INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 4, PP. 855–859
Manuscript received August 19, 2024; revised October, 2024. doi: 110.24425/ijet.2024.152070

A Flexible Way of Coarse Coordinates Estimation for Sodars

Kamil Stawiarski

***Abstract*—The publication presents a flexible approach to implementing coarse coordinate estimation of an object observed with a sodar. This flexibility permits any arrangement of sound sources as well as microphones. Only minimal requirements are imposed on the probing signal, which can particularly be broadband. The algorithms have been tested on both synthetic data and data recorded with an actual device.**

***Keywords*—sodar, DOA estimation, channel estimation**

I. Introduction

SODARS (sonic detection and ranging) are devices similar to sonars, allowing for the detection and estimation of object parameters using acoustic waves, but with a distinction – sodars typically operate in the air. They are used wherever it is required to locate a sound source or determine the location of objects based on the emitted acoustic wave, for example, in alarms [1], robotics [2], or the study of atmospheric parameters [3]. In this regard, they are similar to radars but do not employ electromagnetic waves. All three devices have analogous signal processing. Usually, signal processing is divided into several consecutive steps: signal compression, beamforming, bandpass Doppler filtration, detection, and finally, coordinate estimation [4]. All of these steps are applicable to both pulsed and continuous wave devices. This sort of processing requires certain assumptions and simplifications, such as the narrowband character of the signal. The described approach fits both pulsed and continuous wave devices, allowing for the simplification of the beamforming process by expressing delay as a phase shift. This paper attempts to implement the most general method of processing and investigates whether current processing units are capable of real-time operation. The objective is to derive a universal formula for coordinate estimation in all three spatial dimensions – range, azimuth, and elevation. This approach enables flexibility in signal waveform and antenna pattern domains. The paper is organized as follows: Section 2 provides a theoretical description of the problem, including all mathematical derivations and signal processing descriptions. Assumptions regarding the device, waveform, and propagation path shape are also presented here. Section 3 delves the reader deeper into the implementation method, presents the parameters of the simulation, its conditions and the resulting results. Section 4 describes the operation of the algorithm using actual recorded data, including specifications of the actual device. Section 5 provides theoretical considerations on computational complexity, ways to reduce it, and implementation possibilities in real-time devices. Finally, in the last section, Section 6, a summary of the entire paper is provided, with emphasis on areas where the algorithm could be enhanced.

K. Stawiarski is with Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Department of Signals and Systems, Poland (e-mail: kamil.stawiarski@pg.edu.pl)

II. Theory

A. System description

At the outset, it is crucial to determine the overall shape of the device and the signal channel (path of propagation). The device is assumed to consist of an N-element microphone array, an M-element acoustic transducer array (array of speakers), and a K-element set of objects that disperse the soundreaching them in an omnidirectional manner. The general scheme is depicted in Fig. 1. It illustrates an example of a device constructed based on N microphones ($n_1 \sim n_4$), a single sound source ($M = 1$), and $K = 2$ point objects (k_1, k_2) being observed. The dotted lines represent the paths between the sound source and the observed objects, as well as between the observed objects and the microphones. The total number of signal paths is equal to $M \cdot K \cdot N$. While the microphones and loudspeakers in Fig. 1 are placed collinearly, this arrangement is not mandatory. In general, each object (microphone, speaker, or observed object) can have any position in space, which can be described using the three-dimensional Cartesian coordinate system. Accordingly, the positions of microphones ($\boldsymbol{P}_n$), sound sources ($\boldsymbol{P}_m$), and reflective objects ($\boldsymbol{P}_k$) are expressed in matrix form as follows:

$$\boldsymbol{P}_N = \begin{bmatrix} \boldsymbol{p}_{n,1} \\ \boldsymbol{p}_{n,2} \\ \vdots \\ \boldsymbol{p}_{n,N} \end{bmatrix} \quad \boldsymbol{P}_K = \begin{bmatrix} \boldsymbol{p}_{k,1} \\ \boldsymbol{p}_{k,2} \\ \vdots \\ \boldsymbol{p}_{k,K} \end{bmatrix} \quad \boldsymbol{P}_M = \begin{bmatrix} \boldsymbol{p}_{m,1} \\ \boldsymbol{p}_{m,2} \\ \vdots \\ \boldsymbol{p}_{m,M} \end{bmatrix}. \quad (1)$$

In the formula above, a single row of the matrix represents the coordinates of one microphone/speaker/reflective object.

