Optimizing Sensor Placement in Urban Environments for Time Difference of Arrival Shooter Localization and Event Classification

David-Octavian Iacob*, Philipp Mikus†, Matthias Ospel*,
Luisa Still†, Cyril Blonde-Weinmann*, and Marc Oispuu†
*French-German Research Institute of Saint-Louis (ISL), Saint-Louis, France
†Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE), Wachtberg, Germany
{david-octavian.iacob, matthias.ospel, cyril.blonde-weinmann}@isl.eu
{philipp.mikus, luisa.still, marc.oispuu}@fkie.fraunhofer.de

***Abstract*—This study addresses the optimization of the placement of multiple acoustic sensors for shooter localization and event classification in urban environments. Using ray casting solutions of the Eikonal equation, the shortest propagation paths in the urban model are computed for all source-receiver pairings. Subsequently, the expected Times of Arrival (TOAs) from virtual sources are used to evaluate the localization performance of a given sensor setup using a Monte Carlo approach. Similarly, the modelled signal paths are used to estimate the signal-to-noise ratio (SNR) of the source at the sensor level in order to predict the expected classification performance. Subsequently, a genetic algorithm solves the underlying optimization problem based on these performance metrics and identifies optimal sensor network configurations for shooter localization and event classification within the urban environment. The method is experimentally validated using audio data of propane gas cannon shots recorded at the French-German Research Institute of Saint-Louis.**

***Index Terms*—optimization, variance, sensor placement, classification, localization, ray casting**

I. Introduction

Impulsive sound sources, such as gunshots or explosions, can indicate acute threats that need to be identified and localized, and acoustic sensors can make a valuable contribution for surveillance purposes of impulsive events. For example, acoustic microphone arrays can determine a shooter state based on the directions of arrival (DOA) of the muzzle blast and the shock wave and the corresponding time difference between the aforementioned acoustic events [1]–[5]. To monitor larger areas, (unsynchronized) sensor networks consisting of dislocated individual microphones can be used to determine the shooter location based on time difference of arrival (TDOA) measurements. Aspects of the two-dimensional and three-dimensional TDOA-based localization problem examined in the literature include observability considerations, numerous estimation algorithms, and localization accuracy [6], [7]. It is known that the localization accuracy depends on the number of sensor nodes, the measurement accuracy and, in particular, the placement of the network sensor nodes w.r.t. the sound source location [6], [8]. In [9], a splay configuration with equal angle increments between neighboring sensors has been suggested in order to improve the source localization accuracy.

Enhancing safety measures in urban settings requires the development of effective shooter localization and event classification systems. Compared to free-field conditions, the urban environment significantly influences sound propagation in the complex propagation channel by introducing reflections and diffractions. Consequently, the interpretation of the received signals to enable an accurate shooter localization and event classification is a challenging task. However, if the urban environment is accurately known, source localization can be improved [10], [11] or enabled in non-line-of-sight conditions [12], [13] by exploiting multipath signal propagation effects.

For a given source-receiver pairing, the shortest path of the wave propagation in the considered urban environment can be determined by using ray casting solutions of the Eikonal equation [14].

In order to accurately locate military threats, such as gunshots or explosions, it is important to also classify them accordingly, to mitigate false alarms and to be immune to countermeasures [15]. To that end, the ISL has developed a multi-label AI-based embedded sound signature classification system capable of discriminating between *Firearm*, *Motorised road vehicle*, *Speech* and *Background noise* acoustic samples. Its reliability can be evaluated by traditional metrics that require a human user to choose a classification threshold, such as its sensitivity (true positive rate) or specificity (true negative rate), although more objective classification performance metrics encompass the entire spectrum of classification thresholds. Such metrics include the Receiver Operating Characteristic (ROC), allowing for a qualitative analysis and comparison, or the Area under ROC (AUROC), quantifying the overall performance of a system. These metrics have unique values for each class of interest, and were used to show that the sensor placement has an impact on its classification performance.

Furthermore, the localization of an acoustic source using microphone arrays has been solved by exploiting incomplete measurement data sets [16]–[18], incorporating urban environments [10], and considering imprecise or unknown sensor state information [17]. A genetic algorithm has been proposed to optimize the array sensor locations and to enhance the

localization performance, which was successfully validated in simulations and experimental measurements [19], [20].

Since the locations of the acoustic sensor nodes are of significant importance w.r.t. the achievable surveillance performance, two tasks are considered in this paper: TDOA-based localization and event classification in urban environments. In order to apply the genetic algorithm in [20] to the aforementioned cases, we propose metrics that predict the corresponding surveillance performance for given sensor locations. Firstly, in order to predict the expected TDOA-based localization performance, a Monte Carlo approach is proposed. It involves the computation of solutions of the wave equation within urban models by employing ray-casting and graph traversal algorithms. A variance metric is derived from sound propagation simulations as a cost function for the TDOA sensor placement optimization algorithm. Secondly, for the expected classification performance, the ray-casting method was used to estimate the source-to-sensor propagation distances and the sound absorption due to the propagation. Therefore, if the source intensity is known, the SNR and implicitly the classification performance can be estimated. The resulting classification performance metric is evaluated as the average of the expected AUROC and of the true positive prediction probability. Validation of this approach was conducted through experiments at the ISL, emphasizing the pivotal role of sensor placement in shooter localization and event classification accuracy.

