İSTANBUL UNIVERSITY CERRAHPAŞA

ORIGINAL ARTICLE
Electrica 2025; 25: 1-9

A Novel Acoustic Source Localization Technique for Edge AI Applications: A Lightweight Framework and Implementation for IoT and Smart Sensing Devices

Serhan Yarkan

ISTEC Cybersecurity Inc., R&D Center, İstanbul, Türkiye; Department of Computer Engineering, Fenerbahçe University Faculty of Engineering, İstanbul, Türkiye

Cite this article as: S. Yarkan, "A novel acoustic source localization technique for edge AI applications: A lightweight framework and implementation for IoT and smart sensing devices," *Electrica*, 25, 0099, 2025. doi: 10.5152/electrica.2025.25099.

WHAT IS ALREADY KNOWN ON THIS TOPIC?

- *Recent studies have aimed to adapt direction-of-arrival estimation for edge devices, but many techniques remain unsuitable for microcontrollers due to their reliance on iterative algorithms and floating-point operations.*
- *Classical acoustic direction-finding methods achieve high accuracy but require dense microphone arrays, precise synchronization, and significant computational resources.*
- *Minimalist microphone configurations have shown potential for acoustic localization; however, they often compromise on processing speed or require environment-specific calibration.*

Corresponding author:
Serhan Yarkan

E-mail:
syarkan@ieee.org, serhan.yarkan@fbu.edu.tr

Received: May 22, 2025
Revision Requested: May 31, 2025
Last Revision Received: June 1, 2025
Accepted: June 4, 2025
Publication Date: August 19, 2025

DOI: 10.5152/electrica.2025.25099

ABSTRACT

This paper presents a novel and computationally efficient three-point signal estimation method for acoustic direction finding, designed specifically for low-cost embedded platforms. The proposed approach offers a lightweight alternative to traditional cross-correlation techniques by minimizing computational complexity while preserving high angular resolution. The method was implemented and tested on an STM32F429 microcontroller using a pair of MAX4466 electret microphones arranged on a fixed baseline. The system architecture leverages bare-metal signal processing routines optimized with Acorn RISC Machine Cortex. Microcontroller Software Interface Standard (ARM CMSIS-DSP) libraries, enabling real-time performance on resource-constrained hardware. Extensive experiments were conducted to evaluate the angular estimation accuracy under varying signal-to-noise ratios and source orientations. Results show that the system maintains sub-degree mean square error for source angles up to 30°, with noticeable performance degradation observed at 40° due to the directional response characteristics of the microphone elements. A comprehensive explanation is provided linking this degradation to reduced microphone sensitivity at wider angles of incidence. The proposed solution is ideal for applications requiring embedded acoustic localization, including smart interfaces, vehicular monitoring, and surveillance systems. In addition, the paper discusses the implications of deploying such systems in artificial intelligence (AI)-enabled and security-critical environments, highlighting emerging threats such as adversarial acoustic interference and spoofing attacks. These challenges underscore the importance of resilient and efficient DF methods that can operate reliably within the constraints of embedded systems. The presented work lays the foundation for future research in secure, scalable, and AI-compatible acoustic sensing platforms.

***Index Terms*—**Acoustic, artificial intelligence, direction finding, internet of things

I. INTRODUCTION

Acoustic direction finding (DF) has become an indispensable component across a wide range of civilian and security applications. In the infotainment domain, DF enhances immersive interactions in video conferencing [1] and distance learning [2], enables natural voice control in smart homes and television systems [3, 4], supports customer analytics in retail and advertisement services [5], and augments spatial audio for multi-user gaming [6, 7]. While traditionally employed for infotainment and surveillance, recent advances in artificial intelligence (AI) and growing cybersecurity considerations have significantly expanded its capabilities and applications [8, 9].

In surveillance settings, acoustic DF underpins life-critical systems for driver monitoring—detecting smartphone use [10, 11] and drowsiness [12]—as well as tactical applications such as gunshot and artillery localization [13]. Comprehensive surveys of indoor localization and DF technologies can be found in [9, 14].

Recent advances in AI have dramatically boosted DF performance: convolutional and attention-based networks have been applied to raw multichannel audio for end-to-end angle estimation, yielding robustness to reverberation and diffuse noise [15, 16]. At the same time, the deployment

WHAT THIS STUDY ADDS ON THIS TOPIC?

- *This study introduces a novel three-point signal estimation technique that enables fast and efficient acoustic direction finding without the need for heavy cross-correlation or iterative computations.*
- *The proposed method is successfully deployed on a low-power STM32 microcontroller, showcasing real-time performance with minimal processing overhead on cost-effective hardware.*
- *By combining classical signal processing principles with edge-computing constraints, the study delivers a scalable and secure solution for smart sensing in artificial intelligence–driven and embedded systems.*

of AI-driven DF in security-critical contexts has exposed new vulnerabilities. Adversarial audio attacks can inject imperceptible perturbations that mislead deep models [17], while physics-based signal-injection exploits (e.g., Poltergeist AMpLe) can compromise downstream detection pipelines [18]. Defenses such as SoundFence use physical-layer signatures to detect malicious ultrasonic interference [19]. These AI and cybersecurity considerations underscore the need for resilient, low-cost embedded DF platforms.

