

Acoustic DoA Estimation of Gunshots via ILS-Based Data Association in Nearly Simultaneous Multi-Shooter Scenarios

Wesley G. Barboza, José A. Apolinário Jr., Felipe A. C. Bastos, and Antonio L. L. Ramos

Abstract—Acoustic gunshot localization systems rely on robust direction-of-arrival (DoA) estimation from muzzle blast (MB) and shockwave (SW) acoustic signatures. In multi-shooter scenarios, overlapping arrivals produce ambiguous and physically inconsistent time-difference-of-arrival (TDoA) measurements, severely degrading conventional DoA estimation methods. This work formulates the problem as a geometry-constrained data-association task and proposes a computationally efficient processing pipeline for robust DoA estimation under overlapping-shot conditions. Gunshot arrivals are first detected using generalized cross-correlation with phase transform (GCC-PhaT) and organized into a TDoA candidate matrix. A modified iterative least-squares (ILS) procedure then enforces physical consistency by iteratively rejecting delay combinations associated with high residual errors, yielding mutually consistent TDoA subsets linked to individual sources. A refinement stage reconstructs source-consistent times of arrival and prunes the detection matrix, enabling separation of overlapping MB and SW components without prior source labeling. The method is validated using Monte Carlo simulations and real recordings of near-simultaneous multiple shooters. Under ideal conditions, the proposed framework achieves a DoA root-mean-square error (RMSE) of 1.83° and TDoA association accuracy above 98%. With simulated timing noise corresponding to a 2-sample standard deviation, the RMSE increases to 12.99° , while real-data experiments yield 12.59° , consistent with modeled propagation effects. The proposed GCC-PhaT + ILS framework provides an effective trade-off between localization accuracy and computational complexity, making it well suited for real-time embedded deployment.

Index Terms—robust DoA estimation, microphone array processing, gunshot acoustics, multi-shooter detection, sniper localization, GCC-PhaT, iterative least squares (ILS), acoustic source localization

I. INTRODUCTION

The analysis of acoustic gunshot signals—i.e., gunfire sounds recorded by remote audio sensors—is of significant interest in forensic investigation, public safety, and defense applications. A central objective in this context is the estimation of the direction and, when possible, the location of the shot source [1]. Accurate shooter localization can support forensic reconstruction and, in operational scenarios involving *snipers*, may be critical for survival.

W. G. Barboza, J. A. Apolinário Jr., and F. A. C. Bastos are with the Department of Electrical Engineering (SE/3), Military Institute of Engineering (IME), Rio de Janeiro, Brazil. E-mails: wesley.goncalves@ime.eb.br, apolin@ime.eb.br, and fcaetano@ime.eb.br. A. L. L. Ramos is with the Department of Science and Industry Systems, University of South-Eastern Norway (USN), Kongsberg, Norway. E-mail: antonio.ramos@usn.no.

Direction-of-arrival (DoA) estimation of gunshot signals typically involves three main stages: signal enhancement, event detection, and DoA computation. Considerable effort has been devoted to improving the robustness of these stages, particularly in the detection of individual gunshot events [2], [3]. However, real-world scenarios frequently involve nearly simultaneous gunshots, particularly during conflict escalation, resulting in multiple active sources within short time intervals.

Each gunshot produces two characteristic impulsive acoustic signatures: the muzzle blast (MB) and the shockwave (SW), illustrated in Fig. 1. The MB component typically lasts 3–5 ms and propagates from the shooter’s position to the microphones at the speed of sound [4]. The SW component arises from the projectile’s supersonic motion and is observed only when the sensor lies within the projectile’s Mach cone.

Acoustic sniper localization systems commonly rely on DoA estimation from muzzle blast (MB) and shockwave (SW) arrivals as a key step toward shooter localization [5], [6]. Since the proposed data-association framework operates solely on candidate TDoA measurements, it is independent of the underlying acoustic signature and can therefore be applied to either MB or SW arrivals. In this work, however, the framework is validated exclusively using SW arrivals extracted from real gunshot recordings acquired in a challenging multi-shooter scenario.