II. Localization and Classification Methodology

A. TOA Computation

The propagation of small pressure perturbations in a medium due to an acoustic source is modelled by the linearized inhomogeneous wave equation [21]:

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) p(x,t) = f(x,t). \tag{1}$$

where c is the speed of sound in m/s, t is the time in s, p is the pressure and $f(x,t)$ is the source function. When solely the time component of the wave equation is of interest, it is sufficient to compute the solution of a reduced form of the wave equation, the Eikonal equation. The solution of the Eikonal equation for a given source position maps the travel time of the first wave front to space. The Eikonal equation can be derived from the linearized wave equation with the assumptions, that the amplitude of the wave does not depend on time or frequency and is given by [22]:

$$|\nabla\Phi(x)|^2 = \frac{1}{c^2(x)}, \tag{2}$$

where $\Phi(x)$ is the travel time function. The solution is trivial when line of sight is possible since then the travel time is simply $t = r/c$, where r is the propagation distance. However, when line of sight is not possible, the shortest path of travel time needs to be identified, which can become challenging depending on the shape of the sub surfaces and numbers of polygons the intersected objects are made up by.

In our previous research, the Finite Element Method has been effectively employed to solve wave equations with boundary conditions in the context of source localization problems in urban terrain [14]. However, an alternative approach utilizing the ray solution of the Eikonal equation offers a promising avenue for the efficient computation of travel times for source-receiver pairings. This method presents opportunities for improved computational efficiency and warrants exploration as a potential alternative to existing methodologies and will be used throughout this work. Computation steps for the solver are visualized in Fig. 1.

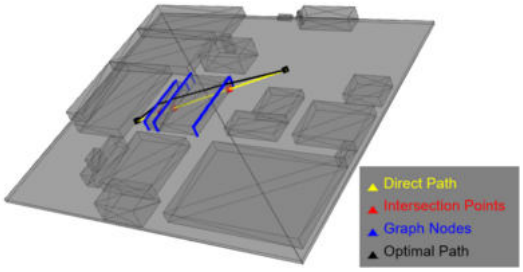

Fig. 1: A visualization of the optimal path of shortest travel time using the ray casting solver given simple polygon shapes.

Firstly, intersections are detected by ray casting along the path from point A to point B. In cases where a direct line of sight is not possible, graph nodes are generated on the hull of impacted objects depending on the neighbouring geometry. Next, an adjacency matrix is initialized using line of sight checks for all point combinations on the hull using ray casting. Subsequently, a graph is established based on the adjacency matrix. Computing the propagation time of a pressure wave along the edges, the corresponding edge weights are assigned. The identification of path in the graph representing the shortest travel time in the weighted graph is accomplished using an A*-search algorithm. The time of flight of the weights is integrated along the ray-path to determine the solution of the Eikonal equation, the wave's travel time. The solver generates a symmetric matrix of TOAs between all pairs of the N_{pt} registered grid points within a given grid in $\mathbb{R}^3$:

$$\mathbf{T} = \begin{bmatrix} 0 & t_{1,2} & \cdots & \cdots & t_{1,N_{pt}} \\ t_{2,1} & 0 & t_{2,3} & \cdots & t_{2,N_{pt}} \\ \vdots & t_{3,2} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & t_{N_{pt}-1,N_{pt}} \\ t_{N_{pt},1} & t_{N_{pt},2} & \cdots & t_{N_{pt},N_{pt}-1} & 0 \end{bmatrix} \tag{3}$$

where $t_{1,j}$ is the time of flight from the first point to the jth point, as well as the corresponding distance matrix $\mathbf{D}$ which contains the shortest paths and the vector of grid points $\mathbf{p} = [p_1, p_2, \cdots, p_{N_{pt}}]$. Finally, transport equations along the ray path are evaluated to compute the amplitude.

B. TDOA matching grid-based localization

The single shooter localization algorithm used in the scope of this project relies on a TDOA matching approach. A net-

work of S_{nodes} time-synchronised sensors is being deployed at various grid points of a pre-modeled urban environment comprised of N_{pt} grid points. Each sensor is capable of independently detecting acoustic events associated to shots and of extracting their timestamp, in an absolute time referential shared by all sensors.

Each sensor-provided TOA is defined as t_i^{sen}, with $i \in \{1, ..., S_{nodes}\}$, and the individual t_i^{sen} scalars corresponding to one event compose the T^{sen} vector of absolute measured TOAs, ordered according to the sensor index i. The minimum TOA value of a vector is defined as $t_{min}^{sen} = \min(t_i^{sen})$ and its sensor index i_{ref} corresponds to the index of the sensor that has generated t_{min}^{sen}. This reference index will be used later in the TDOA matching approach.

Then, based on these values, we generate a vector of TDOAs. The coordinates of this T_Δ^{sen} vector are computed as follows:

$$t_{i\Delta}^{sen} = t_i^{sen} - t_{min}^{sen} \tag{4}$$

The time matching approach searches within the list of registered grid points p_j the point p^* whose corresponding vector of model TDOAs $T_\Delta^{mod}(p_j)$ minimizes the Euclidean norm relative to the sensor-provided T_Δ^{sen} vector:

$$L_2(p_j) = ||T_\Delta^{sen} - T_\Delta^{mod}(p_j)||_2, \tag{5}$$

where $T_\Delta^{sen}, T_\Delta^{mod}(p_j) \in \mathbb{R}^{S_{nodes}}$, $p_j \in \mathbb{R}^2$ and $j \in \{1, ..., N_{pt}\}$.

