Understanding Sound Localization and Augmented Reality for Hearing Accessibility

Mayu Zuniga
Carter G. Woodson High School
Fairfax, United States
mzumo365@gmail.com

Lucas Shen
Carter G. Woodson High School
Fairfax, United States
lucasshen25@gmail.com

Logan Lee
Carter G. Woodson High School
Fairfax, United States
logan.sm.lee09@gmail.com

***Abstract*— Individuals who are deaf or hard of hearing often struggle to locate and identify sounds in noisy and dynamic environments, leading to decreased situational awareness and safety risks. While conventional hearing aids attempt to solve this issue by amplifying sound and helping with speech perception, they often do not reliably improve sound localization, provide limited benefit for individuals with profound or complete deafness, and can be prohibitively expensive for many users. The researchers devised a solution to this problem by developing N1 glasses, augmented reality glasses that use sound localization algorithms to mimic human spatial awareness. Through cross correlation and audio beamforming, this project proposes a novel method to portray sound as visual indicators to increase awareness for the deaf and hard-of-hearing. Using an array of PDM MEMS microphones, the system collects sound from a given sound source and visually displays its direction, utilizing a microLED projector and optical waveguide display. Despite physical and computational constraints, this project was able to accurately imitate human ears in real time with a root-mean-square error of approximately 4.9°. In comparison, the average individual shows a root-mean-square error between 2.8° and 7.3°. These results suggest that real-time sound visualization could serve as a feasible assistive technology to enhance spatial awareness, with potential applications in navigation aids or safety systems for users with hearing loss.**

Keywords—augmented reality, cross-correlation, hearing accessibility, audio beamforming

I. Introduction

In everyday environments, auditory cues such as speech, traffic noise, and warning signals provide critical situational information. For millions of individuals in the deaf and hard-of-hearing (DHH) community, access to this information is limited or unavailable, resulting in reduced environmental awareness and increased safety risks. This challenge is expected to grow; the World Health Organization projects that by 2050, nearly 2.5 billion individuals will have varying degrees of hearing loss [1]. Current assistive technologies such as hearing aids primarily focus on sound amplification rather than spatial awareness, leaving a significant gap in how DHH individuals perceive and locate sounds in their environment. Additionally, the high cost of conventional assistive hearing devices, including hearing aids and cochlear implants, limits accessibility for many potential users. This project explores the cost-accessible design of augmented reality (AR) glasses that detect the direction and type of sounds in the environment and display that information visually in real time. By translating sound into clear visual cues, these glasses aim to provide a novel method of improving situational awareness and independence for DHH users using accessible, modern technology.

II. Background

Human sound localization relies on three primary biological mechanisms: interaural time differences (ITD) that reflect the difference in arrival time of sound between the left and right ear, interaural level differences (ILD) that correspond to differences in sound intensity due to the head blocking or attenuating certain frequencies, and spectral shaping that adds vertical and front-back directional cues through a process called the head-related transfer function (HRTF) [2]. Together, these three cues allow the brain to quickly determine sound location in an environment [3]. When DHH individuals lose these binaural and monaural cues, they lose crucial environmental awareness, driving the development of assistive technologies that can hopefully replicate these localization mechanisms without relying on residual hearing.

Hearing aids attempt to address hearing loss through sound amplification, compensating for auditory deficits by increasing the volume of incoming sound. While hearing aids remain the most widely used assistive technology for individuals with hearing loss, they have significant limitations. Most critically, research has shown that bilateral hearing aids do not reliably preserve localization cues, with aided users performing comparably to or sometimes worse than unaided users in localization tasks [4]. For individuals with unilateral hearing loss, the limitation is even greater; studies have found that monaural listeners are essentially unable to localize sound even when using hearing aids or bone-anchored devices [5]. Even cochlear implants, which provide significant speech perception benefits, leave listeners with localization errors averaging approximately 28 degrees compared to roughly 6 degrees for normal hearing listeners [6]. Beyond technical limitations, adoption rates for hearing aids are low, with only approximately

16 percent of adults aged 20 to 69 who could benefit from using hearing aids having ever used them [7]. Contributing factors to this include cost and stigmatizing mindsets surrounding hearing loss and assistive devices [1]. These gaps in sound awareness and device adoption evidence the need for alternative approaches to providing directional sound information to DHH individuals.

