Security
of Infocommunication Systems
and Internet of Things

2026 Vol 4, No 1

https://doi.org/10.31861/sisiot2026.1.01022

Received 11 June 2026; revised 26 June 2026; accepted 29 June 2026; published 30 June 2026

Aggregation of Conflicting Data in Neural Network-based Sound Source Localization

Olena Kapustian and Bogdan Levchenko*

Department of System Analysis and Decision-Making Theory, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

*Corresponding author (E-mail: bogdan.levchenko.2016@gmail.com)

ABSTRACT Sound source localization is an important task today, especially in the context of the active development of unmanned aerial vehicles. The rapid spread and cheapening of their technologies have created a critical need for the development of reliable detection and tracking systems. Although deep learning-based approaches demonstrate higher accuracy compared to classical methods in complex conditions with strong noise or reverberation, sound source location data obtained from different microphone arrays can be incomplete and contradictory. Given this, there is an urgent need to apply aggregation methods to combine these spatial probability distributions into a single consistent estimate, which will significantly increase the reliability of acoustic detection networks. This study focuses on solving the problem of determining the azimuth of a drone based on multichannel audio recordings. To achieve this goal, a convolutional neural network based on the ResNet-18 architecture was adapted, which is capable of processing spatial matrices generated using the generalized cross-correlation with phase transform method. The classification model achieved a high accuracy rate of 93.75%, confirming the effectiveness of using cross-correlation matrices as input data for spatial localization tasks. To solve the problem of conflicting predictions from independent expert systems, the following probability aggregation methods were tested in this study: linear, geometric, and multiplicative opinion pooling functions, the Dempster-Shafer rule of combination of evidence, and the Dezert-Smarandache proportional conflict redistribution rule. The experimental results showed that while linear pooling effectively preserves the multimodality of distributions, it may not assign the highest probability to the correct spatial sector in situations where the peak predictions of individual sources do not match. On the other hand, the multiplicative rule and the Dempster-Shafer rule successfully isolate a single maximum value during aggregation, but the latter can lead to false-positive conclusions when working with highly conflicting data. The Dezert-Smarandache method proved to be the most reliable solution for environments characterized by strong acoustic noise and correspondingly high model uncertainty, as it proportionally redistributes the conflicting mass without giving premature preference to a single incorrect class. Furthermore, geometric pooling demonstrated significant practical value by finding a compromise between competing estimates and providing better computational efficiency in scenarios with a moderate level of uncertainty. The obtained results form comprehensive recommendations for selecting appropriate data aggregation strategies adapted to specific environmental conditions.

KEYWORDS sound source localization, aggregation, neural network, opinion pooling function.

I. INTRODUCTION

Today, the rapid development and cheapening of technologies have led to the widespread use of unmanned aerial vehicles (UAVs) in various spheres of life. They are actively used for both civilian and military purposes. Drones are capable of conducting aerial reconnaissance, cartography, infrastructure monitoring, strike, rescue operations and many other missions. This has created a critical need for the development of reliable systems for their detection and tracking. At the moment, localization systems based on acoustic signals are actively developing. Their advantages are cheaper equipment compared to radar and the ability to detect targets at low altitudes.

Existing approaches to solving the problem of acoustic localization can be divided into 2 groups: classical methods and deep learning methods [1]. Classical methods include the least squares method [2], triangulation [3], Expectation-Maximization (EM) algorithm for localization under conditions of uncertainty [4], The MUltiple SIgnal Classification (MUSIC) algorithm [5], Generalized Cross-Correlation with Phase Transform (GCC-PHAT) method for estimating the time delay between signals [6]. Despite the active development of modern approaches and machine learning models, classical algorithms are still widely used. One of their advantages over deep learning methods is the ability to mathematically take into account the physical features of sound propagation in space. Also, such algorithms can be simpler and better interpreted than machine learning approaches [1].

Deep learning methods include the use of convolutional and recurrent neural networks [7, 8], residual neural networks [9], attention-based neural networks [10]. The advantages of using these methods over classical ones are their flexibility, accuracy in complex conditions, scalability and speed of decision-making. In the case of strong noise (for example, wind and vehicle traffic) or reverberation, neural networks have better performance than traditional algorithms [10].

Location data of the same sound source can come from different places. It can be incomplete and contradictory. Because of this, there is a need to apply methods of aggregation of location data of the sound source, in order to combine them. This will increase the reliability of acoustic detection systems, as it will allow decisions to be made not only based on evidence from one microphone array, but from several, located in different places and under different conditions.

II. PROBLEM STATEMENT

The problem of sound source localization consists of determining its spatial coordinates based on the received sound signal. Most often, polar coordinates are determined: azimuth, distance, and elevation angle [1].

- Azimuth is the angle between the direction to the north and the direction to the sound source.

- Distance is the range from the microphone to the sound source.

- Elevation angle is the angle between the horizontal plane and the direction to the target. Sometimes, the height of the source above the horizon is determined instead.