The vectors of model TDOAs $T_\Delta^{mod}(p_j)$ are computed based on the vectors of absolute model TOAs as follows:

$$t_{i\Delta}^{mod}(p_j) = t_i^{mod}(p_j) - t_{i_{ref}}^{mod}(p_j), \tag{6}$$

where i_{ref} is the previously-extracted reference sensor index. Finally, each absolute $t_i^{mod}(p_j)$ value is extracted from the pre-computed model as the TOA from a source located at a grid point p_j of coordinates (x_j, y_j) to the sensor of index i. The p^* grid point that minimizes the aforementioned Euclidean distance is considered to be the estimated source location.

C. Sound signature classification

The ISL embedded and multi-label sound signature classification system used in the scope of this study is capable of classifying *Firearm*, *Motorised road vehicle* and *Speech* audio samples, with all other audio samples, including false alarms, being implicitly classified as *Background noise*. The classifier first converts the input audio stream to a time-frequency representation using Continuous Wavelet Transforms [23]. These spectral representations are then converted to images and classified using a multi-layer AlexNET [24] image classifier, offering an excellent trade-off between the classification performance and the computational cost thereof. In the scope of this study, only the performance of the *Firearm* classification was evaluated and compared, with all the other classes (including *Background noise*) being considered negatives.

The system's classification accuracy was evaluated on a previously-collected and manually annotated ISL-FKIE field-collected dataset comprised of 44,760 audio windows, corresponding to over 4 hours of recorded audio data. Using a classification probability threshold of 50%, the system has achieved a sensitivity of 95.0% and a specificity of 97.2% with respect to the classification of *Firearm* audio samples.

As these performance evaluation metrics depend on the chosen classification threshold, a more objective classification performance metric comes in the form of the AUROC, quantifying the area encompassed under the ROC curve of each class of the classifier. With a completely random classifier showcasing an AUROC of 0.5 and a perfect classifier an AUROC of 1.0, our embedded sound signature classification system achieves an AUROC of 0.9961 with respect to the classification of *Firearm* samples.

III. Optimal sensor placement methodology

A. Localization performance prediction

In an environment where the TOAs from a source to each sensor within a modeled grid exactly align with their simulation values, pinpointing the source's location is straightforward. However, variations inevitably arise, leading to discrepancies between the measured TOA values and the corresponding simulation results. The variations stem from various factors, such as differing propagation conditions (e.g., air temperature, humidity, wind speed), analog channel effects, changes in the urban environment, or measurement errors. To gauge the localization accuracy of a sensor setup tasked with localizing impulsive sound sources, we employ a Monte Carlo simulation methodology. This approach allows us to account for the introduced errors and uncertainties inherent in real-world scenarios, providing a more comprehensive assessment of localization performance and its robustness to erroneous measurements. In the scope of this research, the evaluated grid points, source locations and sensor locations are located in **2 dimensions**, with the z coordinate being constant and aligned with the ground level, but the methodology and equations thereof can be easily extrapolated to a 3-dimensional model.

One simulation run is performed for a number K_{loc} of evaluated source locations, and each source location is evaluated for a number of N_{sim} simulation steps. For each simulation step, we retrieve from the propagation model the vector of theoretical TOAs $T^{mod}(p_j^{ref})$ from the evaluated source location $p_j^{ref}(x_j^{ref}, y_j^{ref})$ to each of the S_{nodes} sensors. Then, each theoretical TOA value t_i^{mod} is multiplied by a continuous random error variable X_{err} following a normal distribution of mean $\mu = 1$ and a given standard deviation σ_{err}. The resulting altered TOA vector named $T^{sim}(p_j^{ref})$ is then used as a "simulated sensor data" vector that is used as an input of the localization algorithm. The resulting output simulated sensor location is $p_j^{sim}(x_j^{sim}, y_j^{sim})$.

The simulation procedure evaluates the variance of the predicted source locations relative to the reference locations on each axis. Since the variance decreases as the accuracy

and precision increase, we have defined the localization performance estimation metric as the average value of the inverse of the variance across each axis. The localization performance metric is computed as follows:

$$P_{loc}^{2D} = \frac{1}{2}\left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2}\right) \tag{7}$$

where, for a number of total simulation steps N_t:

$$\sigma_x^2 = \frac{\sum_{j=1}^{N_t}(x_j^{sim} - x_j^{ref})^2}{N_t}; \sigma_y^2 = \frac{\sum_{j=1}^{N_t}(y_j^{sim} - y_j^{ref})^2}{N_t} \tag{8}$$

B. Classification performance prediction

Predicting the classification performance was done in two stages. The first one allowed us to understand the impact of the SNR at the sensor level upon a number of classification performance metrics, whereas the second allowed us to estimate the expected SNR at each sensor level based on the source location and the environment.

A preliminary ISL study has shown that the classification performance varies with respect to the sensor placement. A further investigation of this matter was done within this work by analyzing the correlation between the SNR of each audio sample and its corresponding classification performance. To that end, the audio sample windows of a reference dataset were ordered according to their estimated average SNR and processed using a fixed size moving window. Two classification performance metrics were correlated to the average estimated SNR of each batch of audio windows: the AUROC and the true positive prediction probability.