The literature review reveals that many acoustic direction finding techniques have been widely explored in applications ranging from robotics to IoT, with classical methods (e.g., time difference of arrival [TDOA], Steered-Response Power (SRP), and subspace-based techniques) offering high accuracy but demanding dense arrays, precise synchronization, and substantial computational resources [20, 21]. Recent efforts have optimized direction-of-arrival (DoA) estimation for edge devices [22, 23], though many still rely on iterative searches or floating-point operations ill-suited for microcontrollers. Minimalist microphone configurations have shown promising performances [24] but they often trade off speed or require environment-specific calibration. In contrast, this study proposes a novel three-point estimation technique that enables rapid, accurate direction finding using only three strategic signal features from a compact microphone array, eliminating the need for cross-correlation over large windows or iterative searches. To the author's knowledge, this is the first implementation of such a model explicitly designed for lightweight, real-time, low-power acoustic localization in IoT and embedded systems. The method is implemented on an STM32F429 microcontroller with two MAX4466 microphones, leveraging bare-metal processing and ARM CMSIS-DSP optimizations to achieve real-time performance. Experimental results demonstrate sub-degree mean square error for angles up to 30°, with analysis of performance degradation beyond this range providing insights for future hardware improvements. By bridging classical array processing and edge-computing constraints, this work advances practical, resource-efficient smart sensing systems. The key contributions are as follows:

- A novel three-point signal estimation method is proposed and implemented for lightweight acoustic direction finding, offering a computationally efficient alternative to traditional cross-correlation-based techniques.
- The method is validated on a real-time embedded system, demonstrating its low processing overhead and ease of deployment on cost-effective microcontroller platforms.
- The adaptive structure of the method enhances its resilience and security, supporting the development of robust direction-finding solutions for AI-driven and security-critical applications.

The rest of the paper is organized as follows. Section II outlines the signal and system model for the problem of interest. Section III discusses the implementation of the proposed method on an embedded system as well as the measurement campaign where the implementation is used to collect data. Section IV gives the results. Finally, Section V concludes the paper.

II. SIGNAL AND SYSTEM MODEL

A linear array with M acoustic sensors (i.e., microphones) is assumed to present with equally spaced intervals, say d located on a plane to detect the direction of K acoustic sources. Without loss of generality, let $M=2$ and correspondingly $K=1$, and both source and sensors are coplanar [1]. The single source, say S, is assumed to emit acoustic energy which could spread around a carrier in the far-field of the sensors. The geometric configuration illustrated in Fig. 1 depicts a simplified acoustic direction-finding scenario using a two-element planar microphone array. The microphones, labeled as m_L and m_R, are placed symmetrically on a horizontal sensor plane with a baseline spacing denoted by d. An acoustic source S emits wavefronts that propagate toward the array at an azimuthal angle θ with respect to the array's broadside (normal). Under the far-field assumption, the incident acoustic wavefronts can be modeled as planar and equidistant, arriving at each sensor with a predictable time offset. In this configuration, the wavefront first reaches the microphone that is closer to the source direction-in this case, m_R—resulting in a TDOA between m_L and m_R. This TDOA corresponds to a physical path difference $\Delta d = d\sin(\theta)$, as indicated by the brace annotation in the figure. The planar wavefronts are visualized with dashed lines intersecting both microphones and the direction of wave propagation is marked with curved arrows. A perpendicular reference line (normal) is also drawn from the center of the sensor plane to highlight the angle θ between the incident wavefront and the array axis. This geometry provides the analytical basis for calculating the angle of arrival (AoA) from measured TDOA values and is fundamental to time-domain acoustic localization methods.

Based on the discussion above, under the narrowband assumption, the acoustic pressure that is perceived by each and every sensor could then be written as in (1):

$$S(t)=Re\left\{p(t)e^{J2\pi f_c t}\right\} \tag{1}$$

where $J=\sqrt{-1}$; $p(t)$ denotes the complex baseband equivalent of the transmitted signal; f_c is the carrier frequency; and $\mathbb{R}e\{\cdot\}$ denotes the real part of its input. A representative sketch of the problem is given in Fig. 1.

Any source location that is not on the normal line leads to a delay in the detected acoustic pressure for one of the sensors due to the geometry of the problem. It is known that acoustic waves traverse the propagation medium with a finite speed which can be characterized by (2):

$$\nu=331.3+(0.606)\times\kappa \tag{2}$$

where κ is the ambient temperature in degrees Celsius under the assumption of no humidity assumption. Furthermore, the carrier frequency, f_c, imposes certain restrictions on the identifiable spatial regions due to the wavelength-sensor interval relationship. As stated before, considering the narrowband assumption, d is chosen to satisfy $d=\frac{1}{2}\frac{\nu}{f_c}$ in order to avoid the spatial aliasing problem around the normal in Fig. 1. Furthermore, sensors are assumed to be directional in such a way that only the sources that fall in the upper part of the sensor plane are of interest as indicated in Fig. 1 [2].