Determining the coordinate of a sound source can be considered as a task of predicting either a continuous or a discrete random variable. In the case of a discrete random variable, the space of possible coordinate values is divided into distinct regions [10]. In such a case, it is treated as a classification problem to identify the specific region in which the target sound source is located. When multiple microphone arrays or algorithms process audio data from the exact same source, they can generate their own probability distribution regarding its location.

Let there be N models or expert systems, each of which estimates the probability distribution of some discrete random variable θ (in our case this is one of the coordinates of the sound source). $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ - set of possible values of θ. $(P_1, P_2, \dots, P_n)$, $n = 1, \dots, N$ – vector of distribution functions that the models predict. The task is to construct a function:

$$P_{P_1,\dots,P_n}(P_1, \dots, P_n): [0,1]^n \to [0,1], \quad (1)$$

that would aggregate N distributions into one.

In this work, the azimuth of the drone is considered as a random variable θ. The objective is to extract features from a multichannel audio signal of a drone recording, build a neural network model to determine the azimuth, and investigate the use of various methods for aggregating conflicting probability distributions from several expert systems into a single consistent estimate.

III. FEATURE EXTRACTION USING GCC-PHAT ALGORITHM

For sound source localization, algorithms most often use the following parameters [1]:

- Time Difference of Arrival (TDOA) – a parameter that indicates the difference in the arrival time of a sound signal between individual microphones.

- Direction of Arrival (DOA) – the angle at which the sound signal hits the microphone array.

- Sound pressure – the pressure created by the sound signal from the source acting on the microphone. Unlike the previous parameters, this indicator does not require time synchronization between microphones, but it may be less informative due to the presence of noise coming from other sound sources.

- Signal-to-noise ratio (SNR) – a ratio that shows how heavily the signal is affected by noise. A high SNR corresponds to a cleaner sound, while a low SNR indicates an audio signal with significant noise present. The lower the SNR, the less accurate the localization algorithms can be.

One of the popular algorithms for TDOA estimation is the Generalized Cross-Correlation with Phase Transform (GCC-PHAT) method [11]. It determines the difference in the time of arrival of sound between two or more microphones.

Let there be 2 sound signals at different microphones:

$$x_1(t) = s(t) + n_1(t), \quad (2)$$

$$x_2(t) = \alpha s(t - D) + n_2(t), \quad (3)$$

where $x_1(t)$ and $x_2(t)$ are the first and second signals respectively, $n_1(t)$ and $n_2(t)$ are the noise acting on each of the microphones, α is the amplitude attenuation coefficient of the second signal due to the difference in distance, and D is the time difference of arrival of the signal between the microphones.

Let's move from time t to the signal frequency using the Fourier transform:

$$X_1(\omega) = S(\omega) + N_1(\omega), \quad (4)$$

$$X_2(\omega) = \alpha S(\omega - D)e^{-j\omega\tau} + N_2(\omega). \quad (5)$$

To estimate the value of D, a cross-correlation function is used, which is defined as follows for the time and frequency domains, respectively:

$$R_{x_1,x_2}(\tau) = \mathbb{E}(x_1(t) \cdot x_2(t+\tau)) \quad (6)$$

$$R_{x_1,x_2}(\omega) = X_1(\omega) \cdot {X_2}^*(\omega) = \alpha|S(\omega)|^2 e^{j\omega\tau}. \quad (7)$$

The value of τ that maximizes $R_{x_1,x_2}(\tau)$ is the TDOA estimate. However, the standard cross-correlation function is sensitive to environmental conditions (reverberation and noise) and the spectral characteristics of the audio signal. To solve this problem, a generalized cross-correlation function with phase transform is used. It is defined as follows:

$$\tilde{R}_{x_1,x_2}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{R_{x_1,x_2}(\omega)}{\left|R_{x_1,x_2}(\omega)\right|} e^{i\omega\tau} d\omega \quad (8)$$

Then

$$\tilde{\tau} = argmax\, \tilde{R}_{x_1,x_2}(\tau) \quad (9)$$

The advantage of the GCC-PHAT method is that, due to the normalization of the cross-correlation function's magnitude, the algorithm isolates those frequency components that have the greatest impact on the TDOA estimate [11]. As a result, accuracy is increased in difficult conditions with present noise and reverberation. That is why this algorithm is widely used to extract features from an audio file for localization using neural network models [9].

In our case, the input data for training the model are GCC-PHAT matrices, which consist of the calculated values of the cross-correlation function for individual time intervals. To do this, the audio file is divided into 30 ms fragments with 15 ms coverage, and for each such fragment, the cross-correlation value between the signals received by different pairs of microphones is calculated.