III. Methodology

A. System Architecture

The N1 glasses prototype consists of three primary subsystems: a microphone array for capturing spatial audio, a central processing unit for real-time signal processing, and a waveguide display module for communicating directional information to the user. Fig. 1 illustrates the overall system architecture. Each subsystem was designed to balance performance with the physical constraints of a wearable glasses form factor.

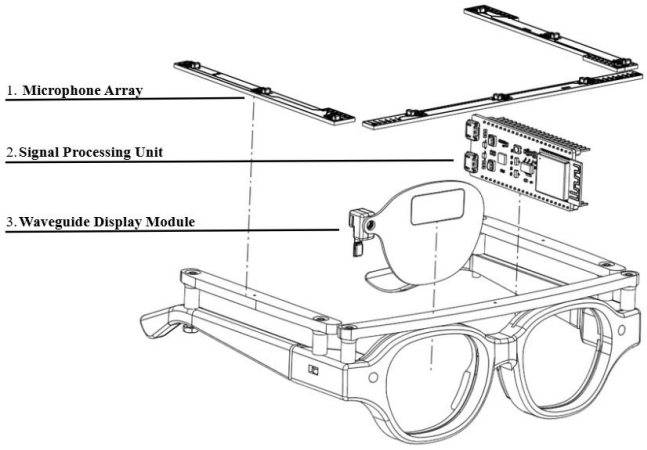


Fig. 1. System Architecture Overview.

The microphone array consists of eight IMP34DT05TR MEMS microphones with a pulse density modulation (PDM) digital output. The microphones are arranged in a U-shaped configuration along the glasses frame, with three along each temple arm and two along the front crossbar, providing broad azimuthal coverage from a form factor similar to standard eyewear. Exact microphone positions in a body-fixed coordinate system with origin at mic 1 (left temple tip) are given in Table 1.

Table 1. **Microphone positions in body-fixed coordinate system**

Mic #	x (mm)	y (mm)	Location
1	0	0	Left temple tip
2	47	0	Left temple mid
3	92	0	Left temple inner
4	103	36	Front bar left
5	103	126	Front bar right
6	92	160	Right temple inner
7	47	160	Right temple mid
8	0	160	Right temple tip

Spacing is non-uniform; adjacent microphones across each temple arm are separated by 45 mm, while the two front microphones are separated by 90 mm. The temple spacing limits spatial aliasing to frequencies below approximately 3,800 Hz, and the front spacing to approximately 1900 Hz, both of which encompass the majority of speech and common environmental sounds [8]. Each microphone is mounted on a custom PCB attached to the glasses frame, with the acoustic port oriented upward. The IMP34DT05TR exhibits sufficient omnidirectional response across the frequency range used for localization to maintain consistent sensitivity across horizontal sound directions. Fig. 2 shows the microphone positions on the prototype.

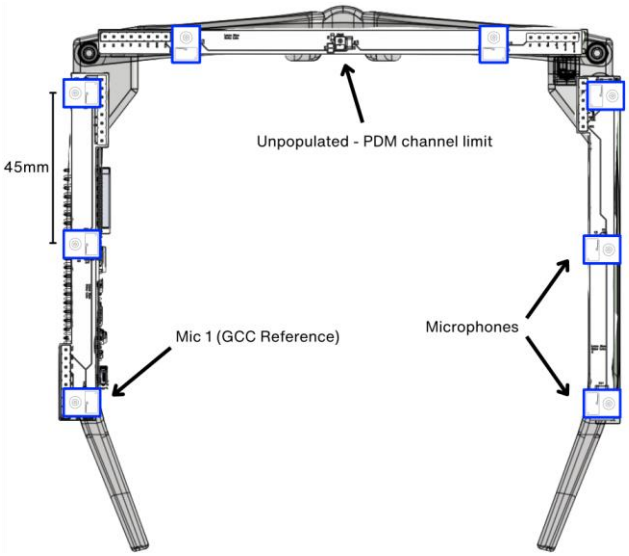


Fig. 2. Microphone array placement.