A general linear, time-varying complex baseband channel impulse response is given by (3):

$$h(t,\tau)=\sum_{k=1}^{N}a_k(t)\delta(t-\tau_k(t))e^{j(2\pi f_k^D(t)+\theta_k(t))} \tag{3}$$

where N is the number of resolvable multipath components; $a_k(\cdot)$ is the magnitude of the k-th component; $\delta(\cdot)$ is the Dirac delta function; $\tau_k(\cdot)$ is the k-th delay; $f_k^D(\cdot)$ and $\theta_k(\cdot)$ are the Doppler shift and phase offset experienced by the corresponding k-th multipath component, respectively, and $j=\sqrt{-1}$. Depending on the scenario, a further extension is possible by considering (4):

$$h(t,\tau)=\sum_{k=1}^{N}a_k(t)\delta(t-\tau_k(t))e^{j(2\pi f_k^D(t)+\theta_k(t))} \tag{4}$$

as a major contribution of the power is obtained by the specular component, $s_k(t)$, along with a relatively weaker contribution of a diffused component, $d_k(t)$, where (5, 6) given by:

$$s_k(t)=\sigma_{s_k}e^{j\left(2\pi f_D\cos\left(\theta_0^{(k)}\right)t+\phi_0^{(k)}\right)} \tag{5}$$

and

$$d_k(t)=\sigma_{d_k}\frac{1}{\sqrt{M_k}}\sum_{m=1}^{M_k}b_m e^{j\left(2\pi f_D\cos\left(\theta_m^{(k)}\right)t+\phi_m^{(k)}\right)} \tag{6}$$

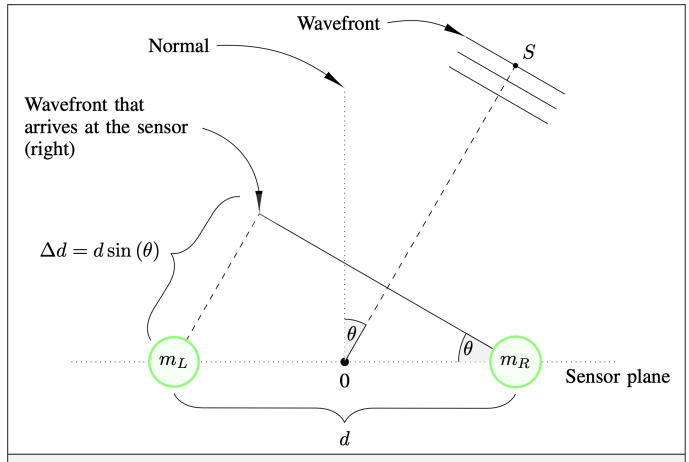

Fig. 1. Geometry of a simple acoustic direction finding system.

This way, the receiver, if possible, could be designed based on the perspective given in (4). This representation also paves way for arranging resolvable multipath components via certain parameters according to the communication scenario to be considered for various purposes. As given in (4), σ_{s_k} is the magnitude of the possibly present specular component; f_D is the maximum Doppler frequency shift observed by the receiver. From the statistical point of view, the power of the specular and diffused components could best be observed under the wide-sense stationary assumption by (7):

$$\begin{aligned}R_{h_k}(\Delta\tau) &= E\left\{h_k(t)h_k^*(t+\Delta\tau)\right\}\\ &= R_{s_k}(\Delta\tau)+R_{d_k}(\Delta\tau)\end{aligned} \tag{7}$$

Once the normalized version is considered, then (8):

$$\begin{aligned}\rho_{h_k}(\Delta\tau)&=\frac{R_{h_k}(\Delta\tau)}{R_{h_k}(0)}=\frac{R_{s_k}(\Delta\tau)+R_{d_k}(\Delta\tau)}{\sigma_{s_k}^2+\sigma_{d_k}^2}\\ &=\frac{K_k}{K_k+1}\rho_{s_k}(\Delta\tau)+\frac{1}{K_k+1}\rho_{d_k}(\Delta\tau)\end{aligned} \tag{8}$$

is obtained.

Given $h(t, \tau)$, the received signal at the baseband is generally expressed in the form of (9):

$$r(t)=x(t)\circledast h(t,\tau)+\omega(t) \tag{9}$$

where $x(t)$ is the transmitted signal at the baseband; $\circledast$ denotes the convolution operator; $\omega(t)$ is the complex baseband additive white Gaussian noise term consisting of both in-phase and quadrature components, $\omega_I(t)$ and $\omega_Q(t)$, respectively, where each component is $\mathcal{N}(0,\sigma_\omega^2/2)$. Without loss of generality, channel statistics are assumed to be stationary (in any sense) for a sufficiently long period of time. Furthermore, in order to have a more tractable analysis, the expression as given in (9) is assumed to be of the following form as given in (10):

$$r(t)=x(t)\circledast h(t)+\omega(t) \tag{10}$$

where the channel, $h(t)$, acts as a linear, time-invariant filter.

Considering the received signal then,

$$R_r(\Delta\tau)=R_x(\Delta\tau)\circledast R_h(\Delta\tau)+\sigma_\omega^2\delta(\Delta\tau)$$
$$\rho_r(\Delta\tau)=\frac{1}{\sigma_r^2}\left(R_x(\Delta\tau)\circledast R_h(\Delta\tau)\right)+\frac{\sigma_\omega^2}{\sigma_r^2}\delta(\Delta\tau) \quad (11)$$
$$=K_k\frac{\left(\rho_{s_k}(\Delta\tau)\circledast\rho_x(\Delta\tau)\right)}{\sigma_x^2\sigma_r^2(K_k+1)}+\frac{\left(\rho_{d_k}(\Delta\tau)\circledast\rho_x(\Delta\tau)\right)}{\sigma_x^2\sigma_r^2(K_k+1)}+\frac{\sigma_\omega^2}{\sigma_r^2}\delta(\Delta\tau)$$