Vol 4, No 1, Paper 01022, pp. 1-8 (2026)

Since the dataset contains 8-channel audio files, there will be $\frac{8\cdot 7}{2} = 28$ such pairs. Thus, the input data of the model are GCC-PHAT matrices of dimension (28, 500), where 500 is the number of points for which the cross-correlation was calculated. An example of such a matrix is presented in Fig. 1.

```
[[-1.24375913e-02 -4.03500581e-03  1.60345547e-02 ... -6.65857736e-03
   4.77500056e-04  3.12924245e-03]
 [-1.69275105e-02  6.60000602e-03 -6.69963751e-03 ... -9.38320905e-03
  -1.57209616e-02  1.77849233e-02]
 [ 1.53957922e-02  3.48629383e-03  1.11413486e-02 ...  9.34705604e-03
  -1.70015711e-02 -7.69120781e-03]
 ...
 [ 1.38236741e-02  7.40628596e-03  1.07603865e-02 ... -8.14299856e-04
   3.20165581e-03 -1.12933088e-02]
 [ 1.01521555e-02  1.14960810e-02  6.41044555e-03 ... -5.68499742e-03
   8.01101979e-03  3.23821930e-03]
 [ 1.07296426e-02 -7.74123985e-03  8.86023976e-03 ... -2.40818597e-03
   1.23649763e-04 -4.25672019e-03]]
```

FIG. 1. Example of GCC-PHAT matrix.

IV. NEURAL NETWORK ARCHITECTURE

The UaVirBASE dataset [12] was selected to train the model. This open-source dataset consists of 132 8-channel audio files featuring sound recordings of a DJI Mavic 3 Cine drone from various locations relative to the microphone array. Each 40-second file represents a distinct drone position. The JSON file describing a given recording contains information about the drone's spatial coordinates (distance, height, azimuth), the weather conditions under which the experiment was conducted, the coordinates of the microphone array, and other supplementary data. A partial example of the JSON file is presented in Fig. 2.


```
{
    "audio_interface_name": "Behringer UMC1820",
    "start_recording_time": "2024-11-15 09:40:02.379",
    "end_recording_time": "2024-11-15 09:40:42.468",
    "additional_info": "",
    "weather_data": {
        "timestamp": "2024-11-15T10:40:20.009000+00:00",
        "measurements": {
            "air temperature": "4.7 C",
            "air humidity": "90 % RH",
            "light intensity": "14823 Lux",
            "barometric pressure": "100810 Pa",
            "wind direction": "230 degrees",
            "wind speed": "1 m/s",
            "rainfall hourly": "0 mm/hour",
            "uv index": 1.1
        }
    },
    "microphone_array": {
        "center_latitude": "52.25532752777778",
        "center_longitude": "20.90053722222222",
        "microphones": [
            {
                "channel": 1,
                "name": "NTG-2",
                "latitude": "52.25534297880067",
                "longitude": "20.90053722222222",
                "height": "1.49",
                "elevation": "10",
                "azimuth": "0.0",
                "distance": "1.72"
            },
            {
                "channel": 2,
                "name": "NTG-2",
                "latitude": "52.25532752777778",
                "longitude": "20.900562463062652",
                "height": "1.49",
                "elevation": "10",
                "azimuth": "90.0",
                "distance": "1.72"
            },
                -
            },
            "drone": {
                "sound_source": "Drone",
                "type": "DJI Mavic 3 cine",
                "movement": "Static",
                "distance": "20",
                "height": "10",
                "azimuth": "0",
                "rotation": "Front",
                "additional_info": ""
            }
```


FIG. 2. Example of JSON file in dataset.

Fig. 3 illustrates the distribution of possible azimuth values where the drone was positioned during the recording.

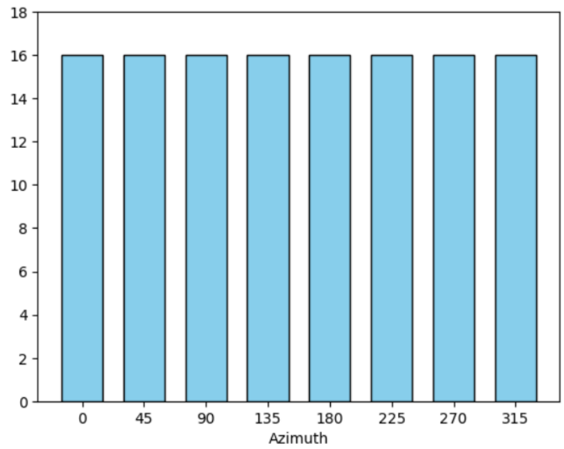

FIG. 3. Possible values of azimuth in dataset.

As can be seen from the distribution, there are 16 drone recordings for each specific azimuth value at 45-degree intervals: 0, 45, 90, 135, 180, 225, 270 and 315.

ResNet-18 was chosen as the base neural network model. It is a residual convolutional neural network consisting of 18 layers. It provides good results in classification tasks for matrix-format data. The first layer of the network was modified to accept single-channel spatial matrices instead of standard color images. At the output, the model produced a probability distribution of the drone being located at one of 8 azimuth values.

The model was trained with the following parameters:
- Number of epochs: 30;
- Batch size: 16;
- Optimizer: Adam;
- Learning rate: 0.001;
- Loss Function: CrossEntropyLoss.

The architecture of the model is presented in Fig. 6.