The processing unit is built around an ESP32-S3 microcontroller, operating at up to 240 MHz. The microcontroller interfaces with the eight PDM microphones through a single I2S peripheral configured in PDM receive mode, using four data lines each carrying a stereo channel pair to yield eight simultaneous audio streams. Audio samples are captured at 16 kHz per channel and buffered in blocks of 512 frames before being passed to the DSP pipeline described in Section III-B. The processed angle estimate is passed to a companion rendering script, which translates the directional estimate into a visual arc transmitted to the display module via HDMI.

The display subsystem uses a Jade Bird Display (JBD) AmµLED Hummingbird Mini II microLED projector coupled with an optical waveguide to overlay visual information onto the user's field of view. The projector produces a monochrome green image at a resolution of 640 x 480 pixels with 4-bit grayscale depth and a 30° diagonal field of view. The waveguide redirects the projected directional arc so that it appears within the user's line of sight while maintaining transparency for normal vision. Both the projector and waveguide are off-the-shelf components.

The electronic components are housed in a custom frame 3D printed in PLA on a Bambu Labs A1 printer. The frame was designed in Onshape to accommodate the microphone array,

processing unit, and display module while approximating the dimensions of conventional eyeglasses. Six iterations of the frame design were produced to address progressive refinements including microphone placement and spacing along the temple arms, PCB mounting and retention, display module alignment with the user's line of sight, and overall fit and weight distribution. Fig. 3 shows an exploded view of the CAD model with each subsystem labeled.

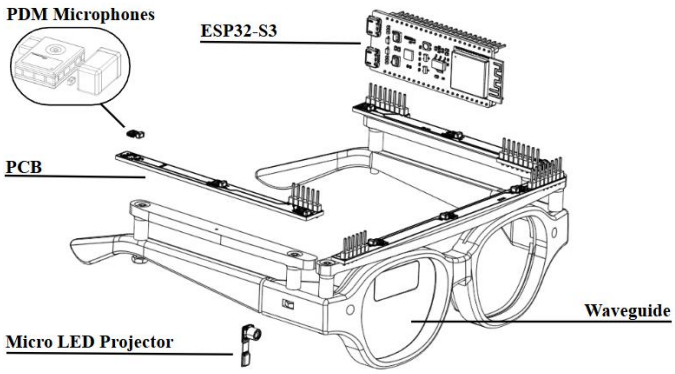


Fig. 3. N1 glasses exploded view.

Custom printed circuit boards were designed on KiCad and iterated to integrate the microphone array into the glasses frame, with the final revision consolidating the IMP34DT05TR microphones, passive components, and solder bridges into a modular system optimized for signal integrity. A custom microLED driver board was additionally designed and manufactured; full integration onto the frame is planned for future work.

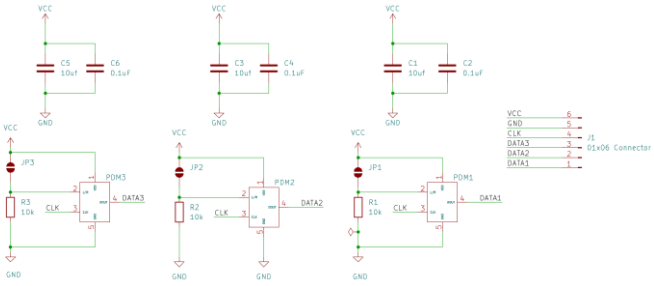


Fig. 4. Right-temple microphone array PCB schematic.