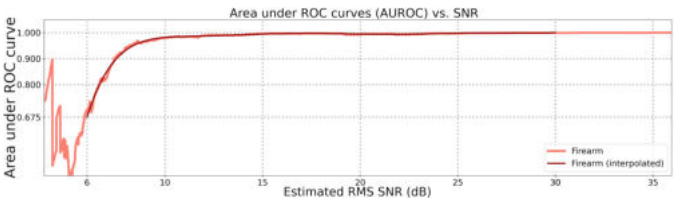

Fig. 2: Correlation between the AUROC and the SNR for the *Firearm* class.

Fig. 2 displays the evaluated AUROC value as a function of the estimated SNR, for a moving window buffer encompassing 256 audio windows of similar SNR. Similarly, the true positive prediction probability was plotted against the SNR using a moving window buffer of size 128, and the obtained data points were interpolated. The curves are displayed in Fig. 3.

The polynomial interpolation of these data points is essential in order to obtain AUROC and true positive prediction probability estimators as functions of the SNR, within a fixed interval of definition. Outside the boundaries of this interval, the estimated performance values are extrapolated to their minimum and maximum evaluated values respectively. This is the first key component of the classification performance estimation process, generating an expected performance metric as a function of an estimated (or known) SNR value.

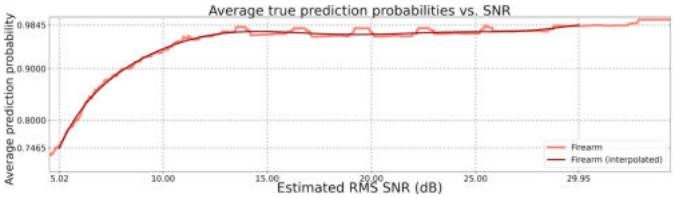

Fig. 3: Correlation between the true positive prediction probability and the SNR for the *Firearm* class.

At this point, the missing link allowing us to predict the classification performance of a sensor or a network thereof based on their placement is the estimation of the SNR for a given sensor geometry in a given environment. This is done by evaluating the transport equations for the source receiver pairings along the identified paths. The attenuation of wave propagation due to meteorological influences was heavily studied since 1960 [25]–[27]. In our study, it has been estimated by modelling the geometric divergence including the atmospheric absorption, where the decrease of intensity with distance is modelled via [26]:

$$A_s(d) = 20\Gamma \log_{10}\left(\frac{d}{d_{ref}}\right) \tag{9}$$

where:

- d is the distance between the acoustic centre of the source and the receiver in m
- d_{ref} is the distance between the acoustic centre of the source and the reference point (the point at which the level of the acoustic source is measured); typically, $d_{ref} = 1$m
- Γ reflects the nature of the wave; in this scenario, $\Gamma = 1$, corresponding to a spherical wave propagation

Subsequently, the atmospheric absorption, derived from two mechanisms (classical effects and relaxation effects) for sine waves are given by [28]:

$$A_a(d) = -20\log_{10}\left(e^{-\alpha d}\right) = ad \tag{10}$$

where:

- α is the attenuation coefficient of the atmosphere in nepers/m
- a is the attenuation coefficient of the atmosphere in dB/m ($a = 0.868\alpha$)

The absorption coefficient of the atmosphere a depends on a number of measures [29], including the ambient air pressure, the relative humidity, the sound frequency and the speed of sound in air. As a reference for the scope of this study, we have considered a worst-case propagation scenario in which the absorption coefficient has a value $a = 0.1$ dB/m, corresponding to a 5 kHz sound wave propagating at 0 °C, in an atmosphere with a relative humidity of 20% and a pressure of 0.94 atm. Finally, using these models for estimating

the sound absorption, we can derive the following equation for expressing the signal-to-noise ratio as a function of the propagation distance between the source and the sensor:

$$SNR(d) = I_{src} - A_s(d) - A_a(d) - I_{bgd} \tag{11}$$

where:

- I_{src} is the sound intensity of the source measured at $d_{ref} = 1$m in dB
- I_{bgd} is the background noise intensity measured at the sensor level in dB

Therefore, by modelling the absorption of the wave emitted from a source location within a known environment to the sensor locations, and by using the above model for predicting the classification performance as a function of the SNR, we can estimate the classification performance of each sensor (either from a single source location or from a set of source locations) within the environment. This performance prediction layer, implemented as a standalone Python library encompassing the sound attenuation model as well as the performance prediction metric functions, is used by the following Genetic Algorithm to find the optimal sensor locations with respect to the sound signature classification performance.