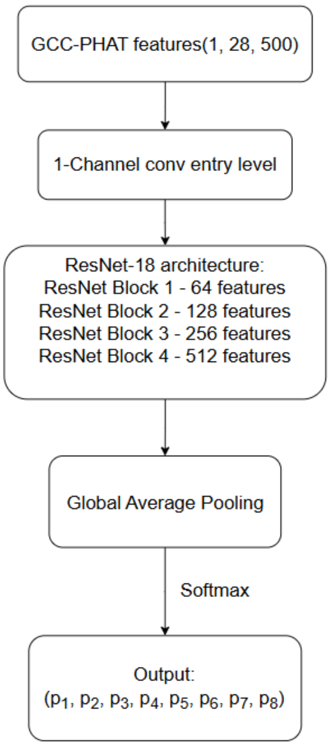

FIG. 4. Architecture of neural network.

V. AGGREGATION METHODS

The choice of a specific function from Eq. (1) depends on the desired set of axioms that must be satisfied [13]. Let us consider the main axioms used.

Indifference preservation. If $P_1(\omega_1) = \ldots = P_1(\omega_m), \ldots, P_n(\omega_1) = \ldots = P_n(\omega_m)$, then $P_{P_1,\ldots,P_n}(\omega_1) = \ldots = P_{P_1,\ldots,P_n}(\omega_m)$. In this case, if each expert system determines the distribution of θ to be uniform, then the aggregated distribution will also be uniform.

Consensus preservation. If for a specific $\omega \in \Omega$ $P_1 = \ldots P_n = P$, then $P_{P_1,\ldots,P_n} = P$. In this case, if all expert systems assign the same probability to a certain value of a random variable, then the aggregated probability estimate will coincide with them.

Independence of events. The aggregated probability estimation of a specific $\omega \in \Omega$ is expressed as a function of $P_1(\omega), \ldots, P_n(\omega)$. This axiom states that the combined probability of an individual value of a random variable depends only on the individual probability estimates of this value by the expert systems, and does not depend on other values of this random variable.

External Bayesianity. For new information L $P_{P_1^L,\ldots,P_n^L} = (P_{P_1,\ldots,P_n})^L$. This axiom states that in the presence of new information for all expert systems, it makes no difference which action is performed first: updating their own individual estimates and then aggregating, or first aggregating based on previous estimates and then updating the final estimate. In both cases, the combined probability distribution will be the same.

Individual-wise Bayesianity. For new information $P_{P_1,\ldots,P_i^L\ldots,P_n} = (P_{P_1,\ldots,P_n})^L$. This axiom is similar to the previous one and states that in the presence of new information for only an individual expert system, it makes no difference which action is performed first: updating the individual estimate and then aggregating, or first aggregating all distributions and then applying the new information [13].

The simplest opinion pooling function is the linear opinion pool. It is defined as follows:

$$P_{P_1,\ldots,P_n}(\omega) = \sum_{n=1}^{N} w_n P_n(\omega) \tag{10}$$

This method calculates the weighted arithmetic mean of the probability estimates for each value $\omega \in \Omega$. This is the only function that simultaneously satisfies the axioms of consensus preservation and independence of events. It also satisfies the axiom of indifference preservation [14].

Geometric opinion pool:

$$P_{P_1,\ldots,P_n}(\omega) = \frac{\prod_{n=1}^{N} P_n(\omega)^{w_n}}{\sum_{\omega'\in\Omega} \prod_{n=1}^{N} P_n(\omega')^{w_n}} \tag{11}$$

This function calculates the weighted geometric mean of the probability estimates and normalizes the obtained estimates. Since the aggregation of the probability estimate for an individual value $\omega \in \Omega$ uses probabilities for other values, the geometric pool does not satisfy the axiom of independence of events. However, it satisfies the axioms of indifference preservation, consensus preservation and external Bayesianity [14].

Multiplicative opinion pool:

$$P_{P_1,\ldots,P_n}(\omega) = \frac{\prod_{n=1}^{N} P_n(\omega)}{\sum_{\omega'\in\Omega} \prod_{n=1}^{N} P_n(\omega')} \tag{12}$$

This function is equivalent to the geometric pool in the case where $w_1 = \ldots = w_n = 1$. This is the only function that simultaneously satisfies the axioms of uniform distribution preservation and internal Bayesianity [14].

When choosing a specific aggregation function, it is necessary to determine the list of required axioms that it must satisfy.

Unlike opinion pooling functions, which operate under conditions where the model estimates a probability distribution, there are also aggregation methods that take into account the model's level of ignorance (when it cannot determine which class an object belongs to). The most popular methods are the Dempster-Shafer rule of combination of evidence and the Dezert-Smarandache proportional conflict redistribution rule. The Dempster-Shafer theory uses the concept of a mass function:

$$m: 2^{\Omega} \to [0, 1] \tag{13}$$

This function assigns a number from 0 to 1 to each subset of the set of possible states Ω, where $m(\emptyset) = 0$, $\sum_{A\in 2^{\Omega}} m(A) = 1$.