In multi-shooter scenarios, overlapping MB and SW signatures result in ambiguous and potentially inconsistent TDoA measurements, significantly degrading the performance of conventional DoA estimation methods—even under ideal detection conditions. This challenge can be interpreted as a data-association problem, in which the observed arrival times must be correctly assigned to their respective sources under geometric constraints. Classical approaches such as wideband steered response power beamforming (WB-SRPB) [5], [7] can resolve multiple sources, but typically require dense arrays and entail high computational cost, limiting their applicability in real-time and embedded systems [8].

In this paper, we propose a structured and computationally efficient approach to address this problem. The method organizes detected arrival times into a candidate TDoA set and employs a modified iterative least-squares (ILS) procedure to enforce geometric consistency by iteratively rejecting incompatible delay combinations. This process enables the identification of mutually consistent TDoA subsets associated with individual sources, allowing reliable DoA estimation even

in the presence of overlapping gunshot signatures.

The remainder of this paper is organized as follows. Section II reviews the classical DoA estimation framework. Section III formulates the overlapping-shot problem and presents the proposed method. Section IV reports simulation and experimental results. Finally, Section V concludes the paper.

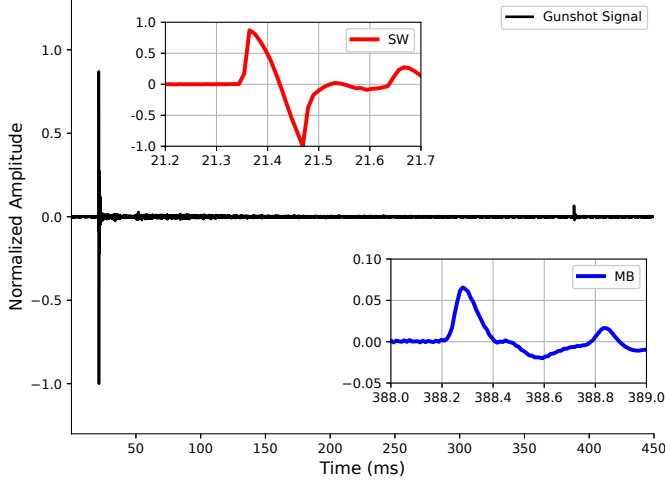


Fig. 1. Typical gunshot signal showing the shockwave (SW) and muzzle blast (MB) acoustic signatures. Magnified views of both components highlight their distinct impulsive characteristics. When observable, the SW typically arrives before the MB because the projectile travels at supersonic speed, whereas the MB propagates at the speed of sound.

II. CLASSICAL DIRECTION-OF-ARRIVAL (DoA) ESTIMATION

Classical DoA estimation relies on time-difference-of-arrival (TDoA) measurements between microphone pairs [3]. In the proposed framework, the TDoA $\hat{\tau}_{ij}$ is estimated using the generalized cross-correlation with phase transform (GCC-PhaT) [9], defined as

$$\hat{r}_{x_i x_j}^G(\tau) = \mathcal{F}^{-1} \left\{ \frac{X_i(e^{j\omega})X_j^*(e^{j\omega})}{|X_i(e^{j\omega})X_j^*(e^{j\omega})|} \right\}, \quad (1)$$

with the delay estimate obtained as

$$\hat{\tau}_{ij} = \arg \max_{\tau} \hat{r}_{x_i x_j}^G(\tau), \quad (2)$$

where the corresponding time delay is $\hat{t}_{ij} = \hat{\tau}_{ij}/f_s$.

Under the far-field assumption, the incident wavefront can be approximated as planar, with its propagation direction represented by the unit vector

$$\mathbf{a}_{\theta, \phi} = [-\sin(\theta) \cos(\phi) \quad -\sin(\theta) \sin(\phi) \quad -\cos(\theta)]^T, \quad (3)$$

where the angles are defined according to the right-handed coordinate system shown in Fig. 2.

