Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 11
TPCEE 2025

Doppler-Based Sound Localization and Its Application in AI-empowered Traffic Warning for Deafness

Jialin Liu *, #, Jiaqi Liu a

Halifax Grammar School, Halifax, Canada

* Corresponding Author Email: anliu@halifaxgrammar.ca, [a]daliu@halifaxgrammar.ca

[#]These authors equally contributed to the work

Abstract. Hearing is essential for information processing and our safety, especially in traffic. For patients with hearing impairments, sound localization devices help identify threatening sound sources and reduce traffic safety risk. Existing methods typically rely on measuring time delay using microphone arrays. Unfortunately, this method faces critical challenges for dynamic sound sources, posing major threats. Addressing this challenge, we propose a wearable sound localization system that combines interaural time differences and Doppler frequency shifts to determine the position, direction, and speed of moving sound sources. The method's feasibility is first evaluated through a combination of theoretical calculations and experimental verifications. A proof-of-concept setup was established by engaging three microphones as the receiver and a toy car emitting a constant tone as the sound source. An AI algorithm based on artificial neural network was further trained using the received sound signal when the source moved at different locations and directions. The results demonstrated accurate detection of Doppler shifts, with classification accuracies of 100% for front/back, 87.5% for distance, and 62.5% for left/right. Finally, the system was integrated into a wristband with feedback motors, providing vibrational alerts based on the detected motion and proximity of sound sources. These results validate the feasibility of using Doppler shifts and machine learning for motion detection in real time. The proposed system offers a portable solution to enhance the awareness of the environment for individuals with hearing impairments and lays the groundwork for future warning devices in traffic safety applications.

Keywords: Sound localization; Doppler effect; Artificial Neural Network; Deafness assistance.

1. Introduction

Auditory perception is one of the most important sensory abilities and plays an essential role in alerting humans to danger, especially as hearing provides 10% of the information we use to perceive the environment [1]. For example, sound localizability is very important to spatial orientation as well as for self-protection. For deaf or single-sided deaf patients, directional audiological unawareness is a major impairment affecting the activities of everyday life, primarily in situations having dangerous on-coming sound-bearing vehicles like autos. As opposed to the normally binaural human being, individuals with single-sided deafness are commonly incapable of the determination of location of the sound. This sensory loss affects their ability to determine where sound is occurring and respond accordingly, making them more likely to be involved in accidents. According to the 2014 National Health Interview Survey in the US, people with a hearing impairment were almost twice as likely to suffer accidental injuries than others with a better hearing [2]. To counter this issue, systems capable of properly identifying sounds in real-time are required to provide warnings and can support individuals who have impaired hearing.

Existing methods employ microphone arrays to estimate the Time Difference of Arrival (TDOA) among sensors [3, 4]. However, the current difficulty in this field is achieving real-time and accurate sound localization in open areas. Moreover, existing techniques cannot deal with or even fail to consider high-speed moving targets or with more than one sound source. There is also the need to account for the determination of the moving direction and velocity of sound sources to detect possible threats. Hence, developing a technology that can accurately locate multiple high-speed moving sound sources would be essential.

In mammals, sound localization is achieved by analyzing acoustic waveforms received at both ears, inspiring microphone arrays that replicates this mechanism [5]. However, laboratory studies often limit variables (e.g., fixed speaker positions, restricted head movements, simplified sounds), limit their applicability to real-world environments involving moving sound sources or receivers [6].

To address this, Baumann et al. developed a binaural system capable of sensing interaural time delays (ITDs) and angular velocity by introducing system rotation, enabling the instantaneous estimation of the sound direction through a differential equation model [7]. Complementing this, Li et al. found that the Doppler effect enhances the plausibility of the simulated moving virtual sound source in virtual 3D audio simulations [8]. Building on this, a binaural Doppler model combining interaural time differences (ITDs), their time derivatives, and known rotational motion can accurately estimate sound direction in the plane of rotation and shows bounded error propagation under controlled motion [9].

Furthermore, Lindgren et al. estimated source position and velocity by modeling the Doppler effect as a nonlinear least squares problem and solves it efficiently using the Gauss-Newton method with variable projection [10].

Similarly, Martín et al. tracked aircraft's position and velocity using a seven-microphone array and Doppler optimization, demonstrating that it did not depend on specific microphone distributions [11]. In parallel, a Moving Time Difference of Arrival (MTDOA) based triangulation method with Doppler correction was proposed to accurately localize moving sound sources with higher resolution and lower computational cost compared to traditional triangulation methods [12].