For aggregating data from multiple sources, a rule of combination is used. For a certain hypothesis A, it has the following form [15]:

$$\begin{gathered} m(A) = \frac{1}{1-K} \sum_{X\cap Y=A} m_1(X)\, m_2(Y), \\ \mathrm{K} = \sum_{X\cap Y=\emptyset} m_1(X) m_2(Y), \end{gathered} \tag{14}$$

where X is the evidence of the first model, and Y is the evidence of the second model. The number K determines the measure of conflict between the sources. The higher the value of K, the more the models contradict each other. A disadvantage of this method is that with a high degree of conflict, it sometimes yields false results. When 2 models each output 2 hypotheses, among which 2 are highly conflicting, applying the rule of combination increases the probability of a third hypothesis. This can lead to false-positive conclusions [15].

The Dezert-Smarandache theory is a modification of the Dempster-Shafer theory. It was created to solve the problem of high contradiction in the data, when the rule of combination does not yield the desired result. The Dezert-Smarandache rule redistributes the conflicting mass among the hypotheses of each model [16]:

$$\begin{aligned} m_{PCR5}(X) &= \\ &= m(X) \\ &+ \sum_{Y\in\Omega, X\cap Y=\emptyset} \Big(\frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} \\ &+ \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)}\Big) \end{aligned} \tag{15}$$

VI. RESULTS

Fig. 5 shows the testing results of the trained neural network in the form of a confusion matrix.

Since the dataset is balanced, the model's performance is evaluated using the percentage of correct answers. It amounted to 93.75%.

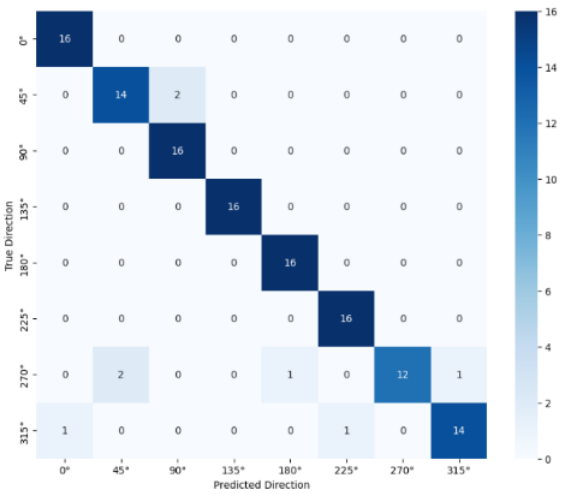

FIG 5. Confusion matrix.

To apply the aggregation methods, various recordings of the drone located at the same azimuth were selected. First, the pooling of distributions that assign the maximum probability to the exact same class was tested. After running these audio files through the neural network, the following results were obtained:

P_1: $(0.0043, 0.0062, 0.0772, 0.8988, 0.0049, 0.0005,$ $0.0002, 0.008)$

P_2: $(0.0125, 0.0048, 0.0056, 0.8895, 0.0058, 0.0006,$ $0.0006, 0.0807)$.

For both recordings, the model predicted the highest probability for class 4 (135 degrees azimuth). The Table 1 presents the aggregation results using linear, geometric, and multiplicative pooling. For geometric and multiplicative pooling, the weights $w_1 = w_2 = 0.5$ were used.

As can be seen from the distribution, linear pooling retained the highest probability for class 4. However, this probability does not exceed the maximum of the two distributions. The next largest are classes 8 and 3, corresponding to the second highest probabilities for distributions 2 and 1, respectively. Geometric pooling assigned a higher probability to class 4 than linear pooling did. At the same time, similarly to linear pooling, it left classes 8 and 3 as the next highest in magnitude, albeit with smaller values. Multiplicative pooling assigns the maximum probability to class 4, which turned out to be close to 1. All other probabilities were estimated by this method to be close to zero.

TABLE 1. Comparison of opinion pooling functions when the model predicts the same class.

P_1	P_2	Linear pooling	Geometric pooling	Multipli-cative pooling
0.0043	0.0125	0.0084	0.00764	0.000067
0.0062	0.0048	0.0055	0.00569	0.000037
0.0772	0.0056	0.0414	0.02167	0.000540
0.8988	0.8895	0.89415	0.93202	0.998513
0.0049	0.0058	0.00535	0.00556	0.000035
0.0005	0.0006	0.00055	0.00057	0
0.0002	0.0006	0.0004	0.00036	0
0.008	0.0807	0.04435	0.02649	0.000806

To test the Dempster-Shafer and Dezert-Smarandache aggregation methods, these distributions were slightly modified to account for the model's level of ignorance. Let's keep only the 2 classes with the highest probability: A_1, A_2 (in our case, these are 135 and 90 degrees in azimuth), and the total probability for the other classes will be considered the measure of the model's ignorance Ω. The measure of conflict K for these distributions amounted to 0.08. In this case, the aggregation results are presented in the Table 2.

TABLE 2. Comparison of Dempster-Shafer and Dezert-Smarandache aggregation methods when the model predicts the same class.