$$\begin{aligned} \boldsymbol{p}_{n,n} &= \begin{bmatrix} n_{n,x} & n_{n,y} & n_{n,z} \end{bmatrix} \\ \boldsymbol{p}_{k,k} &= \begin{bmatrix} k_{k,x} & k_{k,y} & k_{k,z} \end{bmatrix} \\ \boldsymbol{p}_{m,m} &= \begin{bmatrix} m_{m,x} & m_{m,y} & m_{m,z} \end{bmatrix}. \end{aligned} \quad (2)$$

$\boldsymbol{P}_N$ and $\boldsymbol{P}_M$ are a known values, while the positions of the objects are unknown.

It's also assumed that the signal generated by each of the M speakers may be controlled separately. It's important that they

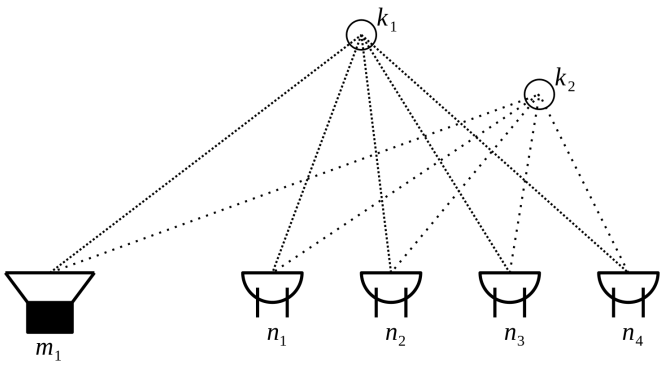

Fig. 1. Simplified schematic of a sodar with one sound source, four microphones, and two objects.

can generate mutually orthogonal signals, which is a crucial aspect for the operation of the entire system. Additionally, the signal at each of the microphones is directly sampled, and all processing is conducted in the digital domain.

B. Waveform

To avoid mutual interference of signals from different sound sources, it's crucial that each speaker emits a signal orthogonal to the others. In real-world scenarios, limited amplitude is also important to prevent power amplifier overdrive. These requirements are fulfilled by the entire family of CAZAC [5] (constant amplitude zero autocorrelation waveform) signals. The chosen waveform for this project is a signal with linearly modulated frequency. It can be expressed as:

$$s(t) = sin\left(t \cdot 2\pi(f_c - \Delta f + \Delta f \cdot t/T_L)\right). \tag{3}$$

In Eq. 3, t is a time, T_L represents the length of the LFM signal, f_c is the central frequency (which occurs at time $t = 0.5 \cdot TL$), and Δf is half of the frequency deviation. This form allows us to limit the signal frequency to the range $[f_c - \Delta f; f_c + \Delta f]$ with a bandwidth of $2\Delta f$. Furthermore, for a wideband signal, it can be assumed that two mutually time-shifted signals are orthogonal.

$$s(t) * s(t + \Delta t) = \begin{cases} 1, & \text{if } \Delta t = 0, \\ 0, & \text{if } \Delta t \neq 0. \end{cases} \tag{4}$$

This property allows achieving orthogonality by emitting signals with additional delays, different for each of the speakers (Δt_m). However, despite similar properties, noise should not be used as a signal. The lack of a deterministic form of the function would make it impossible, in its case, to determine the distance with greater precision than would result from the sampling frequency of the signal in the actual device.

C. Signal path

As depicted in Fig. 1, the signal emitted by the speaker m_m and reaches the object k_k, reflects off its surface, and then reaches each of the microphones. Consequently, the sound propagation time $\Delta t_{m,k,n}$ along the path of speaker m_m, object k_k, and microphone n_n can be calculated by:

$$\Delta t_{m,k,n} = \frac{d_{m,k,n}}{c}, \tag{5}$$

$$\begin{aligned} d_{m,k,n} = \left\|\boldsymbol{p}_{m,m} - \boldsymbol{p}_{k,k}\right\|_2 + \left\|\boldsymbol{p}_{k,k} - \boldsymbol{p}_{n,n}\right\|_2 = \\ \left\|\begin{bmatrix} m_{m,x} & m_{m,y} & m_{m,z} \end{bmatrix} - \begin{bmatrix} k_{k,x} & k_{k,y} & k_{k,z} \end{bmatrix}\right\|_2 + \\ \left\|\begin{bmatrix} k_{k,x} & k_{k,y} & k_{k,z} \end{bmatrix} - \begin{bmatrix} n_{n,x} & n_{n,y} & n_{n,z} \end{bmatrix}\right\|_2, \end{aligned} \tag{6}$$