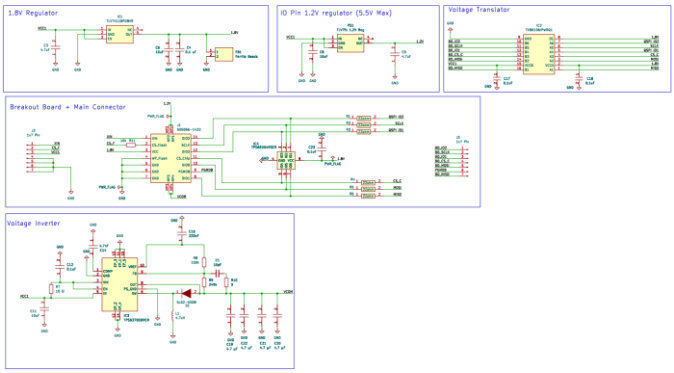


Fig. 5. PCB custom microLED projector driver board schematic.

The total bill of materials for the N1 prototype is approximately $297, comprising $49 for the microphone array, $15 for the processing unit, $193 for the display module, and $5 for the frame and supporting components. Fig. 6 shows the fully assembled prototype.

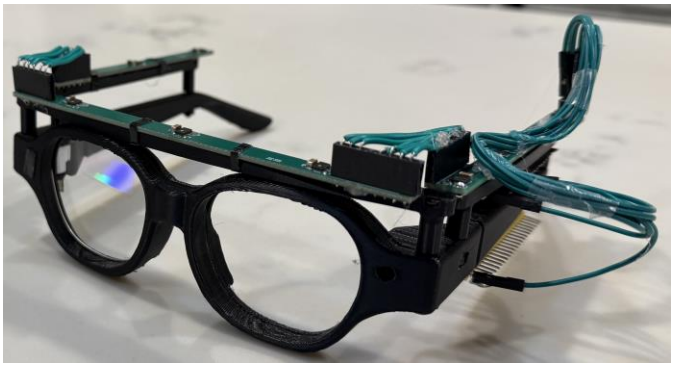


Fig. 6. N1 glasses prototype picture.

This cost is substantially lower than conventional hearing aids, which average approximately $2,500 per pair [9], and cochlear implants, for which device costs alone average $34,000 to $38,000 per hospital before surgical and rehabilitation expenses [10].

B. Sound Source Localization

The N1 signal processing pipeline runs entirely on the ESP32-S3 microcontroller in real time, operating on consecutive 512-sample frames captured at 16 kHz (approximately 32 ms per frame). The pipeline consists of four stages: signal preprocessing, direction-of-arrival (DOA) estimation via a hybrid GCC-PHAT and SRP-PHAT algorithm, delay-and-sum beamforming, and voice activity detection. Fig. 7 depicts the complete signal processing flowchart.



Fig. 7. Signal processing flowchart.

Audio is captured across four I2S PDM data lines, each carrying two channels, for a total of eight simultaneous microphone inputs. The interleaved samples are separated into per-microphone buffers of 512 samples each. Before processing, a level gate evaluates the mean absolute amplitude of one reference microphone; if the amplitude falls below a minimum threshold, the frame is discarded to prevent the system from producing spurious angle estimates during silence. For each remaining frame, a per-microphone DC offset is removed by subtracting the block mean from every sample. A Hann window is then applied to each channel to reduce spectral leakage, and a 512-point radix-2 fast Fourier transform is computed in place. The FFT implementation was written directly for the ESP32-S3 rather than imported from an external library, in order to minimize memory overhead and maintain control over fixed-point behavior on the embedded platform.

The primary DOA estimation method is GCC-PHAT (generalized cross-correlation with phase transform) [11], which estimates the direction of a sound source by measuring the time difference of arrival (TDOA) between microphone pairs. For

each pair, the cross-power spectrum of both signals is computed and normalized so that all frequencies contribute equally regardless of signal strength, then transformed back to the time domain to produce a sharp peak at the lag corresponding to the inter-microphone delay [12]. In the N1 implementation, this computation is band-limited to 200 to 3,500 Hz, covering the speech frequency range while rejecting low-frequency environmental noise and high-frequency content prone to spatial aliasing. A parabolic interpolation step around the correlation peak provides sub-sample delay resolution. Rather than estimating direction from a single pair, the system performs a weighted least-squares fit across all seven pairs formed between the reference microphone and each remaining microphone, solving the direction vector that best explains all observed TDOAs at once. Pairs are weighted by their baseline distance squared and peak correlation strength, so that widely spaced pairs with strong, clean peaks exert greater influence on the final estimate. The output is a single azimuthal angle in degrees, with 0° defined as directly in front of the wearer.