P_1	P_2	Dempster-Shafer	Dezert-Smarandache
0.0772	0.0072	0.00090405	0.00641
0.8988	0.9915	0.99906192	0.99356
0.024	0.0013	0.00003402	0.00003

The first 2 rows of the table correspond to the probabilities for A_1, A_2 and the last one for measure of the model's ignorance Ω. As we can see, the Dempster-Shafer rule of combination of evidence assigned a probability of almost one to the class corresponding to 135 degrees, and all other probabilities are correspondingly close to 0. This result is similar to the results of the multiplicative rule. The Dezert-Smarandache method redistributed the conflict mass more proportionally between A_1 and A_2, and accordingly, the probability of the 2nd class became slightly smaller compared to the Dempster-Shafer rule.

Let's apply these same methods to audio files that represent the same drone at the same azimuth, but for which the model assigns the maximum probability to different classes. The distributions have the following form:

P_1: $(0.0007, 0.0016, 0.0003, 0.4014, 0.5951, 0.0007,$ $0.0001, 0.0001)$

P_2: $(0.0514, 0.187, 0.1613, 0.3408, 0.1983, 0.025,$ $0.0321, 0.0041)$.

The first probability distribution assigns the highest probability to class 5 (180 degrees azimuth), and the second to class 4 (135 degrees). The true class value is 4. The aggregation results of the distributions in this case are shown in the Table 3.

By analyzing the table, we can conclude that linear pooling assigned the highest probability to class 5, while geometric and multiplicative – to class 4. This is due to the

TABLE 3. Comparison of opinion pooling functions when the model predicts the different classes.

P_1	P_2	Linear pooling	Geometric pooling	Multipli-cative pooling
0.0007	0.0514	0.02605	0.00799	0.00014
0.0016	0.187	0.0943	0.02305	0.00117
0.0003	0.1613	0.0808	0.00927	0.00019
0.4014	0.3408	0.3711	0.49298	0.53602
0.5951	0.1983	0.3967	0.45788	0.4624
0.0007	0.025	0.01285	0.00558	0.00007
0.0001	0.0321	0.0161	0.00239	0.00001
0.0001	0.0041	0.0021	0.00085	0

fact that the probability of class 4 in both distributions differs less strongly from each other than for class 5. Geometric and multiplicative pooling generally yielded similar results, but multiplicative assigns low probability values to other classes besides the maximum one.

To test the Dempster-Shafer and Dezert-Smarandache methods, let's select distributions with a high level of conflict (K = 0.994):

P_1: (0.0007, 0.0007, 0.0003, 0.0008, 0.9965, 0.0007, 0.0002, 0.0001)

P_2: (0.0003, 0.0001, 0.0002, 0.9973, 0.0008, 0.0003, 0.0005, 0.0005).

The first distribution assigns the highest probability to class 5 (180 degrees azimuth), and the second distribution to class 4 (135 degrees azimuth). In the same way, we modify these distributions, leaving the 2 classes with the highest probability. Then the aggregation results will look like those presented in the Table 4.

TABLE 4. Comparison of Dempster-Shafer and Dezert-Smarandache aggregation methods when the model predicts the different classes.

P_1	P_2	Dempster-Shafer	Dezert-Smarandache
0.0008	0.9973	0.56415	0.50064
0.9965	0.0008	0.43502	0.49935
0.0027	0.0019	0.00083	0.00001

The Dempster-Shafer rule assigned a higher probability to class A_2 and left a larger measure of model ignorance, while the Dezert-Smarandache method proportionally redistributed this measure of ignorance and assigned almost identical probability values to each class. In this case, the Dezert-Smarandache method is a more effective solution because, given the strong conflict between the two distributions, it does not strongly favor one class, but leaves them almost equally probability. This helps to avoid false-positive results more often.

Based on these experiments, it can be concluded that the linear pooling method is best suited for situations where it is necessary to preserve multimodality and not strongly isolate a single value of the distribution. The multiplicative method and Dempster-Shafer theory, on the contrary, can be useful in cases where it is necessary to isolate a single maximum value during aggregation. However, the Dempster-Shafer theory can give unsatisfactory results when the models strongly contradict each other. The Dezert-Smarandache theory yields similar results to the Dempster-Shafer theory, but it aggregates highly conflicting data better. Therefore, it is advisable to use it in the case of strong noise, when the model is uncertain about its predictions. The advantage of geometric pooling is that it seeks a compromise between models, highlighting one of them, rather than simply averaging like linear pooling. Moreover, it does not lower the probability for other values, besides the most probable one, as drastically as multiplicative pooling does.

VII. CONCLUSION

In this work, a convolutional neural network model was trained based on GCC-PHAT features for the task of predicting the azimuth at which a drone is located, and the following methods of aggregating conflicting data in the form of probability distributions were tested: linear, logarithmic, multiplicative pooling, the Dempster-Shafer rule of combination of evidence, and the Dezert-Smarandache proportional conflict redistribution rule.