could be written. Here, the nature of $x(t)$ and therefore its correlation characteristics, $\rho_x(\cdot)$ are important. In case the human vocal tract is considered to be the acoustic source, several features are prominent such as formant frequencies and temporal decorrelation. In the literature, the temporal decorrelation of an average human speech ranges between 10–100 ms, depending on a vast set of conditions. However, objective measurements reveal that temporal decorrelation is statistically concentrated around 15–35 ms [25]. On the other hand, in case data communications are of interest, temporal decorrelation is driven by the pulse shaping filter. At this point, one could discuss the consequences of the temporal decorrelation of acoustic sources since it is directly related to the operational requirements of the DF process. Considering the source is somehow modeled with a stationary, stable, continuous Langevin model, the temporal decorrelation nature of the source, $\rho_x(\cdot)$, could be expressed when $\Delta\tau\to 0$ as:

$$\rho_x(\Delta\tau)=\xi e^{-\varrho\Delta\tau} \quad (12)$$

When $p_x(\Delta\tau)$ as given in (11) is extended by (12), the following is obtained for $0<\Delta\tau$:

$$\frac{\xi}{\zeta}\left(K_k\rho_{s_k}(\Delta\tau)+\rho_{d_k}(\Delta\tau)\right)\circledast e^{-\varrho\Delta\tau} \quad (13)$$

Note that the convolution operation as given in (13) includes exponential decay due to the stationary, stable, continuous Langevin model. Along with the translation invariant attribute of real exponentials under convolution integral, $p_x(\Delta\tau)$ as given in (13) could be approximated via the closed Newton–Cotes method. First, let $\rho_{h_k^{\text{nom}}}(\cdot)=K_k\rho_{s_k}(\cdot)+\rho_{d_k}(\cdot)$ In case fourth-order approximation (also known as Simpson's 3/8 method) is adopted under the assumption $0\le\Delta_\tau<\Delta T_c$ with ΔT_c being the channel decorrelation duration as in (14):

$$\rho_r(\Delta\tau)=\frac{e^{-\varrho\Delta\tau}\xi\Delta T_c}{8\zeta}\sum_{k=0}^{3}\binom{3}{k}\rho_{h_k^{\text{norm}}}\left(\frac{k\Delta T_c}{3}\right)e^{\varrho\frac{k\Delta T_c}{3}} \quad (14)$$

is obtained. In order to better see both the impact of channel and transmit signal, the expression as given in (13) could be rewritten in light of the expression as given in (12) and as in (15):

$$\rho_r(\Delta\tau)=\frac{\rho_x(\Delta\tau)\Delta T_c}{8\zeta}\sum_{k=0}^{3}\binom{3}{k}\frac{\rho_{h_k^{\text{norm}}}\left(\frac{k\Delta T_c}{3}\right)}{\rho_x\left(\frac{k\Delta T_c}{3}\right)} \quad (15)$$

is obtained. Given the fact that the channel statistics remain the same for two different sensors within a single observation frame, cross-correlation of sensors yields:

$$\rho_{\eta_2}(\Delta\ddot{A})=\frac{\rho_x(\Delta\ddot{A}-\Delta t)\Delta T_c}{8\P}\sum_{k=0}^{3}\binom{3}{k}\frac{\rho_{h_k^{\text{norm}}}\left(\frac{k\Delta T_c}{3}\right)}{\rho_x\left(\frac{k\Delta T_c}{3}\right)} \quad (16)$$

where Δt is the duration for the wavefront to traverse distance sketched in Fig. 1 while satisfying $\Delta t=\frac{\Delta d}{\gamma}$ in conjunction with the expression as given in (2). As given in (12), in case $x(t)$ conveys data (being either analog or digital), it could be concluded that decorrelation time of $\rho_x(\cdot)$ is less than ΔT_c. Thus, the correlation ratio as given in (15) diminishes faster as $k\to 3$. Consider then the expanded version of the expression as given in (16). Then:

$$\rho_r(\Delta\ddot{A})=\frac{\rho_x(\Delta\tau)\Delta T_c}{8^3\P}\left(1+3\frac{\rho_k^{\text{nom}}\left(\frac{\Delta T_c}{3}\right)}{\rho_x\left(\frac{\Delta T_c}{3}\right)}+3\frac{\rho_{h_k^{\text{norm}}}\left(\frac{2\Delta T_c}{3}\right)}{\rho_x\left(\frac{2\Delta T_c}{3}\right)}+\frac{\rho_{h_k^{\text{norm}}}\left(\frac{3\Delta T_c}{3}\right)}{\rho_x\left(\frac{3\Delta T_c}{3}\right)}\right) \quad (17)$$

is obtained. Now, in case $\rho_{h_k^{\text{norm}}}$ is a sum of exponentials (as in the specular and diffused components each having exponential correlation), for instance in the form of:

$$\rho_{h_k^{\text{norm}}}(\Delta\tau)=K_k e^{-\alpha\tau}+e^{-\beta\tau} \quad (18)$$

then after re-expressing the equation in terms of known parameters, the expression as given in (17) could be simplified and:

$$\rho_r(\Delta\tau)=\frac{\xi\Delta T_c e^{-\varrho\Delta\tau}}{8\zeta}\left[K_k\left(1+e^{\frac{(\varrho-\alpha)\Delta T_c}{3}}\right)^3+\left(1+e^{\frac{(\varrho-\beta)\Delta T_c}{3}}\right)^3\right] \quad (19)$$

is obtained. From the expression as given in (17) it is clear that the correlation characteristics could be captured satisfactorily by the first three samples (excluding the zeroth index), namely the samples obtained at $a\Delta T_c$ where $a\in\{1,2,3\}$. Furthermore, from the expression as given in (19), under specific circumstances, the correlation characteristics could solely be captured with the aid of a cubic polynomial in terms of the evaluations at four points located at $a\Delta T_c$ where $a\in\{0,1,2,3\}$.

