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks
Source Author	Maximo Cobos Fabio Antonacci Anastasios Alexandridis Athanasios Mouchtaris and Bowon Lee
Source citation	Cobos, Maximo, Antonacci, Fabio, Alexandridis, Anastasios, … Bowon. (2017, August 17). A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks. Retrieved from https://www.hindawi.com/journals/wcmc/2017/3956282/
Original URL	https://www.hindawi.com/journals/wcmc/2017/3956282/
Source type	Online Publication
Keywords	Sound Source Location
Summary of key points	Evaluates different methods of sound source location algorithms like DOA, TDOA and SRP.
Important Figures	 **Figure 1:** WASN with $M = 3$ nodes and $N = 3$ microphones per node.

<table>
<tr><td></td><td>
</td></tr>
<tr><td>Reason for interest</td><td>Sound Source Location</td></tr>
<tr><td>Notes</td><td>

Most common methods are RSS received signal strength, and TOA or time of arrival
RSS has issues with channel fade and interference
wireless systems provide many advantages as microphone distance and spacing adds more accuracy
Modern methods

energy readings
time-of-arrival (TOA) measurements,
time-difference-of-arrival (TDOA) measurements
direction-of-arrival estimates
by utilizing the steered response power (SRP) function.

DOA systems-

use multiple microphones that each take a DOA reading, which get sent to a processing unit
with multiple sources gets messy
nodes can miss readings which de-sync entire system as well as false alarms

TDOA systems

related to time of flight of wave
needs multiple nodes of sensors which all take reading
then data is processed by generalized cross correlation

</td></tr>
</table>

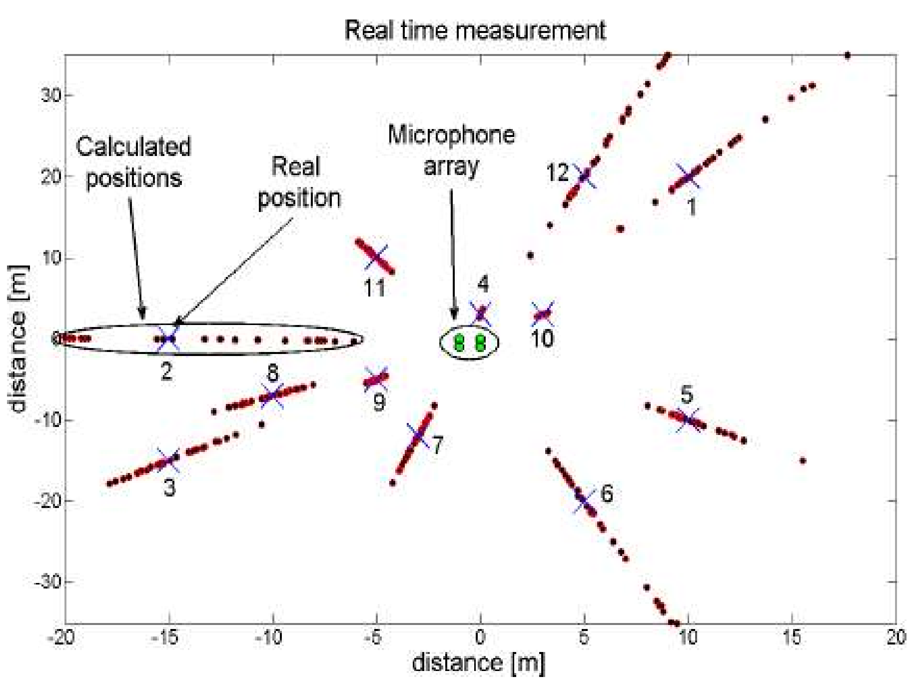
	◦ creates parabola w/ mic as vertex and source laying on the branch ◦ multiple sensors are able to look for intersections ◦ often prone to errors • SRP methods ◦ beamforming technique ◦ compute the output power of a filter-and-sum beamformer steered to a set of possible source locations defined by a grid ◦ point w/ highest value is the estimated location • Paper structured to discuss each process and the math behind it. It evaluates methods of creation and implementation and its advantages and disadvantages.
Follow up Questions	

Abstract Summary:
Currently, acoustic sensor systems with different arrangements form heavy interest from the research community with the specific application to the location of the sound source. Generally, these results are found by looking at the differences in location between microphones. This article aims to evaluate current methods of achieving this location. They discuss time of arrival difference, time difference of arrival, and steered response power. Then they end with the current obstacles and innovations in the field.