The model demonstrated good results in the form of a 93,75% proportion of correct answers. This confirms the effectiveness of using GCC-PHAT features as input data for the model for spatial localization.

To increase the reliability of determining the sound source coordinates, various methods for aggregating probability distributions were applied. Two drone recordings were selected as conflicting data; in each of them, the drone has the same azimuth value, but the model predicts different distributions.

The simplest aggregation method turned out to be linear pooling, which leaves the distribution multimodal. However, as the second application example showed, under conditions where the states with the maximum probability in both distributions do not match, the maximum value in the aggregated distribution may belong to the incorrect class. Geometric and multiplicative rules assigned the maximum probability to the correct class in both cases. At the same time, multiplicative pooling assigned a high probability value to the correct class, while all other values became close to zero. Geometric pooling also significantly increased the probability of the correct class, but it did not decrease the values for the other classes as drastically.

The Dempster-Shafer rule of combination of evidence and the Dezert-Smarandache proportional conflict redistribution rule correctly assigned the maximum probability to the target class in both cases. However, in the case of highly conflicting distributions, the Dempster-Shafer method placed a greater emphasis on one of them, while the Dezert-Smarandache rule assigned almost equal probability to both classes. In real-world scenarios, leaving the probabilities equal can be more practical, since in this case we do not simply trust one of the models, but rather understand the need to additionally verify the evidence of each one and only then make a decision on the result.

The choice of a specific data aggregation method heavily depends on the particular situation and the conditions in which the microphone arrays are operating. In environments with constant interference and noise, an effective solution can be the application of the Dezert-Smarandache proportional conflict redistribution rule, which demonstrated good results under conflicting conditions. Meanwhile, geometric pooling, due to its better computational efficiency, can be effective in cases of lower uncertainty.

To improve the system in the future, new data aggregation rules can be applied and tested under conditions of strong contradiction. Since all drone movement recordings in the dataset are static, there is currently no ability to track its coordinates in motion. In the future, using mathematical modeling methods and the available sound data, it will be possible to simulate the acoustic propagation signal of a moving drone. This will help improve the model and make it more practical.

AUTHOR CONTRIBUTIONS

B.L. – investigation, writing-original draft preparation, methodology, formal analysis; O.K. – supervision, writing-review, editing.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] G. Jekateryńczuk and Z. Piotrowski, “A Survey of Sound Source Localization and Detection Methods and Their Applications,” *Sensors*, vol. 24, no. 1, Art. no. 68, 2024. doi: 10.3390/s24010068.

[2] W. Li, B. Tian, Z. Chen, and S. Xu, “Regularized Constrained Total Least Squares Source Localization Using TDOA and FDOA Measurements,” *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024. doi: 10.1109/LGRS.2023.3340440.

[3] A. K. Agrawal, “Localization Using Angle-of-Arrival (AOA) Triangulation,” in *Proceedings of the International Workshop on Environmental Sensing Systems for Smart Cities*, pp. 14–19, 2025. doi: 10.1145/3742460.3742984.

[4] X. Wang, B. Quost, J.-D. Chazot, and J. Antoni, “Sound Source Localization from Uncertain Information Using the Evidential EM Algorithm,” in *Scalable Uncertainty Management*, 2013, pp. 162–175. doi: 10.1007/978-3-642-40381-1_13.

[5] X. Cao, H. Liu, and S. Wu, “DOA Estimation Based on Online Music Algorithm,” *Journal of Electronics & Information Technology*, vol. 30, no. 11, pp. 2658–2661, 2008. doi: 10.3724/SP.J.1146.2007.00565.

[6] S. Chakrabarty and E. A. Habets, “Multi-Speaker DOA Estimation Using Deep Convolutional Networks Trained With Noise Signals,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 8–21, 2019. doi: 10.1109/JSTSP.2019.2901664.

[7] V. S. Suruthhi, V. Smita, J. R. Gini, and K. I. Ramachandran, “Detection and Localization of Audio Event for Home Surveillance Using CRNN,” *International Journal of Electronics and Telecommunications*, vol. 67, no. 4, pp. 735–741, 2021. doi: 10.24425/ijet.2021.139771.

[8] Y. Zhao, Y. He, H. Chen, Z. Zhang, and Z. Xu, “Three-Dimensional Grid-Free Sound Source Localization Method Based on Deep Learning,” *Applied Acoustics*, vol. 227, Art. no. 110261, 2025. doi: 10.1016/j.apacoust.2024.110261.

[9] G. Liu, Y. Liu, and X. Hong, “Audio Event Recognition by Multitask Learning of Audio Attribute Classification,” in *2021 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*, pp. 8–15, 2021. doi: 10.1109/IC-NIDC54101.2021.9660525.

[10] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, “A Survey of Sound Source Localization With Deep Learning Methods,” *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 107–151, 2022. doi: 10.1121/10.0011809.