Here, one might consider why the correlation model as given in (18) is chosen in the first place. The form as given in (18) is physically motivated for acoustic signals as it separates coherent specular reflections (decaying at rate α) from incoherent diffuse energy (rate β). This is a common model in acoustic propagation physics since direct/early reflections exhibit exponential decay due to absorption and spherical spreading, while late reverberation from dense multipath follows a similar decay law in statistically isotropic environments. Therefore, the model balances fidelity and tractability for far-field narrowband signals.

Another interesting point regarding the expression as given in (18) might be the alternative forms from the perspective of acoustic propagation. As given in (18), the expression represents the normalized channel autocorrelation as a sum of two exponentials. However, there are different alternatives corresponding to different peculiar acoustic propagation scenarios. For instance, when signal reception takes place in enclosed spaces where resonant reflections should be taken into account, then $\rho_{h_k^{\text{norm}}}(\Delta\ddot{A})=e^{-\gamma\ddot{A}}\cos(2\pi f_r\tau)$ would be a natural choice with decay rate γ and fundamental resonant frequency f_γ. On the other hand, in case a highly reverberant environment (e.g., concert hall) is considered, then a slow energy decay model $\rho_{h_k^{\text{norm}}}(\Delta\tau)=\left(1+\frac{\tau}{\tau_0}\right)$ with decay rate η could be a more realistic description.

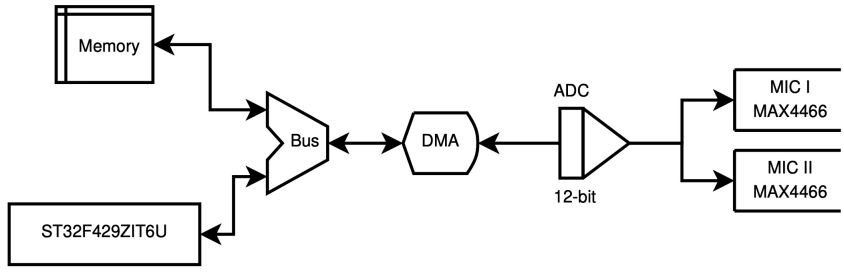

Fig. 2. System used in real-world measurements.

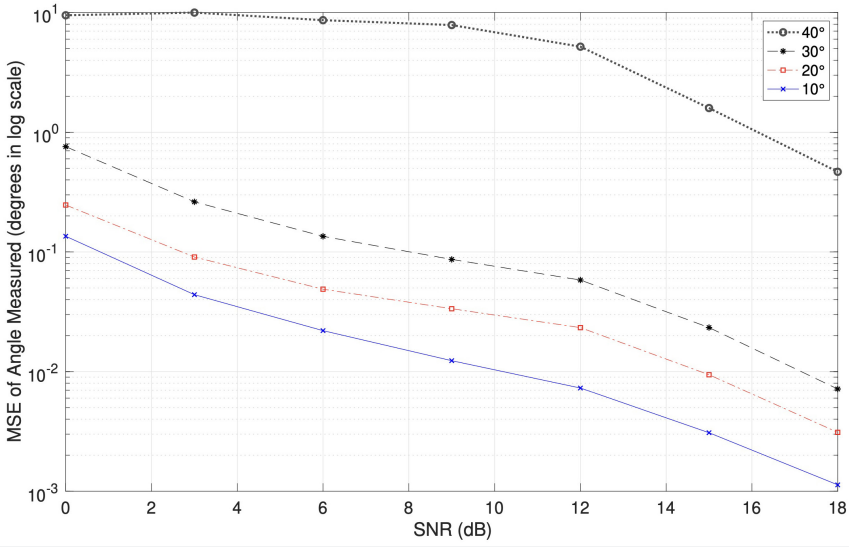

Fig. 3. Measurement results obtained for the normal angles 10–40 degrees.

III. IMPLEMENTATION AND MEASUREMENT CAMPAIGN

A. The Platform Used in the Implementation

The experimental platform employed in this study for acoustic direction finding was built around a low-cost STM32F429 development board, selected for its balance of computational performance and resource efficiency. This microcontroller unit features a 32-bit ARM Cortex - M 4 processor operating at 180 MHz, equipped with 2 MB of flash memory and a 12-bit analog-to-digital converter (ADC) with a sampling rate of 2.4 mega-samples per second (MSPS). These hardware specifications were deemed sufficient for implementing real-time digital signal processing tasks required in acoustic localization. The system utilized multiple MAX4466 electret microphone modules as acoustic sensors, which offer onboard preamplification and adjustable gain. These microphones were selected due to their low noise performance and compatibility with microcontroller-based data acquisition systems, ensuring clean and consistent analog signal capture suitable for direction-finding algorithms.

In order to get a deeper insight into the system used here, specific details might be given in here. The signal acquisition chain was designed with careful attention to analog-digital interface specifications. Each MAX4466 microphone module delivers an output voltage interval of 0.5–2.5V (centered at Vcc/2 = 1.65V for 3.3V supply) with a bandwidth of 20 Hz–20 kHz, aligned with the acoustic range of interest. The analog signals were digitized using the STM32F429's 12-bit ADC configured at a sampling rate of 48 kHz per channel (Nyquist rate = 24 kHz), sufficient to capture acoustic localization features. A 3.3V reference voltage ensured full-scale coverage of the MAX4466's output range.