Another case to consider is when the velocity of a moving source is non-negligible compared to the speed of sound in air. In this scenario, Doppler-induced distortions allow the simultaneous calculation of the relative Doppler stretch and the time delay, which a general particle filter framework can directly use for sound localization without intricate preprocessing of raw signals [13].

Along with physics-based models, the rise of AI has brought data-driven methods into sound localization, where deep learning enables adaptive reasoning but still struggles with generalization, requiring realistic reinforcement learning environments [14].

Based on the above literature survey, we infer that the Doppler-based method holds great promise for moving sound source localization. However, the existing techniques have been designed for fast-moving objects like aircraft, and their suitability in other types of vehicles like automobiles and motorcycles have not been evaluated. Moreover, the reported sound localization instruments are typically bulky and are not suitable for disabled assistive applications. More efforts should be devoted to extending the application scenario and enhancing the wearability of Doppler-based devices.

In this project, we have developed an assistive device that can locate adjacent moving vehicles and providing real-time warning to hearing-impaired individuals. Efficient sound localization was achieved by combining the ITDs and Doppler shift in a microphone array. A machine learning algorithm was also developed for accurate detection of the position, moving direction, and speed of the moving sound source. The developed AI algorithm for risk analysis was also integrated into wearable wristbands for in situ warning functions for the deafness patients. This research advances the understanding of Doppler effects and presents a novel, intriguing application scenario for sound localization.

2. Modeling

Firstly, we simulate the acoustic signal received by the microphone array. As schematically shown in Figure 1. We assume a 2D Cartesian coordinate system. A sound source (with a frequency of f) located at (x0, y0) is moving at a velocity of v with an azimuth angle of θ.

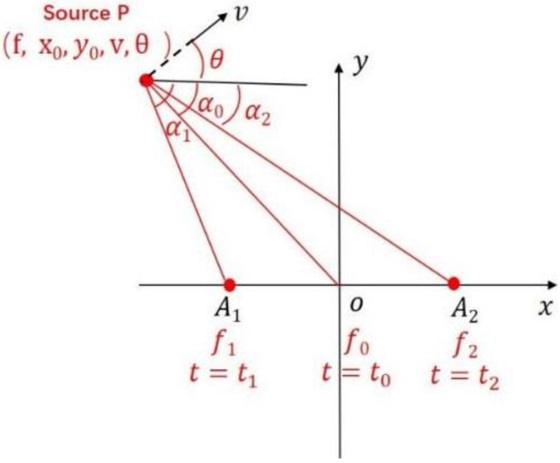

Figure 1. Schematic of the theoretical model.

Three microphones are place on site A1, O, and A2 on the x-axis, with an equivalent spacing of a, in analogous to the case where they are evenly placed on the shoulder of a human. In the conventional ITD method, the time delay among the three sensors is associated with the coordinate of the sound source:

$$t0\text{-}t1=\frac{\sqrt{x_0^2+y_0^2}}{u}-\frac{\sqrt{(x_0+a)^2+y_0^2}}{u} \tag{1}$$

$$t2\text{-}t0=\frac{\sqrt{(x_0-a)^2+y_0^2}}{u}-\frac{\sqrt{x_0^2+y_0^2}}{u} \tag{2}$$

and the position of the sound source can be located. Nonetheless, such a method is unable to directly determine the moving speed and direction of the source, thus is unsuitable for early warning applications. By considering the Doppler frequency shift at each receiver, we further get:

$$f_1=\frac{u}{u+v\cos(\theta+\alpha_1)}f_0 \tag{3}$$

$$f_0=\frac{u}{u+v\cos(\theta+\alpha_0)}f_0 \tag{4}$$

$$f_2=\frac{u}{u+v\cos(\theta+\alpha_2)}f_0 \tag{5}$$

Where

$$\alpha_1=\text{atan}\frac{y_0}{x_0+a} \tag{6}$$

$$\alpha_0=\text{atan}\frac{y_0}{x_0} \tag{7}$$

$$\alpha_2=\text{atan}\frac{y_0}{x_0-a} \tag{8}$$

By combining these three Doppler equations with the two ITD equations, we can easily determine the five parameters of the sound source. Such a method has remarkably extended the scope of sound localization.