Let $\mathbf{p}_i$ denote the position of the i -th microphone relative to the array center and let v denote the speed of sound. Defining $\Delta \mathbf{p}_{ij} = (\mathbf{p}_i - \mathbf{p}_j)/v$, the far-field TDoA model becomes

$$t_{ij} = \mathbf{a}_{\theta, \phi}^T \Delta \mathbf{p}_{ij}. \quad (4)$$

Given the set of estimated delays $\{\hat{t}_{ij}\}$ from all $\binom{M}{2}$ microphone pairs, the DoA can be estimated by solving the least-squares (LS) problem

$$\min_{\mathbf{a}_{\theta, \phi}} \sum_{i < j} (\hat{t}_{ij} - \mathbf{a}_{\theta, \phi}^T \Delta \mathbf{p}_{ij})^2, \quad (5)$$

whose closed-form solution is given by [10], [11]

$$\mathbf{a}_{\theta, \phi} = \mathbf{R}^{-1} \mathbf{q}, \quad (6)$$

where

$$\mathbf{R} = \sum_{i < j} \Delta \mathbf{p}_{ij} \Delta \mathbf{p}_{ij}^T, \quad \mathbf{q} = \sum_{i < j} \hat{t}_{ij} \Delta \mathbf{p}_{ij}.$$

In practice, TDoA estimates may be degraded by noise and overlapping events. Although preprocessing techniques such as median filtering [12] and learning-based detection methods [13], [14] can improve robustness, they do not resolve the ambiguity introduced by multiple simultaneous sources. This limitation motivates the data-association framework proposed in the next section.

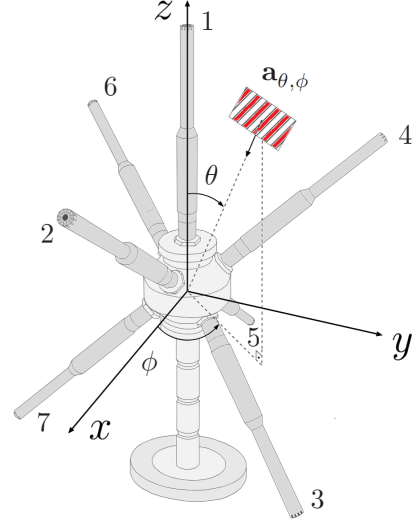


Fig. 2. Geometry of the seven-microphone array and adopted coordinate system. The horizontal (azimuth, ϕ) and vertical (zenith, θ) angles are illustrated. The unit vector $\mathbf{a}_{\theta, \phi}$ denotes the direction of propagation of the incident wavefront.

III. SEPARATION OF GUNSHOT SIGNALS

In multi-shooter scenarios, overlapping gunshot signals yield ambiguous time-of-arrival (ToA) measurements, leading to physically inconsistent time-difference-of-arrival (TDoA) estimates and unreliable DoA estimates. This ambiguity can be formulated as a *data association* problem: given a set of detected arrival times at each microphone, the objective is to correctly assign arrivals to individual sources while enforcing the geometric consistency constraints imposed by the propagation model.

To illustrate this challenge, consider two shooters located at equal distances from the array but observed at directions $(\theta_1, \phi_1) = (10^\circ, 35^\circ)$ and $(\theta_2, \phi_2) = (80^\circ, 35^\circ)$. When the

two shooters fire nearly simultaneously, the detected arrival times across the microphones form two interleaved arrival sequences corresponding to different sources, as shown in Fig. 3. Without prior labeling, temporal ordering alone is insufficient to associate detections with their corresponding sources.

Let $\mathbf{T} \in \mathbb{R}^{M \times n}$ denote the matrix of detected arrival times, where M is the number of microphones and n is the number of sources. Each column corresponds to a candidate set of arrivals, but the correct association between microphones and sources is unknown. From $\mathbf{T}$, a set of candidate TDoAs can be constructed by forming all possible pairwise differences between entries.

The proposed approach resolves this ambiguity by exploiting geometric consistency. Physically valid TDoAs must satisfy the linear model in Eq. (4), which constrains them to a low-dimensional manifold defined by the array geometry. As a result, correctly associated TDoAs yield a common DoA vector and low least-squares residuals, whereas mixed-source associations produce larger residuals.

Based on this observation, we employ a modified iterative least-squares (ILS) procedure to identify consistent subsets of TDoAs. Starting from the full set of candidate delays, the algorithm iteratively estimates the DoA vector $\mathbf{a}_d$ and removes the delay pair that contributes the largest residual error. This pruning process enforces geometric consistency and converges to a subset of TDoAs associated with a single source. This approach can be interpreted as a robust estimation procedure that implicitly rejects outliers arising from incorrect source associations, without requiring explicit clustering or prior labeling.