For real-time data transfer, Direct Memory Access (DMA) was employed in circular buffer mode with dual 512-sample ping-pong buffers (10.67 ms per buffer at 48 kHz). Buffers were triggered by the ADC's end-of-conversion interrupt, minimizing Central Processing Unit (CPU) overhead. This setup achieved continuous streaming while allowing simultaneous processing of one buffer while the other filled—critical for maintaining signal continuity in TDOA calculations. The ADC used 14-cycle resolution (0.58 µs per sample at 180 MHz) with right-aligned 16-bit storage, balancing timing precision and memory efficiency.

The signal processing algorithms were developed in ANSI C for high portability and were built upon the ARM CMSIS-DSP library to take full advantage of the available hardware acceleration. The software was executed in a bare-metal environment without an operating system, thus minimizing overhead and enabling deterministic timing behavior—crucial for time-sensitive audio signal acquisition and processing. The modular architecture of the codebase also permits straightforward migration to an operating system–based platform if needed. The final firmware, including all drivers and the hardware abstraction layer, compiled into an executable of approximately 500

KB. Notably, the same software stack was successfully ported to a lower-end 32-bit Cortex M0 platform, further validating the design's flexibility and computational efficiency. This embedded platform, combined with the low-cost MAX4466 microphone array, thus provides a robust and scalable foundation for real-time acoustic direction-finding applications. The block diagram of the system used in this study is given in Fig. 2.

B. Measurement Campaign and Data Collection

In the experimental setup, the acoustic direction-finding system was deployed in a controlled environment, using two spatially separated MAX4466 electret microphone modules, denoted as m_L and m_R, positioned on a fixed horizontal baseline. These microphones were mounted on a custom-built wooden platform designed to ensure mechanical stability and consistent spatial alignment. The platform consists of a rectangular wooden panel with two precisely drilled holes, spaced 17.5 cm apart, in which the microphone modules were embedded and securely fixed using rubber grommets to isolate mechanical vibrations. The microphones were oriented to face the same direction and positioned at the same vertical height, forming a symmetrical two-microphone array. The acoustic source *S*, representing a broadband impulsive or tonal emitter, was placed at variable azimuthal angles and distances in front of the array, allowing for the measurement of TDOA between the microphones.

The analog outputs from m_L and m_R were directly connected to the dual-channel analog inputs of an STM32F429-based development board, featuring a 12-bit, 2.4 MSPS ADC and a 32-bit ARM Cortex-M4 processor running at 180 MHz. This board served as the central processing unit for digitizing and analyzing the acoustic waveforms in real-time. All signal processing algorithms, including filtering, cross-correlation, and angle of arrival (AoA) estimation, were implemented in ANSI C using the ARM CMSIS-DSP library and executed in a bare-metal environment to ensure deterministic timing behavior. The system was powered via Universal Serial Bus (USB) and enclosed in a lightweight 3D-printed housing for portability and repeatability of the measurements. The measurement data were later transferred to a computer for further analysis and verification. This compact yet robust setup enabled accurate AoA estimation in various test scenarios, validating the effectiveness of the embedded TDOA-based direction-finding system as illustrated by the geometry in the accompanying diagram.

To obtain absolute acoustic power levels in decibels sound pressure level (dB SPL) and compute signal-to-noise ratio (SNR) for the MAX4466-based array, a two-step calibration and measurement procedure was implemented using Audacity as the digital audio workstation.

1) Microphone Calibration: One MAX4466 electret microphone module was mounted in a precision acoustic calibrator producing a 1 kHz tone at 94 dB SPL. The microphone output was recorded in Audacity at the same sampling rate used by the STM32F429 board (2.4 MSPS). Audacity's Amplitude Statistics tool was used to measure the Root-Mean-Square (RMS) level $L_{RMS,FS}$ in decibels relative to full scale (dB_{FS}). The calibration constant *C* (in dB) was then computed as

$$C = L_{\text{cal}} - L_{\text{RMS,FS}} \left(L_{\text{cal}} = 94\,\text{dBSPL} \right)$$

2) Signal and Noise Measurement: Under identical gain and environmental conditions, separate recordings of the signal-plus-noise and the ambient noise floor were made. Audacity was again used to extract the RMS levels $L_{\text{sig+noise,FS}}$ and $L_{\text{noise,FS}}$ (both in dB_{FS}). These were converted to dB SPL via

$$L_{\text{sig+noise}} = L_{\text{sig+noise,FS}} + C, L_{\text{noise}} = L_{\text{noise,FS}} + C.$$

The true signal level L_{sig} (in dB SPL) was then obtained by subtracting the noise contribution in the power domain:

$$L_{\text{sig}} = 10\log_{10}\left(10^{L_{\text{sig+noise}}/10} - 10^{L_{\text{noise}}/10}\right),$$

and the resulting SNR (in dB) was computed as

$$\text{SNR} = L_{\text{sig}} - L_{\text{noise}}.$$

IV. RESULTS

Measurement results for the campaign and the proposed method are given in Fig. 3. As can be seen from the figure, across all tested source angles, the measurement results exhibit a clear trend: as SNR increases from 0 dB to 18 dB, the mean square error (MSE) of the estimated arrival angle is observed to decrease consistently. This monotonic improvement is explained by the enhanced ability of the system to resolve the time difference of arrival between the two MAX4466 microphone elements as noise diminishes. Even at the lowest SNR levels, the MSE—plotted on a logarithmic scale—remains finite, indicating the robustness of the embedded cross-correlation algorithm. As the SNR climbs, each performance curve shifts downward, confirming that higher-quality acoustic inputs result in more accurate angle-of-arrival estimates.