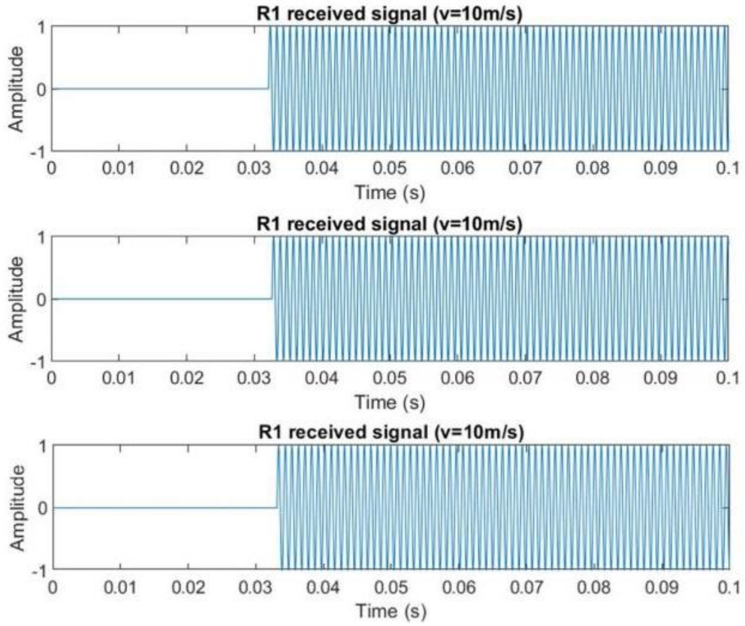

Figure 2. Simulated waveforms at site A_1, O, and A_2.

To verify the feasibility of the method, we have simulated the signal at each of the receiver using Matlab. In the simulation, the wave frequency f= 1000 Hz is used. The source is placed at (x0=10 m, y0=5 m), moving at a speed of v =10 m/s and a θ=45º. The spacing of microphones, a=0.2 m, is set to simulate the case where they are placed on the shoulders. u= 343 m/s is used as the speed of sound at room temperature. The simulated waveforms are shown in Figure 2. A gradual phase shift can be observed by comparing the waveforms at the three positions. The differences in simulated frequencies at each site, 1028.4575 Hz, 1028.5331 Hz, and 1028.609 Hz for A1, O, and A2, also promises a notable Doppler shift at different positions.

3. Experimental Details

The experiment uses the following equipment: three rechargeable wireless microphones, one wireless speaker, and a toy car. The equipment is positioned as illustrated in the figure.

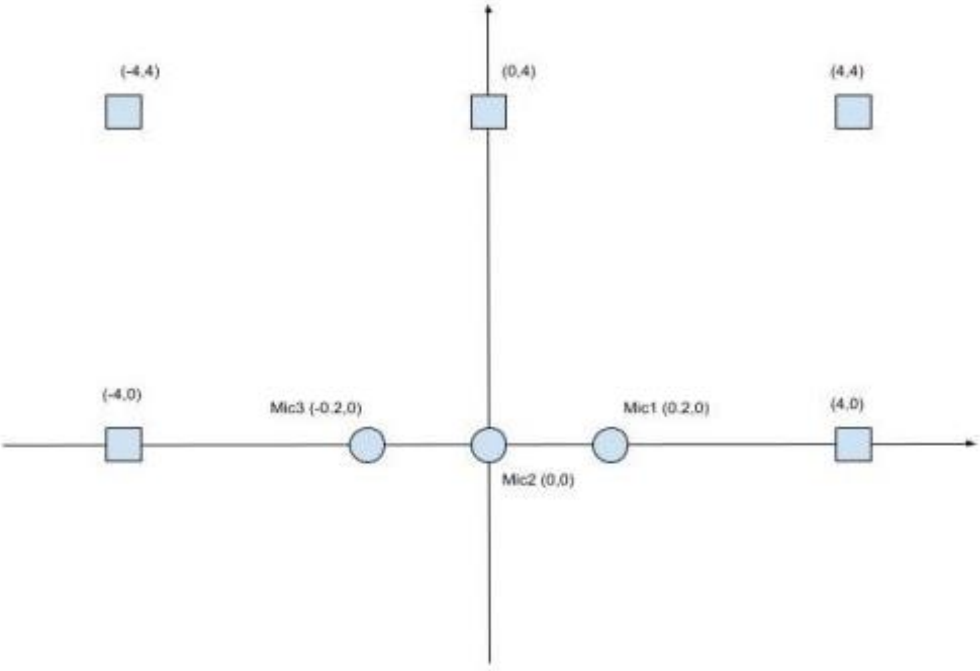

Figure 3. Positioning of equipment.