The secondary DOA method is SRP-PHAT (steered response power with phase transform) [13]. Unlike GCC-PHAT, which estimates a TDOA first and converts it to an angle, SRP-PHAT directly evaluates candidate directions by measuring how well each one explains the observed cross-correlations across multiple microphone pairs. The implementation selects 12 of the 28 possible microphone pairs to maximize geometric diversity while remaining computationally feasible on the ESP32-S3. For each candidate direction, the algorithm computes the expected inter-microphone delay based on array geometry and the speed of sound, then interpolates the precomputed GCC-PHAT correlation function at that predicted lag. The interpolated values are summed across all 12 pairs, weighted by baseline distance squared, to produce a steered response power score for that direction. A two-stage search is used: a coarse scan at 3° increments across the full 360°, followed by a fine scan at 1° resolution within ±8° of the coarse maximum. To reduce frame-to-frame jitter, an exponential moving average with a smoothing factor of 0.70 is applied to the correlation values across consecutive processing blocks, stabilizing the SRP spatial power surface over time.

Neither algorithm performs reliably across all angles due to the non-circular geometry of the microphone array. The eight microphones are distributed along the glasses frame in an approximately rectangular arrangement rather than a uniform circle, which produces uneven geometric coverage across the azimuthal plane and reduces localization reliability in certain angular regions for both algorithms. Early versions of the system exhibited significant accuracy gaps in these regions. To address this, the system employs a hybrid approach. GCC-PHAT runs continuously on every frame and produces an angle estimate regardless of direction. This estimate is used to determine which angular zone the sound source currently occupies. If the GCC-PHAT estimate falls between 157° and 281° (the rear hemisphere) or within a forward window between 330° and 43° (a 73° region passing through 0°), the GCC-PHAT output is used directly, as the least-squares fitting produces stable estimates in these regions. If the estimate falls in the remaining lateral angles, the system switches to the SRP-PHAT output, where the exhaustive spatial search better accommodates the pair geometries in those regions. While GCC-PHAT is reliable enough in the lateral zones to correctly identify that the sound source is within SRP-PHAT's operating range, its angular resolution in those areas is coarser, meaning it has difficulty distinguishing between nearby angles. Some elevation in error at 45°, 225°, and 270° is consistent with the reduced angular resolution near the boundaries of the GCC and SRP-PHAT operating zones, where the determining algorithm has greater difficulty distinguishing between closely spaced angles.

To further improve temporal consistency, a beamformed spectral mask provides a feedback mechanism between consecutive processing blocks. After each DOA estimate, a delay-and-sum beamformer is steered toward the previous block's output angle [14]. The beamformed signal is transformed into the frequency domain, and each bin's magnitude is compared to the approximate median magnitude across the half-spectrum. Bins where the beamformed signal exceeds the median receive an upward weight, while bins below the median are attenuated. This per-bin mask is applied as a multiplicative factor inside the GCC-PHAT cross-spectrum computation on the subsequent block, biasing the correlator toward frequency content consistent with the currently tracked direction. The weighting coefficient is set to 0.45, and the mask is clamped between 0.6 and 1.6, which allows the mask to reinforce frequency bins that carry directional energy while remaining bounded enough to prevent the system from locking onto a false direction when the source moves. The result is an improvement in estimate continuity across consecutive frames, particularly in environments where intermittent noise would otherwise cause the raw DOA estimate to fluctuate. The beamformed signal additionally serves as the audio stream evaluated by the companion VAD process, which operates on the spatially filtered output rather than a raw microphone channel.

The raw DOA output can fluctuate between consecutive frames due to noise, reverberation, and transitions between the two estimation algorithms. A stability filter operating on a circular buffer of the 15 most recent angle estimates suppresses this variability. The filter computes a circular mean and circular standard deviation using the unit-vector method, which correctly handles the discontinuity at 0°/360°. If the circular standard deviation exceeds 90°, the current output is suppressed entirely, as the recent history indicates that the estimate is too unreliable to present to the user. If a new estimate deviates from the circular mean by more than 29° while the history is otherwise stable, the filter snaps the output to the circular mean rather than following the outlier. Only estimates that pass both criteria are forwarded to the display as the final output angle.