A closer examination of individual source angles reveals a distinct pattern: for 10°, 20°, and 30° settings, the MSE curves overlap closely, and the maximum angular error is never observed to exceed approximately 0.5° (on a linear scale), even at the lowest SNR. This consistency is explained by the fact that within this angular range, the MAX4466 microphones maintain sufficient sensitivity and sub-sample time delays are reliably extracted by the STM32F429-based processing pipeline. However, at 40°, the performance curves diverge sharply under low-SNR conditions, with the MSE rising well above the levels observed at smaller angles. This degradation is attributed to the directivity pattern of the MAX4466: as the incident wavefront deviates further from the array's broadside, microphone sensitivity falls off, producing weaker signals and poorer time-delay estimates when noise dominates. Consequently, at larger off-axis angles such as 40°, the combination of reduced sensor response and low SNR leads to markedly worse localization accuracy.

V. CONCLUSION

In this study, a novel three-point signal estimation technique was proposed and implemented for acoustic direction finding on a lightweight embedded platform. The method provides a computationally efficient alternative to conventional full-scale cross-correlation-based approaches, significantly reducing the processing burden while maintaining high estimation accuracy. Experimental evaluations using a MAX4466 microphone array and STM32F429 microcontroller demonstrated the system's effectiveness across varying SNR regimes and angular configurations. In particular, the system maintained sub-degree mean square error performance for source angles up to 30°, with degradation at 40° explained by the directional response limitations of the microphone hardware.

One of the critical aspects of the model proposed here is the impact of various acoustic propagation characteristics on the performance. Therefore, it is important to discuss the effects of such characteristics on the performance of the proposed model. While the model as given in (18) treats the channel as a sum of two exponentials, alternatives like a Rician-weighted form or power-law decay may better capture specific acoustic environments. For example, in highly reverberant environments (e.g., indoor settings), a power-law decay $A \propto \tau^{-\cdot}$ could replace the expression as given in (18), extending the effective correlation window and improving TDOA robustness at the cost of increased computational complexity. Conversely, a damped oscillator model introduces frequency-dependent correlation oscillations, which may degrade AoA estimation if source spectra exhibit narrowband features near f_r. Validating these alternatives against real-world measurements (e.g., anechoic vs. reverberant chambers) is critical to investigate their impact on system performance.

The proposed approach proves highly suitable for deployment in resource-constrained acoustic sensing applications, where real-time performance, energy efficiency, and system simplicity are critical. Furthermore, the integration of this embedded DF solution into broader AI-enabled or cybersecurity-sensitive platforms opens new possibilities for robust, low-cost acoustic perception in domains ranging from smart home interfaces to vehicular safety and tactical surveillance. Future work will focus on extending this system to multi-microphone arrays, integrating AI-driven anomaly detection, and investigating security hardening techniques to defend against adversarial audio threats in mission-critical environments.

While this study focuses on core localization performance, future work must address emerging cybersecurity threats to acoustic DF systems deployed in adversarial settings. Key threat vectors include spoofing attacks, where malicious actors inject deceptive audio (e.g., pre-recorded or synthesized signals) to distort TDOA estimates and induce false AoA readings or adversarial noise, such as low-power ultrasonic interference designed to disrupt correlation-based estimation by exploiting hardware nonlinearities or algorithmic blind spots.

To mitigate these risks, two lightweight defensive strategies suitable for embedded platforms could be proposed: (i) Signal Authentication: Embedding known acoustic preambles (e.g., chirp sequences or spread-spectrum codes) to verify signal legitimacy before processing. This leverages the STM32F429's real-time correlation capabilities to reject untrusted inputs with minimal overhead. (ii) Statistical Anomaly Detection: Monitoring temporal/spatial consistency of AoA estimates (e.g., via Huber loss or clustering) to flag outliers induced by adversarial manipulation. Given the system's low computational footprint (<5% CPU load for baseline processing), such checks could be implemented without hardware upgrades.

These countermeasures would provide enhancements to the model proposed here without any structural changes in the design philosophy of efficiency and resilience, ensuring the framework remains viable for security-critical applications like surveillance or access control. Future validation will quantify robustness against physical-layer attacks like Poltergeist and adversarial perturbations.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – S.Y., ; Design – S.Y.; Supervision – S.Y.; Resources – S.Y.; Materials – S.Y.; Data Collection and/or Processing – S.Y.; Analysis and/or Interpretation – S.Y.; Literature Search – S.Y.; Writing – S.Y.; Critical Review – S.Y.

Declaration of Interests: The author has no conflict of interest to declare.

Funding: The author declared that this study has received no financial support.