Introductions of articles were read and notes were written in the notes section and figures added.

Article #3 Notes: Real-Time sound source localization

Article notes should be on separate sheets

Source Title	Real-Time sound source localization
Source Author	Mandlik, M., Nemec, Z., & Dolecek, R
Source citation	Mandlik, M., Nemec, Z., & Dolecek, R. (2012). Real-Time sound source localization. Retrieved from https://ieeexplore.ieee.org/document/6233370
Original URL	https://ieeexplore.ieee.org/document/6233370
Source type	Online article
Keywords	real time sound source localization
Summary of key points	Explains the math behind TDOA methods, and the advantages and disadvantages to certain arrangements of microphones. They used a square arrangements which did well for their testing.
Important Figures	
Reason for interest	More on setting up microphone array for sound source localization
Notes	Microphone arrays require three or more microphones in a square,

	line or star arrangement. The microphones have to be at a known location and when sound is received they are sent to a central hub for processing and time comparison. More microphones means more accuracy in measurements. All microphones must near identical in how they receive and amplify sound. In an indoors settings the microphones receive reverb from sound bouncing off walls and other objects. Capacitor microphones and linear amplifiers were used.
Follow up Questions	The articles describes frames that are used to divide the incoming signals and I don't quite understand the meaning and significance of this. Are the frames what allows the computer to measure the time differences of arrival?

3 Terms/concepts that I need to investigate further:

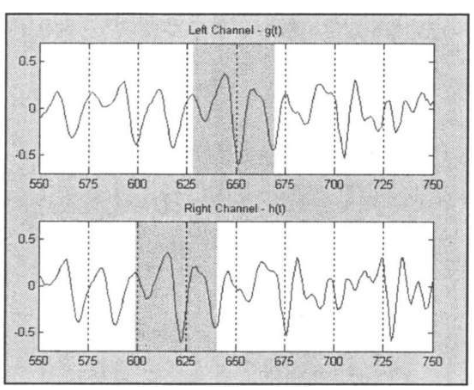
Microphones arrangements in the array

Processing the information without the need for high computational power

How to process and make sense of the data received from the microphones.

Article #4 Notes: A recurrent neural network for sound-source motion tracking and prediction

Article notes should be on separate sheets

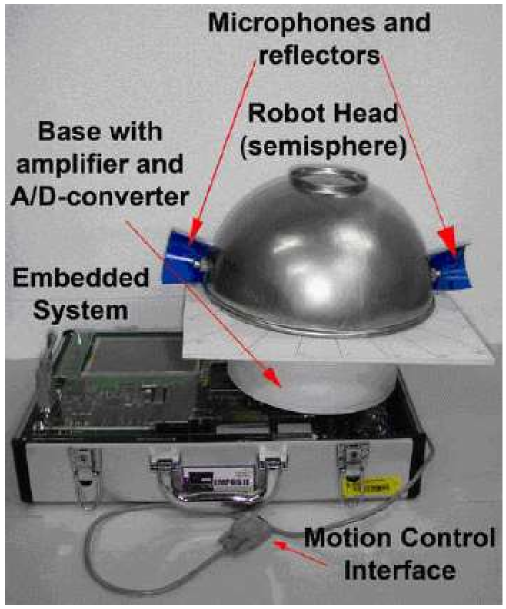
Source Title	A recurrent neural network for sound-source motion tracking and prediction
Source Author	J. C. Murray, H. Erwin and S. Wermter
Source citation	J. C. Murray, H. Erwin and S. Wermter, "A recurrent neural network for sound-source motion tracking and prediction," *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, Montreal, Que., 2005, pp. 2232-2236 vol. 4. doi: 10.1109/IJCNN.2005.1556248. Retrieved from https://ieeexplore.ieee.org/document/1556248
Original URL	https://ieeexplore.ieee.org/document/1556248
Source type	Conference Paper
Keywords	Sound source + neural network
Summary of key points	Created a motion detection and sound locating system. Used a RNN to then predict where the sound would come from next
Important Figures	
Reason for interest	Accomplishes a very similar task to what I hope to accomplish but