C. Genetic Algorithm (GA)

The popularity of GAs stems in part from their broad applicability and ease of use compared to classical optimization methods [30]. On one hand, GAs are advantageous for the given problem, because they allow different metrics to be exchanged easily without the need to calculate derivatives. On the other hand, they provide robust search in complex and discrete search spaces. GAs make use of a multi-hypothesis approach by operating with a population consisting of a predefined number of solutions to the problem, in this case specific sensor arrangements called individuals. The following approach is based on the work in [20] and consists of the following steps, with he sequence of operations also depicted in Fig. 4:

1) *Initialization*: The population contains I individuals. Let each individual $\mathbf{P}_i = \begin{bmatrix} p_1 & \dots & p_N \end{bmatrix}$, $i = 1, ..., I$, $p_i \in \mathbb{R}^2$ be the matrix containing the 2D coordinates of the N sensor positions. Here, the initial population is randomly initialized within the provided sensor area and its initial fitness is evaluated (see step 5).
2) *Parent Selection*: To preserve diversity in the population and reduce the selection pressure, tournament selection is performed with a tournament size of k. This way, k individuals are selected at random and the individual with the highest fitness rating is chosen as a parent. This process is repeated I times.
3) *Crossover*: The crossover operation is used to create I new individuals that make use of information from the parents, while also exploring new areas of the search space. Here, we form the sensor arrangement of an offspring individual by selecting each sensor position to be one of the sensor positions of a randomly selected parent. This way, good sensor positions can be preserved while new arrangements are evaluated.
4) *Mutation*: In order to find new sensor positions and therefore further increase the diversity in the population, mutated versions of the parents and the offspring are created, where the positions are replaced by random grid points in the surrounding area within a certain radius.
5) *Fitness Evaluation*: The fitness function $f(\mathbf{P}_i) : \mathbb{R}^{2\times N} \rightarrow \mathbb{R}$ maps an individual to a scalar value which determines its fitness. The fitness is determined by averaging over all possible target positions for the given sensor geometry.
6) *External Selection*: Finally, the individuals that advance to the next generation are selected. Out of the current population, the offspring and the mutated versions of parents and offspring, the fittest I individuals are chosen to form the population of the next generation.

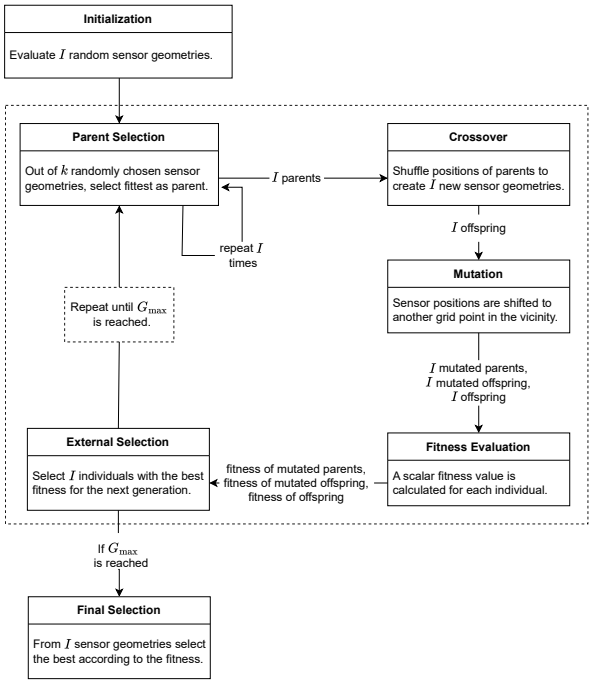

Fig. 4: Sequence of operations of the genetic algorithm.

After initialization, the steps 2-6 are repeated until the predetermined maximum generation $G_{\max}$ is reached. Afterwards, the individual with the highest fitness value is selected as the result of the optimization problem.

IV. Experimental Methodology

In order to show the impact of optimal sensor placement setups and the ability of our algorithms to generate such optimal setups, we have devised an experiment where the performance of two optimally-placed acoustic sensor setups was evaluated and compared against those of two corresponding sub-optimally placed sets of sensors. One pair of setups, optimal and sub-optimal, aimed at evaluating the TOA-based localization performance, whereas the other focused on

the classification performance. A total number of 6 acoustic sensors were used for both localization and classification purposes.

The optimal placement setups were generated using our GA-based tool, which relies on a pre-computed grid point model of the experimental environment, generated around the 3D model shown on the left of Fig. 5. This model is used to determine the TOAs and the propagation distances between every pair of grid points, as described in Sec. II-A. In total, a number of 302 grid points and 91,204 wave propagation paths were evaluated, for a grid resolution of 5.91 m. The calculated look-up tables are then passed to the optimization algorithm, which is used to identify the sensor setups that ensure the best overall localization and classification performances respectively.

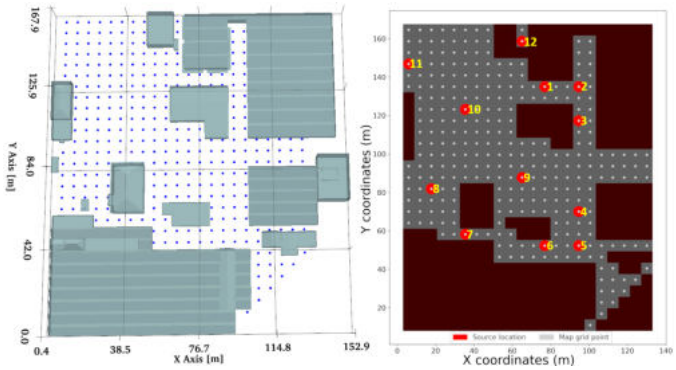

Fig. 5: 3D representation of the test environment and of its underlying grid point model with source positions in red.

A number of 12 source locations were considered for this experiment, spread around the test environment, as shown on the right side of Fig. 5 using red circles with yellow indices.