[11] “Gccphat – generalized cross-correlation,” *MATLAB Help Center*. [Online]. Available: https://www.mathworks.com/help/phased/ref/gccphat.html

[12] G. Jekateryńczuk, R. Szadkowski, and Z. Piotrowski, “UaVirBASE: A Public-Access Unmanned Aerial Vehicle Sound Source Localization Dataset,” *Applied Sciences*, vol. 15, no. 10, Art. no. 5378, 2025. doi: 10.3390/app15105378.

[13] L. Elkin and R. Pettigrew, *Opinion Pooling*. Cambridge, U.K.: Cambridge University Press, 2024. doi: 10.1017/9781009315203.

[14] F. Dietrich and C. List, “Probabilistic Opinion Pooling,” in *The Oxford Handbook of Probability and Philosophy*, A. Hájek and C. Hitchcock, Eds. Oxford, U.K.: Oxford University Press, 2016, pp. 519–542. doi: 10.1093/oxfordhb/9780199607617.013.37.

[15] G. Shafer, “Dempster’s Rule of Combination,” in *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton University Press, 1976, pp. 57–73. doi: 10.2307/j.ctv10vm1qb.7.

[16] J. Dezert and F. Smarandache, “An Introduction to DSmT for Information Fusion,” *New Mathematics and Natural Computation*, vol. 8, no. 3, pp. 343–359, 2012. doi: 10.1142/S179300571250007X.

Olena Kapustian

Head of the Department of System Analysis and Decision-Making Theory, Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Dr. Sci. (Phys&Math), Senior Researcher. Research interests: optimal control theory, decision-making theory under uncertainty, minimax estimation theory, systems analysis.

ORCID ID: 0000-0002-2629-0750

Bogdan Levchenko

Graduating in 2026 from the Taras Shevchenko National University of Kyiv with a master’s degree in "System Analysis". Research interests: artificial intelligence, decision-making systems and methods.

ORCID ID: 0009-0000-7109-4022

Агрегування суперечливих даних при нейромережевій акустичній локалізації джерела звуку

Олена Капустян, Богдан Левченко*

Кафедра системного аналізу та теорії прийняття рішень, Київський національний університет імені Тараса Шевченка, Київ, Україна

*Автор-кореспондент (Електронна адреса: bogdan.levchenko.2016@gmail.com)

АНОТАЦІЯ Локалізація джерела звуку є важливою задачею на сьогоднішній день, особливо в умовах активного розвитку безпілотних літальних апаратів. Стрімке поширення та здешевлення їх технологій створили критичну потребу в розробці надійних систем виявлення та відстеження. Хоча підходи на основі глибинного навчання демонструють вищу точність порівняно з класичними методами в складних умовах із наявністю сильного шуму або реверберації, дані про місцезнаходження джерела звуку, отримані з різних мікрофонних решіток, можуть бути неповними та суперечливими. З огляду на це виникає гостра необхідність у застосуванні методів агрегації для об'єднання цих просторових розподілів ймовірностей у єдину узгоджену оцінку, що дозволить значно підвищити надійність мереж акустичного виявлення. Це дослідження зосереджено на розв’язанні задачі визначення азимута безпілотника на основі багатоканальних звукових записів. Для досягнення цієї мети було адаптовано згорткову нейронну мережу на базі архітектури ResNet-18, яка здатна обробляти просторові матриці, згенеровані за допомогою методу узагальненої крос-кореляції з фазовим перетворенням. Модель класифікації досягла високого показника точності, який склав 93,75%, що підтверджує ефективність використання матриць взаємної кореляції як вхідних даних для задач просторової локалізації. Щоб вирішити проблему суперечливих прогнозів від незалежних експертних систем, у цьому дослідженні було протестовано наступні методи агрегації ймовірностей: лінійну, геометричну та мультиплікативну функцію об'єднання думок, правило комбінування свідчень Демпстера-Шейфера та правило пропорційного перерозподілу конфліктів Дезера-Смарандаке. Результати експериментів показали, що хоча лінійне об'єднання ефективно зберігає мультимодальність розподілів, воно може не призначити найвищу ймовірність правильному просторовому сектору в ситуаціях, коли пікові прогнози окремих джерел не збігаються. З іншого боку, мультиплікативне правило та правило Демпстера-Шейфера успішно виділяють єдине максимальне значення під час агрегації, проте остання може призводити до хибнопозитивних висновків під час роботи з даними, що мають високий ступінь конфлікту. Метод Дезера-Смарандаке виявився найбільш надійним рішенням для середовищ, що характеризуються сильним акустичним шумом і відповідно, значною невпевненістю моделі, оскільки він пропорційно перерозподіляє конфліктну масу, не надаючи передчасної переваги одному хибному класу. Крім того, геометричне об'єднання продемонструвало значну практичну цінність, знаходячи компроміс між конкуруючими оцінками та забезпечуючи кращу обчислювальну ефективність у сценаріях із помірним рівнем невизначеності. Отримані результати формують комплексні рекомендації щодо вибору відповідних стратегій агрегації даних, адаптованих до конкретних умов навколишнього середовища.

КЛЮЧОВІ СЛОВА локалізація джерела звуку, агрегування, нейронна мережа, функція об’єднання думок.