	with limitations.
Notes	● Use cross correlation to find the azimuth angle of the sound ○ look at the two sound waves from the left and right mic ○ compare them and look for the same signals except with time delay between them ○ If it finds one the time difference is calculated and used to find the azimuth angle ● RNN uses back propagation weight adjusting ○ learns temporal movements ○ keeps a short term memory in a sense ○ looks for patterns in movements and tries to predict them ○ Requires two angles to be inputted for the prediction ○ waits for two angle readings and can determine the third based on the difference in movement ○ short memory done by only looking at recent data ○ only responds after two consecutive movements, can’t get biased by older data ● 2 simulation environments were used to test the device ○ one with randomly generated points ○ one with manually generated points ● device functioned for points moving up to 34 degrees per second and was rather successful
Follow up Questions	Does this experiment assume a constant or fluctuating speed for the sound source? Would that have an effect on the RNN’s ability to predict the correct point.

Article #5 Notes: Sound Source Localization system based on neural network for mobile robots

Article notes should be on separate sheets

Source Title	Sound Source Localization system based on neural network for mobile robots
Source Author	Yang Geng, Jongdae Jung and Donggug Seol
Source citation	Yang Geng, Jongdae Jung and Donggug Seol, "Sound-source localization system based on neural network for mobile robots," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, 2008, pp. 3126-3130.doi: 10.1109/IJCNN.2008.4634240. Retrieved from https://ieeexplore.ieee.org/document/4634240
Original URL	https://ieeexplore.ieee.org/document/4634240
Source type	Conference Paper
Keywords	Sound Source Localization + Neural Network
Summary of key points	Sound source localization methods take a long time to compute, and the use of neural network to infer positions poses a much more efficient solution to the problem.

Important Figures	
Reason for interest	Uses an interaural system similar to human detection of sound location. Tests the validity of only using 2 mics in comparison to an array of 3 or more.

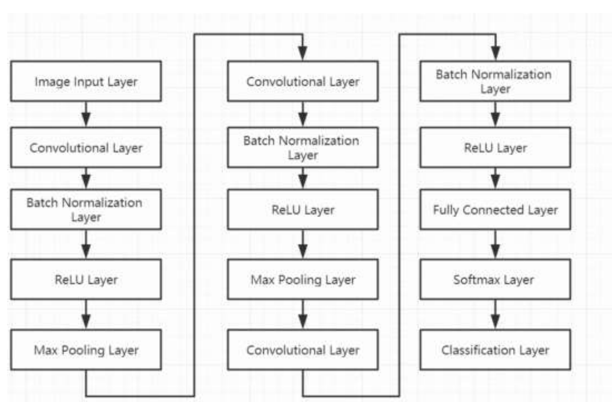
Notes	● 2 microphones are all that's necessary
	    ○ naturally humans only have 2 ears
	    ○ said that vision impaired people can tell the location of sound
	● Cross correlation with interaural time difference
	    ○ caused too high of error up to 5 degrees
	● Used supervised machine learning to accomplish task
	    ○ Large amount of data available
	    ○ Input data is very difficult to computationally compute into correct output data
	    ○ NN solution avoids this as the computer only needs to be taught the correct solution
	    ○ took sound amplitude over time as input for the model
	    ○ model used gradient descent to improve
	● Data was collected by playing sounds from a fixed distance of 1 meter away,
	    ○ sounds were played from various angles and the input of the right mic and left mic was recorded over the three seconds
	● Error was minimal with no inference more than one degree off.

Follow up Questions	How was the neural network designed?

	What type of neural network is this? How much data was collected to train the model?

Article #6 Notes: Detection Sound Source Direction in 3D Space Using Convolutional Neural Networks

Article notes should be on separate sheets

Source Title	Detection Sound Source Direction in 3D Space Using Convolutional Neural Networks
Source Author	X. Yue, G. Qu, B. Liu and A. Liu
Source citation	X. Yue, G. Qu, B. Liu and A. Liu, "Detection Sound Source Direction in 3D Space Using Convolutional Neural Networks," *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, Laguna Hills, CA, USA, 2018, pp. 81-84. doi: 10.1109/AI4I.2018.8665693. Retrieved from https://ieeexplore.ieee.org/document/8665693
Original URL	https://ieeexplore.ieee.org/document/8665693
Source type	Conference Paper
Keywords	Sound Source Localization + Neural Network
Summary of key points	The use of a convolutional neural network could create a "filter" for input signals and provide the output signal with minimal computation. It can be applied to almost any other same array and provides a strong deployable solution.
Important Figures	