Despite attempts to use the GA for identifying bad sensor configurations, the optimization process consistently converged to setups where all sensors were placed in the same location. This behaviour is expected, as this scenario implies that TDOA localization would fail, and the reliably classified total area would be minimum. Subsequently, sub-optimal reference configurations were selected among random configurations respecting a minimum inter-sensor distance of 40 m.

A. Generating the sensor placement setups

1) Localization: When searching for the optimal localization sensor placement, the relative TOA error is assumed to have a standard deviation of $\sigma = 3\%$ and 50 Monte Carlo runs were evaluated per source position. The locations of 6 sensors were optimized per scenario. Moreover, a population size of 150 was used and evaluated over 50 generations.

The optimization result is shown in the right side of Fig. 6, with the left side displaying the randomly-selected sub-optimal setup. The light gray dots represent the grid points, the blue circles show the chosen sensor locations in each setup, whereas the light green areas indicate the best localization performance (as opposed to the dark red areas, corresponding to a poor performance). It can be noted that the sensor locations lie at the edge of the sampled location grid and are well separated from each other. This is consistent with results on the optimal placement of TOA sensors in free-field scenarios, which have shown that for equal sensor measurement noises, equiangular sensor separation w.r.t. the source position is optimal [9], [31]. The displayed arrangement ensures a similar separation for most of the possible target positions.

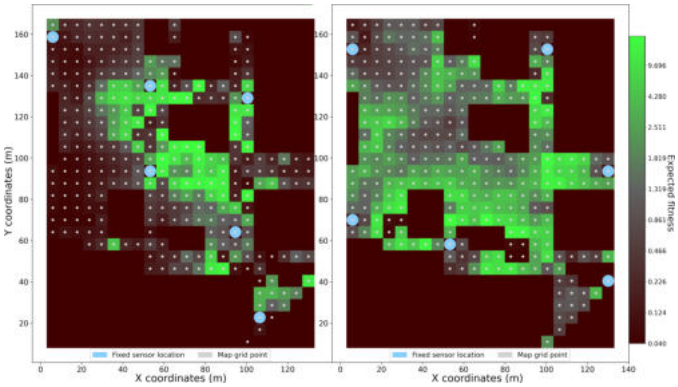

Fig. 6: The sub-optimal (left) and optimal (right) localization sensor placement setups and their expected performance.

2) Classification: For the optimization with the classification-based metric, a sound source intensity of 100 dB and a background noise intensity of 60 dB was assumed. A population size of 250 was evaluated over 60 generations.

The results are displayed on the right side of Fig. 7, with the left side displaying the randomly-selected sub-optimal setup and a colour-coding similar to that of Fig. 6. A single sensor leads to an improved accuracy omni-directionally around it and decreasing with greater distance from the emitter. Additionally, the improvement in classification performance gained by one sensor is independent from the other sensors. This leads to the sensors being distributed almost evenly over the area.

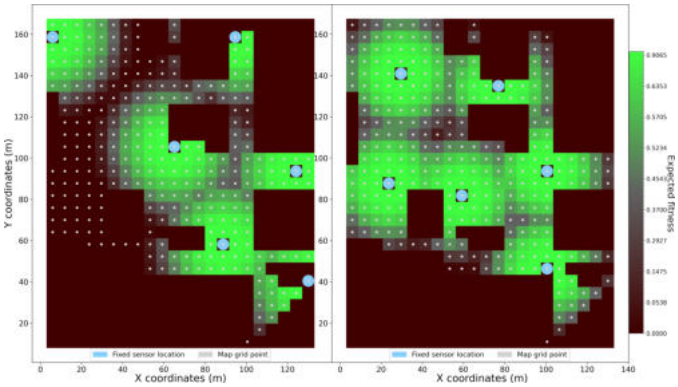

Fig. 7: The sub-optimal (left) and optimal (right) classification sensor placement setups and their expected performance.

B. Performance evaluation

To verify whether the localization and classification performance improves for the optimized sensor locations, the

proposed method was applied to determine sensor positions in an urban environment and measurement data was collected and evaluated. A total of 4 sensor configurations were evaluated, two corresponding to the sub-optimal and optimal localization performance setups and the other two to the sub-optimal and optimal classification performance setups. The acoustic source was a propane gas cannon, capable of generating loud impulsive sounds similar to the blast wave of shots. For statistical analysis, 10 shots were fired from each of the 12 source position for every sensor geometry under test.

Each sensor consisted of a GRAS 40PL microphone, a PCB 480E09 ICP sensor signal conditioner and a Tascam DR-100 MkII audio recorder. Each input signal was recorded at a sampling frequency of 48 kHz and a depth of 24-bit per channel. The sensors were manually synchronised and calibrated at the beginning and at the end of each experimental round. The recordings have been annotated in order to identify the TOAs of shots and the ground truth with respect to the sound classification task, based on a a semi-automated and human validated annotation and event registration process. The resulting annotation files were then post-processed to generate localization and classification outputs by the corresponding algorithms.

V. Results

A. Localization performance

Fig. 8 shows a comparison of the unoptimized and optimized sensor array configurations used for the surveillance of the $21,780$ m^2 area. The localization error is plotted against the percentile of most accurate measurements (top) and for each source position using all measurements (middle). The inverse variance metric is plotted for both configurations versus the source index.