REFERENCES

1. M. Bekrani, A. W. H. Khong, and M. Lotfizad, "A clipping-based selective-tap adaptive filtering approach to stereophonic acoustic echo cancellation," *IEEE Trans. Aud. Speech Lang. Process.*, vol. 19, no. 6, pp. 1826–1836, 2011. [CrossRef]
2. Y. A. Huang, J. Benesty, and G. W. Elko, *Microphone Arrays for Video Camera Steering*. USA: Kluwer Academic Publishers Publishers, 2000, pp. 239–259.
3. S. Mun, S. Shon, W. Kim, and H. Ko, "Robust speaker direction estimation with microphone array using nmf for smart TV interaction," in IEEE International Conference on Consumer Electronics (ICCE), 2015, 2015, pp. 112–113.
4. H. Lee, T. H. Kim, J. W. Choi, and S. Choi, "Chirp signal-based aerial acoustic communication for smart devices," in IEEE Conference on Computer Communications, 2015, pp. 2407–2415. [CrossRef]
5. W. Huang et al., "Walkielokie: Sensing relative positions of surrounding presenters by acoustic signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16. New York, NY, USA, 2016, pp. 439–450. [CrossRef]
6. R. Shea et al., "Location-based augmented reality with pervasive smartphone sensors: Inside and beyond pokemon Go!," *IEEE Access*, vol. 5, pp. 9619–9631, 2017. [CrossRef]
7. M. Beig, B. Kapralos, K. Collins, and P. Mirza-Babaei, "An introduction to spatial sound rendering in virtual environments and games," *Comput. Games J.*, vol. 8, no. 3–4, pp. 199–214, 2019. [CrossRef]
8. K. Li, R. Zhang, B. Liang, F. Guimbretière, and C. Zhang, "Eario: A low-power acoustic sensing earable for continuously tracking detailed facial movements," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 1–24, 2022. [CrossRef]
9. Y. Bai, L. Lu, J. Cheng, J. Liu, Y. Chen, and J. Yu, "Acoustic-based sensing and applications: A survey," *Comput. Netw.*, vol. 181, 2020. [CrossRef], p. 107447, 2020.
10. J. Yang et al., "Detecting driver phone use leveraging car speakers," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*. New York, NY, USA: ACM, 2011, pp. 97–108. [CrossRef]
11. C. Bo, X. Jian, X.-Y. Li, X. Mao, Y. Wang, and F. Li, "You're driving and texting: Detecting drivers using personal smart phones by leveraging inertial sensors," in *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*. New York, New York, USA: ACM Press, 2013, pp. 199–202. [CrossRef]
12. Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "D3-guard: Acoustic-based drowsy driving detection using smartphones," in New York: IEEE Infocom, 2019 - IEEE Conference on Computer Communications, 2019, pp. 1225–1233.
13. A. Dagallier et al., "Long-range acoustic localization of artillery shots using distributed synchronous acoustic sensors," *J. Acoust. Soc. Am.*, vol. 146, no. 6, pp. 4860, 2019. [CrossRef]
14. M. Liu, L. Cheng, K. Qian, J. Wang, J. Wang, and Y. Liu, "Indoor acoustic localization: A survey," *Hum. Centric Comput. Inf. Sci.*, vol. 10, no. 1, pp. 1–24, 2020.
15. P.-A. Grumiaux, S. Kitic, L. Girin, and A. Guerin, "A survey of sound source localization with deep learning methods," *J. Acoust. Soc. Am.*, vol. 152, no. 1, pp. 107, 2022. [CrossRef]
16. H. Pujol, É. Bavu, and A. Garcia, *"Beamlearning: an end-to-end deep learning approach for angular localization of sound sources," arXiv Preprint ArXiv:2104.13347*, 2021.
17. T. Du, S. Ji, J. Li, Q. Gu, T. Wang, and R. Beyah, "Sirenattack: Generating adversarial audio for end-to-end acoustic systems," in Oakland Workshop on Machine Learning and Security (MLS), 2019. [CrossRef]
18. J. Ji et al., "Poltergeist: Acoustic adversarial machine learning against object detectors," in *Proc. 2021 IEEE Symposium on Security and*, 2021, pp. -.

19. J. Lou, Q. Yan, Q. Hui, and H. Zeng, *"Soundfence: Securing ultrasonic sensors in vehicles using physical-layer defense," in arXiv Preprint ArXiv:2105.07574*, 2021.
20. M. Brandstein, and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin: Springer, 2001.
21. J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin: Springer, 2008.
22. Y. Zhao, X. Fan, J. Liu, Y. Li, L. Yao, and J. Wang, "Efficient 2D-DOA estimation based on triple attention mechanism for L-shaped array," *Sensors (Basel)*, vol. 25, no. 8, p. 2359, 2025. [CrossRef]
23. B. Sherlock, N. Morozs, J. Neasham, and P. Mitchell, "Ultra-low-cost and ultra-low-power, miniature acoustic modems using multipath tolerant spread-spectrum techniques," *Electronics*, vol. 11, no. 9, p. 1446, 2022. [CrossRef]
24. M. Calis, S. van de Par, R. Heusdens, and R. C. Hendriks, "Localization based on enhanced low frequency interaural level difference," in *IEEE ACM Trans. Aud. Speech Lang. Process.*, vol. 29, pp. 3025–3039, 2021. [CrossRef]
25. K. K. Paliwal, J. G. Lyons, and K. K. Wójcicki, "Preference for 20–40 ms window duration in speech analysis," in 4th International Conference on Signal Processing and Communication Systems, 2010, 2010, pp. 1–4.

Serhan Yarkan received the B.S. and M.Sc. degrees in computer science from Istanbul University, Istanbul, Turkey, in 2001 and 2003, respectively, and the Ph.D. degree from the University of South Florida, Tampa, FL, USA, in 2009. He was a Post-Doctoral Research Associate with the Department of Computer and Electrical Engineering, Texas A&M University, College Station, TX, USA, from 2010 to 2012. He was at Istanbul Commerce University from 2012 to 2025. Since 2023, he has also been affiliated with ISTEC Cybersecurity Inc., R&D Center, Istanbul, Turkey. Serhan Yarkan's current research interests include statistical signal processing, cognitive radio, wireless propagation channel measurement and modeling, cross-layer adaptation and optimization, and interference management in next-generation wireless networks, and underground mine channels and disaster communications and machine learning and artificial intelligence.