In the current prototype, VAD runs as a separate process on a companion device rather than on the ESP32-S3, operating in parallel with the embedded pipeline. Full integration of VAD into the on-device pipeline is planned for future work.

C. Visualization of Localization Output

The final stage of the N1 pipeline translates the numerical DOA output into visual cues rendered in the user's field of view through the glasses' display. The objective is to communicate

the direction of a detected sound source intuitively without diverting the user's attention from their environment. The display updates every processing frame, producing a new visual output approximately every 32 ms.

The current prototype renders directional cues on a Jade Bird Display microLED projector driven via HDMI, with output rendered in software and transmitted to the projector in real time. The display occupies a small portion of the user's visual field at the edge of vision, consistent with the projector's 30° diagonal field of view. The waveguide maintains optical transparency, allowing the user to see the physical environment normally while directional cues are overlaid within the peripheral visual field.

The directional cue is rendered as an arc segment that rotates around a central point in correspondence with the estimated sound direction. The arc spans approximately 72° and is drawn at a radius of 300 pixels from the display center, fading toward transparency at both endpoints to produce a smooth, curved indicator. As the DOA estimate changes, the arc repositions accordingly, giving the user a compass-like reference for where the sound originates. If no valid estimate is produced within 250 ms, the arc is suppressed until a new estimate is available. When the companion VAD process detects speech in the beamformed audio stream, a text label reading 'Speech' is rendered on the display adjacent to the directional arc. Because VAD currently runs on a separate device rather than on the ESP32-S3, the speech indicator operates with slightly higher latency than the directional cue and is not yet fully integrated into the embedded pipeline.

The visual design was informed by a survey administered to 68 participants across English and Spanish versions, including individuals with single-sided hearing loss, bilateral hearing loss, and deaf or hard-of-hearing identity. Participants were asked to rate the usefulness of the concept, identify preferred visual cue formats, and specify preferred display placement. Directional arcs were among the most preferred visual cue formats, selected by 27 of 68 respondents, alongside arrows (26). The majority of respondents preferred cues positioned at the edge of vision, which informed the current display placement within the peripheral visual field. Further structured evaluation under realistic conditions is a direction for future work. Fig. 8 shows the waveguide display output as seen by the user, with the directional arc indicating a sound source behind the wearer and a speech label confirming detected voice activity.

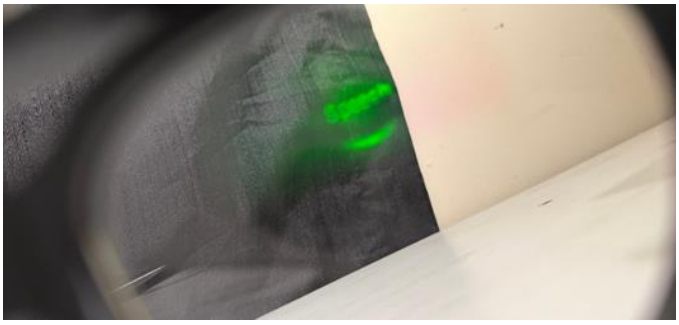


Fig. 8. Waveguide display output showing directional cue and speech indicator.

IV. Results & Analysis

A. Data Collection

A prerecorded male speech sample was played from a loudspeaker positioned 0.5 m from the N1 glasses at 60 dB (measured at the glasses). The speaker was moved around the glasses in 45-degree increments, and the angle output by the N1 glasses was recorded for three repeated trials at each position. For each trial, the angle reported by the N1 glasses was compared to the known speaker angle, the ground truth. Table 2 summarizes the results, with averages computed as the mean of absolute error values.