Once a consistent subset is obtained, the corresponding global ToAs are reconstructed using Eq. (4) with respect to a reference microphone. The matching entries in $\mathbf{T}$ are then removed, thereby isolating one source. The procedure is repeated until all sources are separated.

Algorithms 1 and 2 summarize the proposed procedure for the two-source and general n -source cases, respectively.

Algorithm 1: ILS-based separation for two sources

Input: Detection matrix $\mathbf{T} \in \mathbb{R}^{M \times 2}$
Output: Separated TDoA subsets and corresponding DoA estimates
for $s = 1$ **to** 2 // extract one source per iteration
 do
 Construct the candidate delay-pair set $\mathcal{P}$ from $\mathbf{T}$;
 while $|\mathcal{P}| > N_{\min}$ **do**
 Estimate $\mathbf{a}_{d,s}$ using Eq. (6);
 Compute residuals $r_{ij} = (t_{ij} - \mathbf{a}_{d,s}^\top \Delta \mathbf{p}_{ij})^2$;
 Remove the pair with largest residual from $\mathcal{P}$;
 Compute the final DoA estimate $\hat{\mathbf{a}}_{d,s}$ from the remaining pairs in $\mathcal{P}$;
 Reconstruct the source-consistent ToAs associated with $\hat{\mathbf{a}}_{d,s}$;
 Remove the corresponding detections from $\mathbf{T}$;

A. Complexity Analysis

Let M denote the number of microphones and n the number of sources. The initial number of candidate delay pairs scales

Algorithm 2: ILS-based separation for n sources

Input: Detection matrix $\mathbf{T} \in \mathbb{R}^{M \times n}$
Output: Set of estimated DoA vectors $\mathcal{R} = \{\hat{\mathbf{a}}_{d,1}, \dots, \hat{\mathbf{a}}_{d,n}\}$
Initialize $\mathcal{R} \leftarrow \emptyset$;
Set $N_s \leftarrow n$ // sources to be processed
while $N_s > 1$ **do**
 Construct the candidate delay-pair set $\mathcal{P}$ from $\mathbf{T}$;
 Apply the ILS pruning procedure to estimate $\hat{\mathbf{a}}_d$;
 Store $\hat{\mathbf{a}}_d$ in $\mathcal{R}$;
 Reconstruct the source-consistent ToAs and remove the corresponding detections from $\mathbf{T}$;
 Update $N_s \leftarrow N_s - 1$;
Estimate the final DoA vector $\hat{\mathbf{a}}_d$ from the remaining entries in $\mathbf{T}$;
Store $\hat{\mathbf{a}}_d$ in $\mathcal{R}$;

as $N = O(nM^2)$, since all microphone pairs are considered for each source candidate. At each iteration of the pruning stage, the algorithm computes residuals for all active pairs and removes the worst candidate. The cost per iteration is therefore $O(k)$ for k active pairs. Summing over all iterations yields a worst-case cost of $O(N^2)$ for single-source extraction.

Repeating the procedure for all n sources leads to an overall worst-case complexity bound of

$$T(n, M) = O(n^3 M^4). \quad (7)$$

In practice, however, early pruning of inconsistent pairs and the typically small number of simultaneous sources ($n \ll M$) significantly reduce the effective complexity. This suggests that the proposed approach is suitable for real-time implementation on resource-constrained platforms.

IV. EXPERIMENTAL RESULTS

The proposed method was evaluated using the seven-element microphone array shown in Fig. 2, with coordinates listed in Table I. Although applicable to both MB and SW arrivals, all results reported here use SW arrivals. Validation was conducted in three stages: (i) ideal detection, (ii) simulated detection errors, and (iii) real-world recordings, including 1000 simulations with randomly generated source directions.