<table>
<tr><td>Reason for interest</td><td>The use of a neural network applied to this problem</td></tr>
<tr><td>Notes</td><td>
TDOA algorithms struggle with noisiness and reverberation
GCC-PHAT algorithm for detecting time delays between different mics
Adapted to function in 3D space
means mics have to arranged symmetrically in all three dimensions
means the distance between the center and the mics are all equal

Create a cube around the microphones, and distribute 100 points across its surface
Each point represents a set of data that can be classified as that point for its direction/angle

GCC-PHAT is calculated between each pair of microphones and used as the inputs to the CNN
Image input layer
accepts matrix of inputs

Convolutional layer
does sizes and amount of filters done on the data

Batch Normalization Layer
normalizes the gradients to make computation easier

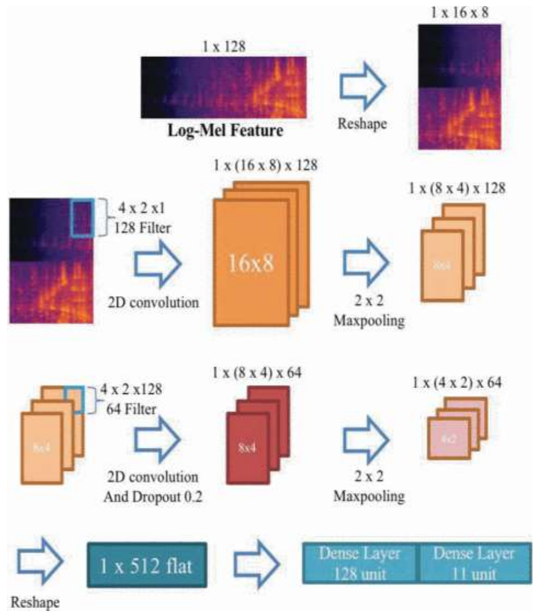
Classification Layer
final layer
generates output and loss based on probabilities

Simulation done in a 4m by 4m by 4m room
each mic located .1 meters from center
Maximum time delay was calculated between pairs of microphones
50 directions of each surface was tested
</td></tr>
<tr><td>Follow up Questions</td><td>Why was the cube method used over a spherical one?
How large were the training sets?
How long is the processing time?</td></tr>
</table>

Article #7 Notes:

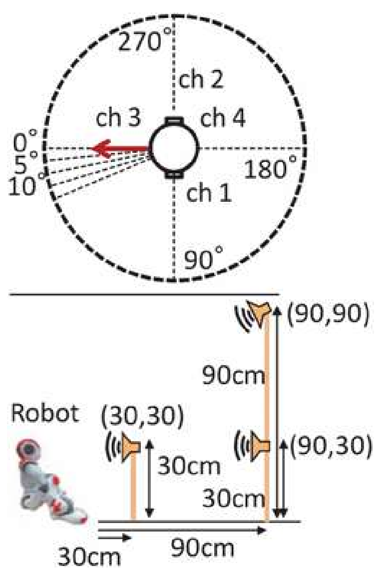
Article notes should be on separate sheets

Source Title	Sound Event Detection Based on Beamformed Convolutional Neural Networks Using Multi-microphones
Source Author	J . Kim, K. Noh, J. Kim and J. Chang
Source citation	J . Kim, K. Noh, J. Kim and J. Chang, "Sound Event Detection Based on Beamformed Convolutional Neural Network Using Multi-Microphones," *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, Guiyang, 2018, pp. 170-173.doi: 10.1109/ICNIDC.2018.8525597
Original URL	**https://ieeexplore.ieee.org/document/8525597**
Source type	Conference Paper
Keywords	Convolutional Neural Network and Sound
Summary of key points	Use neural networks in sound event detection to prevent robots from acting on sounds that normal humans would not and benefit the hearing impaired. They wanted to develop a preprocessing method to remove noise and reverberation from the sound for more accurate detection.