Table 2. **Difference between estimated angle and true angle of sound stimulus**

True Angle(°)	Trial 1 (°)	Trial 2 (°)	Trial 3 (°)	Average (°)
0	0.7	9.1	-1.3	3.7
45	-5.4	3.2	-6.7	5.1
90	3.5	0.3	2	1.9
135	2.3	1.6	4.7	2.9
180	-9.4	-0.8	0.3	3.5
225	1.6	-5.5	9.2	5.4
270	2.8	-8.5	7.8	6.4
315	1.6	-1.4	-2.3	1.8

Then, the root mean square error was calculated with the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n=24}(\theta_{estimated} - \theta_{true})^2}$$

where $\theta_{estimated}$ is the angle in degrees reported by the N1 glasses, θ_{true} is the known angle of the speaker, and n is the total number of trials. For each of the 8 positions tested across 3 trials, the difference between the estimated and true angle was squared, averaged across all 24 measurements (n = 24), and square-rooted to yield a single error value in degrees: 4.9°.

B. Comparative Performance Analysis

The system's performance was evaluated against the biological benchmark of human auditory localization. The N1 system's RMSE of 4.9° falls within the 2.8° to 7.3° range reported for normal-hearing listeners [16], indicating that the microphone array and DSP pipeline achieve localization accuracy comparable to biological spatial hearing. Both benchmarks were measured under controlled, low-noise conditions; performance of both human listeners and the N1 system would likely degrade in reverberant or noisy environments.

C. Environmental Robustness and Limitations

Although initial testing was conducted in a controlled environment, the integration of audio beamforming allowed

the system to filter meaningful acoustic information from background noise. Elevated error was observed at 225° and 270°, with average deviations of 5.4° and 6.4° respectively. As described in Section III-B, this reflects the reduced angular resolution of GCC-PHAT at the boundaries of the hybrid switching zones. Variance across trials at individual angles was also notable; at 0°, errors ranged from 0.7° to 9.1°, and at 270° from -8.5° to 7.8°, suggesting that frame-to-frame instability is not fully resolved by the stability filter, contributing to within-angle variability in addition to the systematic angular effects. Real-world conditions including reverberation, background noise, multiple simultaneous sources, and moving sources would likely degrade both GCC-PHAT accuracy and beamformer performance [15]; future testing should evaluate performance with multiple simultaneous sound sources and ambient background noise. The 32 ms processing frame described in Section III-B yields an estimated end-to-end latency from sound onset to display update of under 100 ms under typical operating conditions. Although hardware and PCBs underwent multiple iterations, the current hardware has not achieved the size and ergonomic profile required for practical daily use, which remains an engineering challenge for future iterations.

V. Conclusion and Discussion

This project demonstrates that augmented reality glasses with an embedded microphone array and DSP pipeline can achieve sound localization accuracy comparable to human spatial hearing. With an RMSE of 4.9°, the N1 system falls within the 2.8° to 7.3° range observed in normal-hearing listeners [16], suggesting that real-time sound visualization is a feasible approach to restoring spatial awareness for DHH individuals. AR provides an intuitive means of presenting this information, overlaying directional cues onto the user's field of view without requiring the user to attend to a separate device. Recent work has shown that AR-based sound displays can improve comfort and quality of life for DHH users [17], and demand for such technologies remains high; 73.1% of DHH survey participants reported being very or extremely interested in greater access to environmental sounds even while using existing assistive devices [18]. User preferences vary across visual and haptic feedback modalities [6], underscoring the importance of flexible, user-centered design in future systems.

At a bill of materials cost of approximately $297, the N1 prototype demonstrates that wearable sound localization hardware can be developed at a fraction of the cost of conventional hearing aids or cochlear implants. The arc-based directional cue design was informed by a survey of 68 participants, including DHH individuals, whose preferences for display placement and visual cue format shaped the current interaction architecture. Limitations of the current prototype include single-source controlled testing conditions, a hardware form factor not yet suitable for daily wear, and VAD operating as a separate companion process rather than on the embedded platform. As AR hardware continues to become more accessible and embedded DSP capabilities improve, wearable sound visualization systems have the potential to meaningfully enhance situational awareness and independence for DHH individuals.