TABLE I
COORDINATES OF THE MICROPHONES RELATIVE TO THE ARRAY CENTER

Mic.	x (m)	y (m)	z (m)
1	0.031	0.040	0.266
2	-0.085	0.084	0.186
3	-0.204	0.014	0.176
4	-0.208	-0.084	0.154
5	-0.085	-0.175	-0.174
6	0.231	0.004	0.016
7	0.119	0.184	-0.170

Performance was evaluated using the DoA root-mean-square error (RMSE),

$$\text{RMSE}_{\text{DoA}} = \sqrt{\frac{1}{2K} \sum_{i=1}^K [(\theta_i - \hat{\theta}_i)^2 + (\phi_i - \hat{\phi}_i)^2]}, \quad (8)$$

where θ_i and ϕ_i denote the true zenith and azimuth angles for the i -th observation, respectively, while $\hat{\theta}_i$ and $\hat{\phi}_i$ denote the corresponding estimated angles.

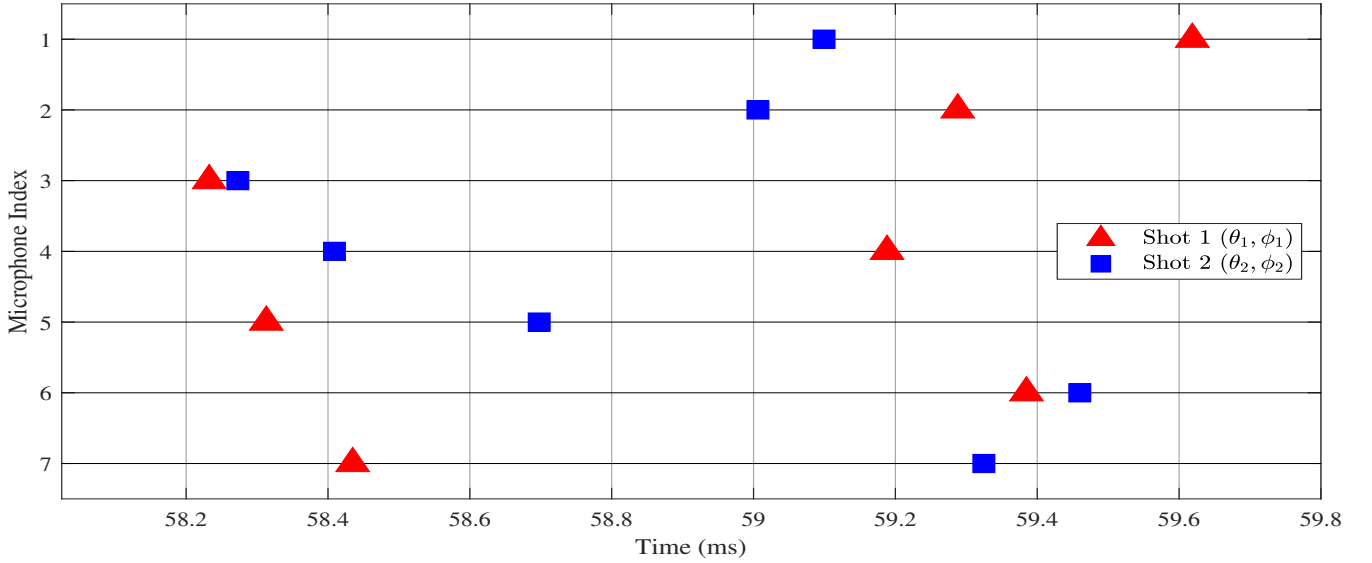


Fig. 3. Detection times of two nearly simultaneous gunshots across the array. The interleaving of arrival sequences leads to data-association ambiguity.

The TDoA association accuracy is defined as

$$\%TA = \frac{N_A}{N_T} \times 100, \quad (9)$$

where N_A denotes the number of correctly associated TDoAs and N_T is the total number of evaluated TDoAs.

A. Ideal Detector

Table II shows the results of evaluating the method under ideal detection, i.e., without timing errors. For non-overlapping shots, a baseline RMSE of 0.53° is obtained. For 1000 simulated pairs of nearly simultaneous shots (2000 total shots), the proposed method achieves a DoA RMSE of 1.83° with a time-association accuracy of 98.11%. Out of 14000 candidate TDoAs (7 microphones, two sources), 13736 were correctly selected. Moreover, 92.6% of the estimated directions exhibit errors below 2° , indicating a strong concentration of estimation errors around small angular deviations. These results indicate that the proposed method introduces only a limited degradation relative to the ideal single-source case, while successfully resolving the source ambiguity.