Important Figures	
Reason for interest	Used machine learning to attack a similar problem to me.
Notes	Parameterized multi-channel non-causal Wiener Filter(PMWF) ● preprocessing method ● removes noise and reverberation ● peak normalization used as well to minimize the amplitude difference between different sound sources Deep Learning ● Used MFCC, MBE, LMBE ● Settled on log mel band(LMBE) ● shows highest accuracy among studies ● Used librosa and FFT windows of 370ms ● used CNN and DNN hybrid Post Processing ● used median filtering to reduce small noise Confirmed that sound event based detection needed filtering to be viable
Follow up Questions	Why was an FFT window of 370ms chosen?

Article #8 Notes: Template

Article notes should be on separate sheets

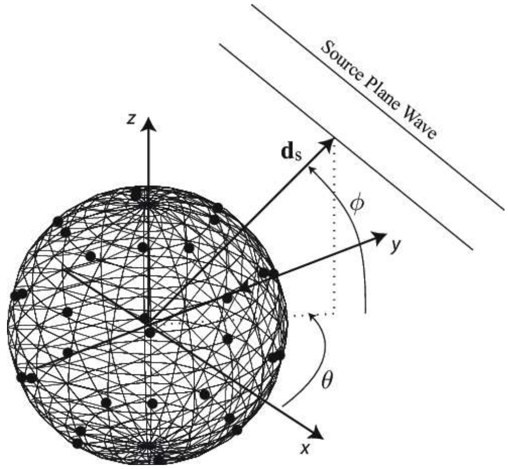
Source Title	Sound source localization based on deep neural networks with directional activate function exploiting phase information
Source Author	R. Takeda and K. Komatani
Source citation	R. Takeda and K. Komatani, "Sound source localization based on deep neural networks with directional activate function exploiting phase information," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 405-409. doi: 10.1109/ICASSP.2016.7471706
Original URL	https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7471706
Source type	Conference Paper
Keywords	Sound source Localization + Neural Network
Summary of key points	Used DNN naively and compared them to a version that they have adjusted and looked for improvements in accuracy.
Important Figures	
Reason for interest	Used a DNN approach on steered vectors to attempt to solve this problem.

Notes	Steering Vectors • representation of intensity and time delay differences • based on reference points in space to the robots • uses SVs as input for the DNN Discriminative Machine Learning • estimates posterior probability • edited by adding hierarchical integration of directional information • added a directional activator that can deal with complex numbers Modified DNN outperformed MUSIC, PHAT, and naive DNN implementations
Follow up Questions	How was hierarchical integration added?

Article #9 Notes: Real-time Sound Source Localization on an Embedded GPU Using a Spherical Microphone Array

Article notes should be on separate sheets

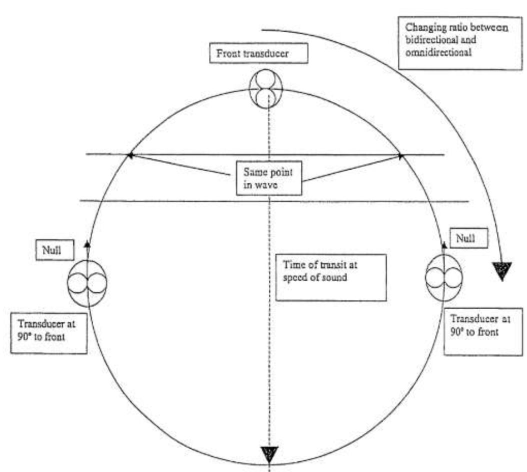
Source Title	Real-time Sound Source Localization on an Embedded GPU Using a Spherical Microphone Array
Source Author	Belloch, Jose & Cobos, Maximo & Gonzalez, Alberto & Quintana-Ortı́, Enrique.
Source citation	Belloch, Jose & Cobos, Maximo & Gonzalez, Alberto & Quintana-Ortı́, Enrique. (2015). Real-time Sound Source Localization on an Embedded GPU Using a Spherical Microphone Array. Procedia Computer Science. 51. 201-210. 10.1016/j.procs.2015.05.226.
Original URL	https://www.researchgate.net/publication/282536580_Real-time_Sound_Source_Localization_on_an_Embedded_GPU_Using_a_Spherical_Microphone_Array
Source type	Article
Keywords	Real-time and Sound Source Localization
Summary of key points	Reverberations easily disturb sound source localization processes, and current need for SSL applications is necessary. Current algorithms need to use Generalized Cross Correlation which is computationally intensive and difficult to run on embedded devices.
Important Figures	