Acknowledgment

This work was awarded two PTSO Grants by Carter G. Woodson High School; won first place, Grand Prize Nominee, and Grand Prize Runner-up at the Fairfax County Regional Science Fair; received honorable mention at the Virginia State Science and Engineering Fair. Special thanks to Dr. Douglas Oberle for his mentorship throughout the research.

References

[1] S. Chadha, H. Kamenov, and A. Cieza, "The world report on hearing, 2021," *The Lancet*, vol. 397, no. 10278, pp. 1842–1843, May 2021, doi: 10.1016/S0140-6736(21)00517-X.

[2] F. Arosio, L. Brayda, and R. Actis-Grosso, "Auditory localization: a comprehensive practical review," *Frontiers in Psychology*, vol. 15, Art. no. 1408073, Jul. 2024, doi: 10.3389/fpsyg.2024.1408073.

[3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, rev. ed. Cambridge, MA, USA: MIT Press, 1997.

[4] T. Van den Bogaert, T. J. Klasen, J. Wouters, S. Doclo, and M. Moonen, "Horizontal localization with bilateral hearing aids: Without is better than with," *Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 515–526, Jan. 2006, doi: 10.1121/1.2139653.

[5] S. Ausili et al., "Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: A review," *American Journal of Audiology*, vol. 31, no. 3, pp. 819–834, Sep. 2022, doi: 10.1044/2022_AJA-22-00006.

[6] M. D. Fletcher et al., "Haptic sound-localisation for use in cochlear implant and hearing-aid users," *Scientific Reports*, vol. 10, Art. no. 14171, Aug. 2020, doi: 10.1038/s41598-020-70379-2.

[7] National Institute on Deafness and Other Communication Disorders, "Quick statistics about hearing," NIDCD, 2024. [Online]. Available: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing

[8] X. Zhang et al., "Design of small MEMS microphone array systems for direction finding of outdoor moving vehicles," *Sensors*, vol. 14, no. 3, pp. 4384–4398, Mar. 2014, doi: 10.3390/s140304384.

[9] C. L. Nieman and D. L. Bhatt, "Hearing aid affordability in the United States," *Disability and Rehabilitation: Assistive Technology*, vol. 17, no. 3, pp. 246–252, 2022, doi: 10.1080/17483107.2020.1822449.

[10] R. A. Bartholomew, V. K. Rathi, K. Suresh, R. K. V. Sethi, D. J. Lee, and R. Xiao, "Hospital-negotiated pricing of cochlear implants," *Otolaryngology–Head and Neck Surgery*, vol. 169, no. 3, pp. 734–737, Sep. 2023, doi: 10.1002/ohn.280.

[11] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976, doi: 10.1109/TASSP.1976.1162830.

[12] K. W. Wilson and T. Darrell, "Learning a precedence effect-like weighting function for the generalized cross-correlation framework," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2156–2164, Nov. 2006, doi: 10.1109/TASL.2006.872601.

[13] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Dept. Div. Eng., Brown Univ., Providence, RI, USA, 2000.

[14] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988, doi: 10.1109/53.665.

[15] J.-M. Yang, C.-H. Lee, S. Kim, and H.-G. Kang, ”A robust time difference of arrival estimator in reverberant environments,” in *Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009)*, Glasgow, Scotland, Aug. 24–28, 2009, pp. 864–868.

[16] S. Kerber and B. U. Seeber, ”Sound localization in noise by normal-hearing listeners and cochlear implant users,” *Ear and Hearing*, vol. 33, no. 4, pp. 445–457, Jul.–Aug. 2012, doi: 10.1097/AUD.0b013e318257607b.

[17] T. Asakura, ”Augmented-reality presentation of household sounds for deaf and hard-of-hearing people,” *Sensors*, vol. 23, no. 17, Art. no. 7616, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/17/7616

[18] L. Findlater, B. Chinh, D. Jain, J. Froehlich, R. Kushalnagar, and A. C. Lin, ”Deaf and hard-of-hearing individuals’ preferences for wearable and mobile sound awareness technologies,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland, May 2019, pp. 1–13, doi: 10.1145/3290605.3300276.