TABLE II
SIMULATION RESULTS FOR SIMULTANEOUS SHOTS (IDEAL DETECTOR)

Parameter	Value
Total number of shot pairs	1000
N_A / N_T	13736 / 14000
$\%TA$	98.11%
Base RMSE ($^\circ$)	0.53
DoA RMSE ($^\circ$)	1.83
Error $< 2^\circ$	92.6%

B. Detector with Simulated Error

To evaluate robustness, timing errors were introduced by perturbing detection instants with a zero-mean Gaussian variable $e \sim \mathcal{N}(0, \sigma_e^2)$, modeling degraded detection accuracy due

to noise and interference. Table III summarizes the results. The “Base” column corresponds to the error obtained without source overlap, while “ILS MOD” reflects the additional error introduced by the proposed association mechanism. The “Total” column represents the overall DoA error. As expected, performance degrades as σ_e^2 increases. However, the additional error introduced by the association stage remains limited. For example, at $\sigma_e^2 = 2$, the total RMSE is 12.99° , while only 5.33° is attributable to the separation process. Thus, detection uncertainty remains the dominant source of error.

TABLE III
DOA RMSE FOR SIMULATED NOISY SIGNALS

σ_e^2	$\%TA$	Base	ILS MOD	Total
1	98.05	2.69	1.82	4.51
2	96.67	7.66	5.33	12.99
2.5	96.41	13.07	5.35	18.42
3	95.79	17.87	8.49	26.36
4	94.54	25.03	7.97	33.00

C. Real Data

The method was further validated using real recordings acquired at the Forte São João shooting range (Rio de Janeiro, Brazil). The dataset contains near-simultaneous gunshots from multiple shooters for which independently measured ground-truth positions were unavailable. DoA estimation was performed using the SW component. Fig. 4 shows a representative recording that illustrates the challenging conditions encountered in practice, including overlapping arrivals, multipath reflections [15], microphone saturation, and ambiguous onset times. These effects complicate both arrival detection and source association.

A total of 11 overlapping events were analyzed (8 involving two sources and 3 involving three sources), corresponding to 25 individual shots. Reference directions were estimated from

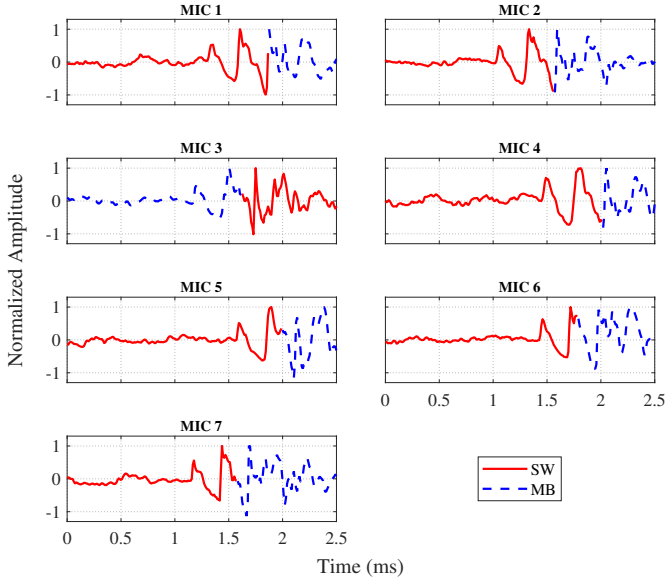


Fig. 4. The color coding indicates the separation of two nearly simultaneous shots. Multipath propagation, SW/SW and SW/MB overlap, saturation, and background noise complicate signal detection and source association.

isolated-shot segments recorded under the same experimental conditions and used as a baseline for evaluating relative performance under realistic operating conditions.

Table IV summarizes the results, showing an overall DoA RMSE of 12.59° , consistent with the simulated noisy conditions and confirming the robustness of the proposed method under realistic propagation effects.