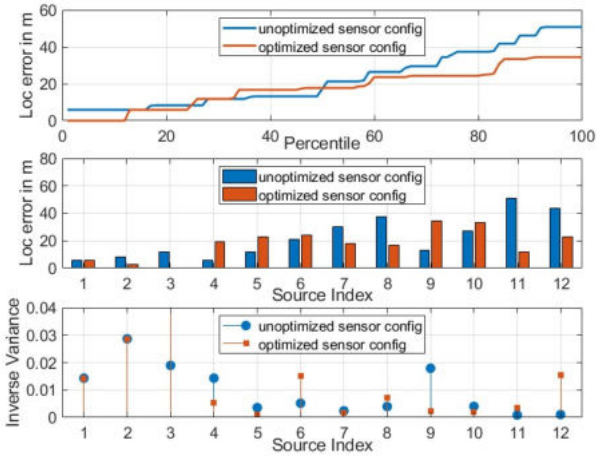

Fig. 8: Localization error and inverse variance: comparison of the performance of the unoptimized and optimized sensor array configurations in experiment.

An average localization error of 22.35 m and inverse variance metric of $2.77 \times 10^{-3}\,\mathrm{m}^{-2}$ have been determined using the unoptimized sensor configuration. Subsequently, an average localization error of 17.69 m and inverse variance metric of $4.69 \times 10^{-3}\,\mathrm{m}^{-2}$ have been determined using the optimized sensor configuration. The optimized sensor configuration appears to perform consistently across the quantiles of data, whereas the unoptimized sensor configuration appears to have systematically disproportionately large errors compared to the percentile rank above 50. It can be concluded that the optimized sensor configuration is likely more robust and performs more consistently.

The localization error shows a correlation with the inverse variance metric. Consequently, it tends to be minimal when the variance metric is high and maximal when the inverse variance metric is low. Notably, the inverse variance metric is infinite for the optimal sensor configuration for source position 3, as the source node has been localized exactly for every shot.

B. Classification performance

The collected audio data was evaluated by comparing the ground truth of the recorded audio files to the outputs provided by the sound signature classification system, with respect to the *Firearm* class. Two classification performance metrics were evaluated: the AUROC and the true positive prediction probability. On top of that, the ROC curves corresponding to each sensor setup were traced, allowing for a qualitative comparison of the two scenarios.

TABLE I: Comparison of the evaluated classification performance metrics.

Evaluated metric	Sub-optimal setup	Optimal setup	Relative improvement
AUROC	0.9491	0.9631	2.8 %
T+ pred. prob.	79.30%	85.41%	6.11%

Table I shows, for each of the two setups, the corresponding metric values. The evaluated data shows a performance improvement both in terms of AUROC ($+2.8\%$ relative to a random classifier) and of the true positive prediction probability ($+6.11\%$).

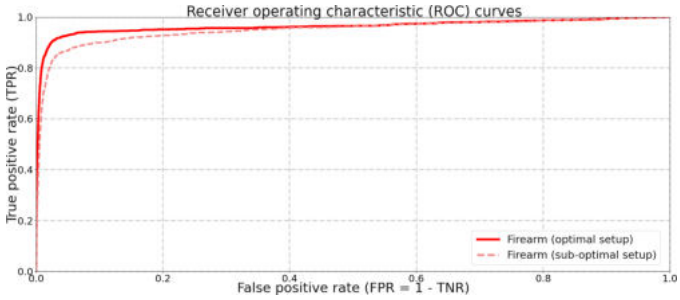

Fig. 9: Comparison of the ROC curves of the sub-optimal (dotted) and optimal (continuous) sensor setups.

The comparison of the ROC curves of the two classification scenarios (dotted for the sub-optimal and continuous for the optimal one) show a similar improvement, that can be observed in Fig. 9. The curve of the optimal setup is closer to that of a

perfect classifier, particularly for low classification thresholds, implying a better sensitivity for similar specificity values.

VI. Conclusions

The proposed approach integrates numerical acoustics and sensor information fusion principles, contributing to an interdisciplinary framework for addressing the optimization of sensor placement in complex urban scenarios. Using the proposed optimization technique for sensor placement in an urban environment, we have observed significant enhancements in localization accuracy and reliability, based on time differences of arrival. The discrepancy in localization accuracy was particularly notable for higher percentiles, with an improvement of up to $16.4\,\mathrm{m}$ in accuracy compared to the non-optimal setup. Notably, the performance improvement was more pronounced for the localization setup compared to the classification setup, reflecting the challenges posed by urban environments in inducing localization errors. It has been demonstrated that the optimal sensor setups surpassed their sub-optimal counterparts in both localization and classification tasks. The experiments have validated the proposed optimization approach for sensor array optimization for the surveillance of urban environments.

Acknowledgment

We would like to thank the BAAINBw for funding the project of optimal sensor placement in urban enviroments and for supporting the cooperation between Fraunhofer FKIE and the French-German Research Institute of Saint-Louis.

References

[1] L. M. Kaplan, T. Damarla, and T. Pham, "QoI for passive acoustic gunfire localization," in *2008 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, 2008, pp. 754–759.

[2] J. George and L. M. Kaplan, "Shooter localization using soldier-worn gunfire detection systems," in *14th International Conference on Information Fusion*, 2011, pp. 1–8.

[3] J. Bedard and S. Pare, "Ferret: a small arms fire detection system: localization concepts," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement II*, E. M. Carapezza, Ed., vol. 5071, International Society for Optics and Photonics. SPIE, 2003, pp. 497 – 509.