As shown in the figure, the microphones are depicted as the circle and placed on a horizontal line with microphone 2 at the origin, O. Microphone 1 and 3 are placed at points A1 and A2, respectively, and are a distance of a = 0.2 m away from Microphone 2.

4. Experimental Results

4.1. Static Source Experiment

With the equipment fully tested and the Doppler effect verified, we begin collecting data on a static sound source emitting a constant sine wave at 1000 Hz. This setup is used to validate time delay measurements.

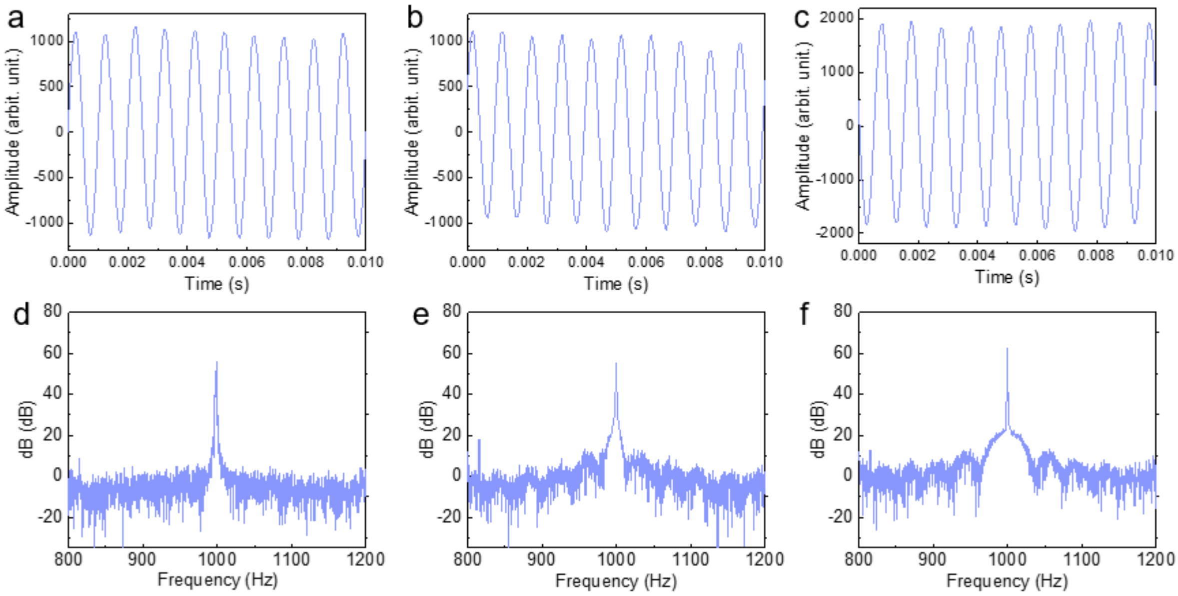

Figure 4. The received waveforms at the three microphones and their respective frequency spectrums.

Figure 4a, 4b, 4c shows the waveforms of the sound received at microphones 1, 3, and 2 respectively. A clear time delay can be observed, corresponding to the difference in distances between the microphones and the sound source. The recorded waveforms accurately reflected the sinusoidal shape of the signal, with a high signal-to-noise ratio (SNR). This delay can provide a measure of a phase shift that can be later used to estimate sound direction.

Figure 4d, 4e, 4f presents the corresponding frequency spectrum obtained from Fast Fourier Transform (FFT) of the recorded signals. All three microphones have a strong peak close to 1000 Hz, consistent with the emitted tone from the speaker. The frequency deviation between the received signals is less than 0.04%. Such a high accuracy along with a SNR exceeding 50 dB highlight the microphones' ability to record clean signals without receiving many significant disruptions, validating the system's accuracy in static conditions.

This static setup confirms the system's ability to accurately determine the small-time differences between microphones and detect precise frequencies, forming the foundation for later dynamic source experiments.