TABLE IV
DoA RMSE FOR REAL SHOOTING DATA

Scenario	Shots	RMSE ($^\circ$)
2 simultaneous	8	13.63
3 simultaneous	3	10.51
Total	11	12.59

V. CONCLUSION AND FUTURE WORK

This paper presents an ILS-based data-association framework for DoA estimation in scenarios with multiple nearly simultaneous gunshots. The proposed method addresses the ambiguity introduced by overlapping acoustic signatures by selecting physically consistent TDoA combinations prior to DoA computation. Simulation results demonstrate that the method achieves high source-separation accuracy (above 98% under ideal detection conditions) with only limited degradation in DoA estimation. Under noisy detection conditions, the results show that the dominant performance limitation arises from detection uncertainty rather than from the proposed association mechanism. Experiments with real recordings yielded a DoA RMSE of 12.59° , consistent with simulation results under perturbed conditions. These results indicate robustness to signal overlap, reflections, and sensor-proximity effects.

The analysis also highlights that performance is strongly influenced by the quality of the detection stage, particularly when MB/SW overlap is present. While the proposed method effectively resolves data association ambiguities, further improvements in detection are necessary to enhance overall accuracy. Future work will focus on integrating advanced detection and estimation techniques within the proposed framework. In particular, learning-based approaches such as NGCC-PhaT [14] and subspace-based methods [6] will be investigated to improve TDoA estimation and robustness in low-SNR and highly overlapping scenarios. Additional field experiments will also be conducted to further validate the approach under operational conditions.

REFERENCES

- [1] J. C. Freytag and B. M. Brustad, "A survey of audio forensic gunshot investigations," in *Proceedings of the AES 26th International Conference*, July 2005, pp. 131–134.
- [2] P. C. Prandel, I. L. Freire, and J. A. Apolinário Jr., "Detecção e estimação em tempo real de direção de chegada de sinais impulsivos," in *XXX Simpósio Brasileiro de Telecomunicações*, 2012.
- [3] I. L. Freire and J. A. Apolinário Jr., "DoA of gunshot signals in a spatial microphone array: performance of the interpolated generalized cross-correlation method," in *Proc. 5th Argentine School of Micro-Nanoelectronics, Technology and Applications*, 2011.
- [4] R. C. Maher, "Modeling and signal processing of acoustic gunshot recordings," in *Proceedings of the 12th Digital Signal Processing Workshop (DSP)*, September 2006, pp. 257–261, IEEE.
- [5] A. L. L. Ramos, S. Holm, S. Gudvangen, and R. Otterlei, "Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics," in *Proceedings of SPIE*, 2011, vol. 8019, p. 8019U.
- [6] A. M. C. R. Borzino, J. A. Apolinário Jr., and M. L. R. Campos, "Consistent DOA estimation of heavily noisy gunshot signals using a microphone array," *IET Radar, Sonar & Navigation*, vol. 10, no. 9, pp. 1519–1527, 2016.
- [7] 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, 1988.
- [8] Çağlar Akman, "Multi shooter localization with acoustic sensors," M.s. thesis, Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, September 2017, Supervisor: Prof. Dr. Mehmet Kemal Leblebicioğlu.
- [9] 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, August 1976.
- [10] Y. Chan, R. Hattin, and J. B. Plant, "The least squares estimation of time delay and its use in signal detection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 3, pp. 217–222, 1978.
- [11] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.
- [12] A. M. C. R. Borzino, J. A. Apolinário Jr., M. L. R. de Campos, and L. W. P. Biscainho, "Signal enhancement for gunshot DOA estimation with median filters," in *IEEE 6th Latin American Symposium on Circuits & Systems (LASCAS)*, 2015, pp. 1–4.
- [13] Héctor A. Cordourier Maruri, Ngoc Q. Vu, and Mark D. Plumbley, "GCC-PHAT Cross-Correlation Audio Features for Simultaneous Sound Event Localization and Detection (SELD)," in *DCASE 2019 Workshop*, New York, USA, 2019.
- [14] Orchi Dorchi, "Multiple-Peak Extraction from GCC-PHAT for Overlapping Source TDOA Detection," 2022, <https://arxiv.org/html/2408.17166v1>.
- [15] A. L. L. Ramos, Sverre Holm, Sigmund Gudvangen, and Ragnvald Otterlei, "The multipath propagation effect in gunshot acoustics and its impact on the design of sniper positioning systems," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense XII*, Edward M. Carapezza, Ed. International Society for Optics and Photonics, 2013, vol. 8711, p. 87110A, SPIE.