	
Reason for interest	Involved the use of embedded systems, which will be used for my device.
Notes	Most SSL devices are dependent on GCC algorithm ● Very computationally intensive ● Generally unable to compute on embedded systems ● Found that CUDA cores in GPUs are efficient at solving problem Use spherical design with 32 microphones capsules ● High accuracy ● a lot of data to compute, creates need to evaluate different algorithms SRP-PHAT Algorithm ● Effective but computationally heavy algorithm ● take DFT to give you the FFTs ● Then Cross Power Spectrum is done for each pair of microphones to find the difference ● Then a phase transform is taken and an IDFT is performed to give the final azimuth angle Simulation ● testing was done in a 10 by 10 by 5m room ● added white noise to test samples to simulate louder and reverberant environments Results ● as resolution of sound increased the algorithm's efficiency and accuracy increased ● response time remained under 50ms for most resolutions meaning that real-time was achieved ● Accuracy decreased as noisiness increased as expected ● Error never exceeded one degree which made the algorithm very effective.

Follow up Questions	What structural material and attachment method was used so they did not interfere with the sound recording in the sphere? Why was 32 capsules necessary for the array?

Article #10 Notes: US8090117B2

Article notes should be on separate sheets

Source Title	Microphone Array and digital signal processing system
Source Author	James Cox
Source citation	Cox J. (2012) US Patent No. US8090117B2. Retrieved from https://patents.google.com/patent/US8090117
Original URL	https://patents.google.com/patent/US8090117
Source type	Patent
Keywords	Microphone Array + Patent
Summary of key points	Used a spherical shape consisting of microphone pairs to create a sound source localization system.
Important Figures	 FIG. 1

Reason for interest	Microphone array design
Notes	Open geometry spherical shape ● Allows the use of many inexpensive microphones ● place them in opposite pairs for easy differentiation ● Meant for 3D space and location in all directions ● isolate from reverberation and other sound sources ● Strengths are cheap creation with high accuracy and modularity ● 3 dimensional symmetry Many systems copy the middle ear drum mechanism
Follow up Questions	Does the quality of the microphones affect its effectivity?

Article #11 Notes: Microphone Array

Article notes should be on separate sheets

Source Title	Microphone Array
Source Author	Craven P., Law M., Travis C.
Source citation	Craven P., Law M., Travic C. (2013) US Patent No. US8406436B2 Retrieved from https://patents.google.com/patent/US8406436B2/en
Original URL	https://patents.google.com/patent/US8406436B2/en
Source type	Patent
Keywords	Microphone Array + Patent
Summary of key points	Created a device that improved audio capture quality by using a specifically designed microphone array.
Important Figures	
Reason for interest	Microphone Array Design
Notes	
Follow up Questions	

Article #12 Notes: Augmented elliptical microphone array

Article notes should be on separate sheets

Source Title	Augmented elliptical microphone array
Source Author	Meyer, J., Elko G.
Source citation	Meyer J. Elko G. (2010) European Patent No EP 2168396A2 Retrieved from https://patents.google.com/patent/EP2168396A2/fi
Original URL	https://patents.google.com/patent/EP2168396A2/fi
Source type	Patent
Keywords	Microphone Array + Patent
Summary of key points	
Important Figures	
Reason for interest	Microphone Array Design
Notes	
Follow up Questions	

Article #10 Notes: Template

Article notes should be on separate sheets

Source Title	
Source Author	
Source citation	
Original URL	

Source type	
Keywords	
Summary of key points	
Important Figures	
Reason for interest	
Notes	
Follow up Questions	

Article #n Notes: Template

Article notes should be on separate sheets

Source Title	
Source Author	
Source citation	
Original URL	
Source type	
Keywords	
Summary of key points	
Important Figures	
Reason for interest	
Notes	
Follow up Questions	

https://esteemhearing.com/about-hearing-loss/articles/hearing-loss-affecting-well/
https://www.anxiety.org/anxiety-more-common-in-deaf-people
https://www.ncbi.nlm.nih.gov/pubmed/16950865
https://link.springer.com/article/10.1007/s00127-018-1638-3