4.2. Dynamic Source Experiment

Using the validated static configuration, the next phase introduces a moving sound source to study Doppler shifts and their implications for detecting direction and classifying distance. In this setup, the wireless speaker is placed on a toy car, moving at a constant speed along trajectories parallel to the x-axis towards the microphones. The emitted tone remains constant at 1000 Hz during each trial.

Clear frequency shifts towards higher frequencies (~1004 Hz) were observed in Figure 5, confirming the fidelity of the Doppler method.

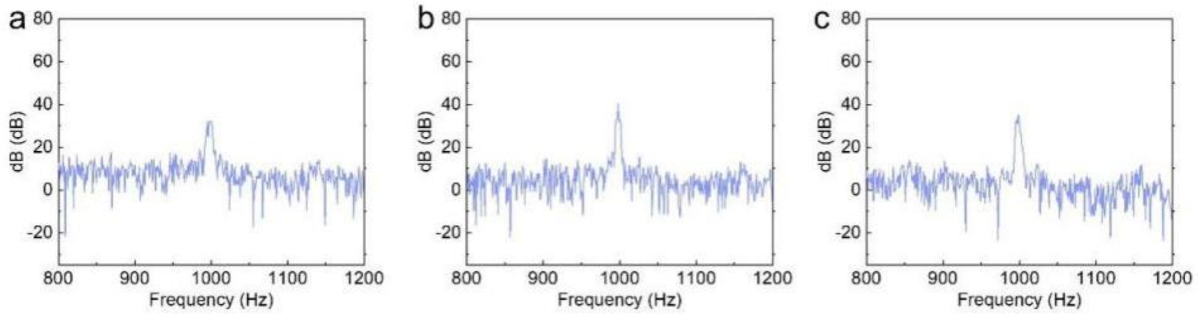

Figure 5. The FFT spectrum of the three microphones with moving sound source.

Figure 5a, 5b, 5c presents the corresponding frequency spectrum obtained from Fast Fourier Transform (FFT) of the recorded signals as the toy car is moving towards the microphones. The microphones have recorded similar frequencies with each other at a bit above 1000 Hz, confirmed by the Doppler effect equations. Thus, this proves the system's ability to function with a moving source.

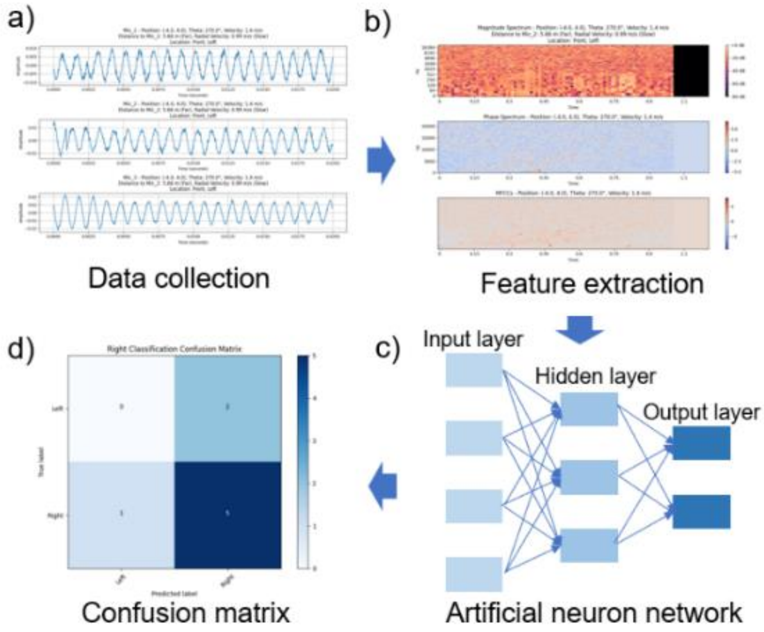

Figure 6. The process of training the data obtained by the microphone signals using a neural network.

Based on the setup, the sound localization was conducted by integrating the machine learning algorithms. The schematic of the process is shown in Figure 6a. The waves for each dynamic trial are recorded by the microphones, where an example is shown in Figure 6a. These signals are processed using feature extraction techniques, as shown in Figure 6b, including the magnitude spectrum, the phase spectrum, and MFCCs (Mel-frequency cepstral coefficients).

The magnitude spectrum represents the distribution of the amplitude of a signal's frequency shifts.

The phase spectrum represents the relative shifts of the frequency components.