[4] J. A. Mazurek, J. E. Barger, M. Brinn, R. J. Mullen, D. Price, S. E. Ritter, and D. Schmitt, "Boomerang mobile counter shooter detection system," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, E. M. Carapezza, Ed., vol. 5778, May 2005, pp. 264–282.

[5] R. W. Osborne, Y. Bar-Shalom, J. George, and L. Kaplan, "Data fusion from multiple passive sensors for multiple shooter localization via assignment," in *17th International Conference on Information Fusion (FUSION)*, 2014, pp. 1–7.

[6] R. Kaune, J. Hörst, and W. Koch, "Accuracy analysis for TDOA localization in sensor networks," in *14th International Conference on Information Fusion*, 2011, pp. 1–8.

[7] R. Kaune, "Accuracy studies for TDOA and TOA localization," in *2012 15th International Conference on Information Fusion*, 2012, pp. 408–415.

[8] Z. Xu, Y. Guo, and J. H. Saleh, "Multi-objective optimization for sensor placement: An integrated combinatorial approach with reduced order model and Gaussian process," *Measurement*, vol. 187, p. 110370, 2022.

[9] K. Doğançay and H. Hmam, "On optimal sensor placement for time-difference-of-arrival localization utilizing uncertainty minimization," in *2009 17th European Signal Processing Conference*, 2009, pp. 1136–1140.

[10] L. Still, M. Oispuu, and W. Koch, "Accuracy Study on Target Localization Using Acoustic Bearing Measurements Including Urban Reflections," in *2022 25th International Conference on Information Fusion (FUSION)*, 2022, pp. 1–8.

[11] F. Ribeiro, D. Ba, C. Zhang, and D. Florêncio, "Turning enemies into friends: Using reflections to improve sound source localization," in *2010 IEEE International Conference on Multimedia and Expo*, 2010, pp. 731–736.

[12] J. Even, Y. Morales, N. Kallakuri, C. Ishi, and N. Hagita, "Audio ray tracing for position estimation of entities in blind regions," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 1920–1925.

[13] K. Takami, T. Furukawa, M. Kumon, and L. C. Mak, "Non-field-of-view indoor sound source localization based on reflection and diffraction," in *2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2015, pp. 59–64.

[14] C. Blondé-Weinmann, P. Hamery, D. Iacob, and M. Ospel, "Pertinence of finite element method for impulse localization in urban environments," in *Proceedings INTERNOISE*, Nantes, 2024.

[15] K. W. Lo and B. G. Ferguson, "Simultaneous classification and ranging of direct fire weapons using an asynchronous acoustic sensor network," in *2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2011, pp. 425–430.

[16] L. Still, M. Varela, W.-D. Wirth, and M. Oispuu, "Shooter Localization with a Microphone Array Based on a Linearly Modeled Bullet Speed," in *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2019, pp. 1–6.

[17] L. Still, M. Oispuu, and W. Koch, "Accuracy Study on Shooter Localization Using Incomplete Acoustic Measurements," in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 2020, pp. 1–8.

[18] L. Still and M. Oispuu, "Field Experiments on Shooter State Estimation Accuracy Based on Incomplete Acoustic Measurements," in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2020, pp. 121–126.

[19] L. Still, M. Oispuu, and W. Koch, "Optimal Sensor Placement for Shooter Localization Using a Genetic Algorithm," in *2021 IEEE 24th International Conference on Information Fusion (FUSION)*, 2021, pp. 1–8.

[20] ——, "Optimal Sensor Placement for Shooter Localization within a Surveillance Area," in *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2021, pp. 1–6.

[21] E. Skudrzyk, *The foundations of acoustics: basic mathematics and basic acoustics*. Springer Science & Business Media, 2012.

[22] V. Pereyra, "Ray tracing methods for inverse problems," *Inverse Problems*, vol. 16, no. 6, p. R1, 2000.

[23] H. Rafat and N. Holger, "An ANSI C implementation of decimated and undecimated 1D Fast Discrete Wavelet Transforms," 2012.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNET classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[25] A. Ziemann, M. Barth, and M. Hehn, "Experimental investigation of the meteorologically influenced sound propagation through an inhomogeneous forest site," *Meteorologische Zeitschrift*, vol. 22, no. 2, pp. 221–229, 2013.

[26] L. C. Sutherland and G. A. Daigle, "Atmospheric sound propagation," *Handbook of acoustics*, vol. 28, pp. 305–329, 1998.

[27] K. Attenborough and T. Van Renterghem, *Predicting outdoor sound*. CRC Press, 2021.

[28] M. J. Crocker, *Handbook of noise and vibration control*. John Wiley & Sons, 2007.

[29] H. Bass, H.-J. Bauer, and L. Evans, "Atmospheric absorption of sound: Analytical expressions," *The Journal of the Acoustical Society of America*, vol. 52, no. 3B, pp. 821–825, 1972.

[30] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, ser. Addison Wesley series in artificial intelligence. Addison-Wesley, 1989.

[31] S. Xu, M. Rice, and F. Rice, "Optimal TOA-Sensor Placement for Two Target Localization Simultaneously Using Shared Sensors," *IEEE Communications Letters*, vol. 25, no. 8, pp. 2584–2588, 2021.