MFCCs captures the important characteristics of an audio and is used for efficient machine learning.

The extracted features serve as inputs to an artificial neural network (ANN), as shown in Figure 6c, which is trained to classify three aspects of motion: front-back direction, left-right direction, and distance to the microphone array. The model's structure consists of an input layer of the extracted features, followed by one or more hidden layers, and an output layer for the respective classification or regression tasks.

Training results are shown in the table below. The model has achieved stable convergence with low Mean Absolute Error (MAE) and Weighted MSE on validation data. For each distance, front, and left classification, the respective accuracy and binary cross-entropy were tracked and recorded in Figure 7.

Distance Classification maintained a constant accuracy around 87.5%, indicating strong generalization

Front Classification reached 100% accuracy rapidly

Left Classification reached 62.5% final accuracy, proving it to be more challenge. This is confirmed by figure 7d.

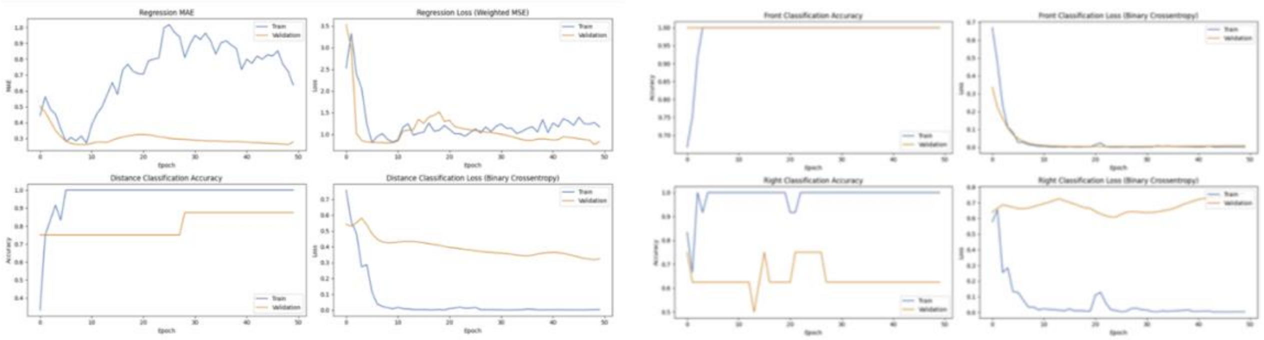

Figure 7. The training history of the model.

Table 1. The respective prediction accuracy of front/back classification, left/right classification, and distance detection.

Prediction	Front-back	Left-right	Distance
Accuracy	100%	62.5%	87.5%

These results demonstrate the system's ability to classify direction and distance of a moving source based on audio data and neural network inference. Although left and right classification remains less accurate than the other categories, the overall performance confirms the possibility of using this system to detect the motion of a sound source.

5. Application in Wearable Traffic Warning System

The designed system provides a foundation for real-time motion detection, classifying the direction and distance of a moving source. The success in training such as AI algorithm has provided a solid foundation for developing a traffic warning system for hearing impaired individuals. Based on the above experimental results, we further integrate a physical wristband with vibration motors to warn users in complex environments.

To ensure the portability and usability, we designed and 3D-printed holders for the entire system, as shown in Figure 8. Figure 6a shows the holder case for the ESP32 board, shown in figure 9, which is a circuit board in which the controls the motors and tells them when to vibrate. Figure 6b shows the cap for the ESP32 case, preventing it to fall out of place. Figure 6c shows the platform on which the motors are placed. The entire system utilizes four motors and one ESP32 board, with everything placed on the wristband.

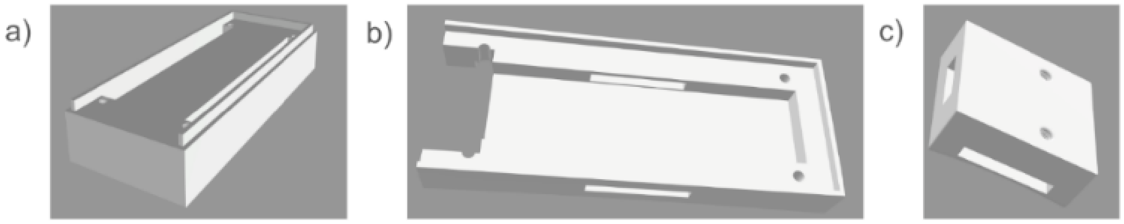

Figure 8. The 3D printing model of the holder and caps for the vibration motors and esp32 board.

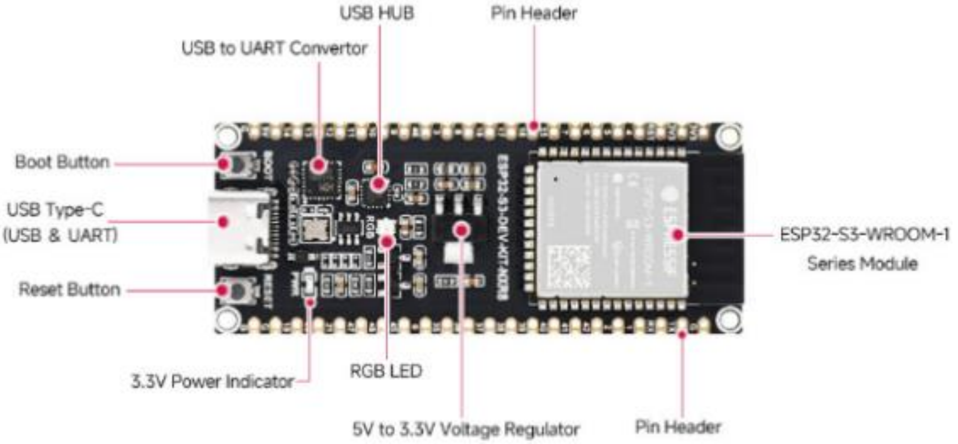

Figure 9. Model of the ESP 32 board used.

When an incoming sound is determined to be closer than 5 meters, the system triggers the vibration motors at a speed of 255Hz while a source further than 5 meters would trigger a vibration speed of 100Hz. This signal would be sent to the front/back/left/right motors depending on the direction of the acoustic sound source. This configuration is useful for individuals with unilateral hearing loss, and even pedestrians or cyclists navigating urban areas. The design also allows for flexible retraining of the classification model to adapt to new environments or new purposes. The final product is shown in figures 10a and 10b below.

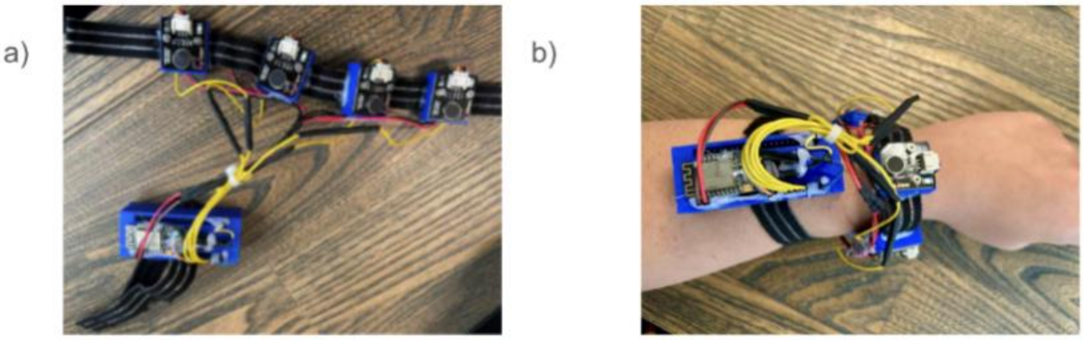

Figure 10. The final wristband product produced

In summary, this system provides robust motion classification, using the trained model and responding with vibrations on the motors to signal the users. It also provides the foundation for real-time alerting, allowing for safety.

6. Conclusion and Outlook

In conclusion, this work presents a microphone-array-based system that can classify the direction and distance of a moving sound source using Doppler shifts and neural network inference. Using a static source, we validated the system's ability to measure time delays and frequency shifts. For dynamic sources, the system successfully extracted relevant features and utilized them to achieve high classification accuracies for front-back (100%) and distance (87.5%), and achieved moderate performance in left-right (62.5%) classifications. We have further integrated such an algorithm into a wearable wristband to provide vibration warnings of incoming vehicles or other noise makers.

7. Methods

Before beginning each trial, all three microphones were fully charged and connected to the same wireless receiver to ensure aligned data acquisition. To verify each device's functionality, all microphones were tested using a steady tone of 1000 Hz that was emitted from the speaker, and the corresponding sound at each receiver was recorded and analyzed using a self-written Python code. All trials were done after ensuring the functionality of all three microphones.

In analogous to a moving sound source, the wireless speaker is placed on the toy car using a tape to stabilize the sound source. The speaker continuously emits a pure tone of 1000 Hz, as the car is set in motion along a straight track aligned parallel to the x or y axes. The four driving directions, left, right, up, and down, were tested individually by driving the car in each respective direction relative to the positive x-axis.

To confirm that the Doppler effect can be monitored in the experiment, a preliminary Fast Fourier Transform (FFT) analysis is performed. By analyzing the frequency spectrum of the recorded audio signals as the toy car moves toward or away from the microphones, the shifts in frequency due to motion are observed. The frequency shift in the recorded signal is commuted using short-time FFT windows of 0.2 seconds. To increase the frequency resolution of FFT analysis, the peaks in frequency spectrum are fitted with Gaussian equation. An increase in peak frequency is observed when the car approaches the microphones and a decrease when the car moves away from the microphones, consistent with the expected Doppler effect pattern. This validates the setup and the ability to capture real-time frequency variation when the car moves in the predetermined directions.

This setup confirms the functionality of the equipment and the system's potential to detect real-time frequency shifts of a moving source.

References

[1] Man, D., Olchawa, R., The Possibilities of Using BCI Technology in Biomedical Engineering. In: Hunek, W., Paszkiel, S. (eds) Biomedical Engineering and Neuroscience. BCI 2018. Advances in Intelligent Systems and Computing, vol 720. Springer, Cham.

[2] Lin HW, Mahboubi H, Bhattacharyya N., Self-reported Hearing Difficulty and Risk of Accidental Injury in US Adults, 2007 to 2015. JAMA Otolaryngol Head Neck Surg. 2018, 144 (5): 413 – 417.

[3] M. Pollefeys and D. Nister, Direct computation of sound and microphone locations from time-difference-of-arrival data, IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 2008, 2445 - 2448.

[4] J. Nikunen and T. Virtanen, Time-difference of arrival model for spherical microphone arrays and application to direction of arrival estimation, 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 2017, 1255 – 1259.

[5] Grothe, B., Pecka, M., & McAlpine, D., Mechanisms of sound localization in mammals. Physiological Reviews, 2010, 90 (3): 983 - 1012.

[6] Chantai Laroche, Review of the Literature on Sound-Source Localization and Applications to Noisy Workplaces, Canadian Acoustics, 1994, 22 (4): 13.

[7] Baumann, C., Rogers, C., & Massen, F., Dynamic binaural sound localization based on variations of interaural time delays and system rotations. The Journal of the Acoustical Society of America, 2015, 138 (2): 635 - 650.

[8] Li, S., Moghaddam Nia, S., & Peissig, J. (2016). Impact of Doppler Effect, Echo and Reverberation on the Externalization and Plausibility of Binaural Rendered Moving Sound Sources Presented via Headphones. Fortschritte der Akustik-DAGA, 42, 1004 - 1007.

[9] Baumann, C., Rogers, C., & Massen, F. (2015). Dynamic binaural sound localization based on variations of interaural time delays and system rotations. The Journal of the Acoustical Society of America, 138 (2), 635 - 650.

[10] Lindgren, D., Guldogan, M. B., Gustafsson, F., Habberstad, H., & Hendeby, G. (2013). Acoustic Source Localization in a Network of Doppler Shift Sensors. 16th International Conference on Information Fusion. Presented at the 16th

International Conference on Information Fusion, 9 - 12 July 2013, Istanbul, Turkey. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-97309.

[11] Martín, S. R., Genescà, M., Romeu, J., & Pàmies, T. (2013). Aircraft tracking by means of the Acoustical Doppler Effect. Aerospace science and technology, 28 (1), 305 – 314.

[12] Miao, F., Yang, D., Wen, J., & Lian, X. (2016). Moving sound source localization based on triwangulation method. Journal of Sound and Vibration, 385, 93 - 103.

[13] Chen, G., & Lu, Y. (2020). Fast-Moving Sound Source Tracking with Relative Doppler Stretch. IEEE Access, 8, 221269 – 221277.

[14] Jekateryńczuk, G., & Piotrowski, Z. (2024). A Survey of Sound Source Localization and Detection Methods and Their Applications. Sensors, 24 (1), 68.