where $\|_2 \cdot \|^2$ is the L2 norm operator and c is the signal propagation velocity (typically around 343 meters per second for sound in the air). Given the above relationships, the signal recorded by the n_n-th microphone ($w_n(t)$) can be expressed as a superposition of the signals reflected by the objects:

$$w_n(t) = \sum_{m=1}^{M} \sum_{k=1}^{K} a_{m,k,n} s\left(t - \Delta t_m - \Delta t_{m,k,n}\right) + \nu(t), \tag{7}$$

where $a_{m,k,n}$ is the unknown amplitude of the signal after passing through the path from speaker m to object k to microphone n, and $\nu(t)$ is additive Gaussian noise representing noise and interference in the system. Due to the known constant sampling period, the sampled signal in digital form at the n-th microphone can be expressed as a vertical vector $w_n[j]$ of length $S_L = T_L \cdot f_s$:

$$w_n[j] = \sum_{m=1}^{M} \sum_{k=1}^{K} a_{m,k,n} s\left(j/f_s - \Delta t_m - \Delta t_{m,k,n}\right) + \nu[j], \tag{8}$$

where j denotes the sample number and f_s is the sampling frequency. Equation 8 can be rewritten in matrix form as:

$$\boldsymbol{w}_n = \boldsymbol{S}_n \boldsymbol{a}_n + \boldsymbol{\nu}, \tag{9}$$

$$\boldsymbol{S}_n = \begin{bmatrix} \boldsymbol{s}_{1,1,n} & \cdots & \boldsymbol{s}_{M,1,n} & \boldsymbol{s}_{1,2,n} & \cdots \boldsymbol{s}_{M,K,n} \end{bmatrix},$$

$$\boldsymbol{a}_n = \begin{bmatrix} a_{1,1,n} \\ \vdots \\ a_{M,1,n} \\ a_{1,2,n} \\ \vdots \\ a_{M,K,n} \end{bmatrix},$$

where $\boldsymbol{s}_{m,k,n}$ is a vertical vector of values $s\left(j/f_s + \Delta t_m + \Delta t_{m,k,n}\right)$ representing the sample numbers. The signal recorded by a set of microphones can be represented by a matrix $\boldsymbol{W}$:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdots & \boldsymbol{w}_N \end{bmatrix}. \tag{10}$$

D. Algorithm

Due to convergence with the Cramér-Rao [6] constraint, the basic variant of the algorithm is based on the maximum-likelihood estimator, which under Gaussian data distribution is equivalent to the least squares method [7] and can be represented by the formula:

$$\hat{\boldsymbol{P}}_K = \underset{\tilde{\boldsymbol{P}}_K}{\operatorname{argmin}} \left(\sum_{n=1}^{N} \left\| \boldsymbol{w}_n - \boldsymbol{S}_{n,\tilde{\boldsymbol{P}}_K} \boldsymbol{a}_n \right\|^2 \right). \tag{11}$$

In Eq. 11, $\boldsymbol{S}_{n,\tilde{\boldsymbol{P}}_K}$ is the matrix of unweighted signals recorded by the n-th microphone when objects are placed in $\tilde{\boldsymbol{P}}_K$ coordinates. Assuming that the values of the vector $\boldsymbol{a}_n$ can be estimated using the Moore-Penrose inverse [8], the equation will take the form:

$$\hat{\boldsymbol{a}}_{n,\tilde{\boldsymbol{P}}_K} = \left(\boldsymbol{S}^T_{n,\tilde{\boldsymbol{P}}_K} \boldsymbol{S}_{n,\tilde{\boldsymbol{P}}_K} \right)^{-1} \boldsymbol{S}^T_{n,\tilde{\boldsymbol{P}}_K} \boldsymbol{w}_n, \tag{12}$$

$$\hat{\boldsymbol{P}}_K = \underset{\tilde{\boldsymbol{P}}_K}{\operatorname{argmin}} \left(\sum_{n=1}^{N} \left\| \boldsymbol{w}_n - \boldsymbol{S}_{n,\tilde{\boldsymbol{P}}_K} \hat{\boldsymbol{a}}_{n,\tilde{\boldsymbol{P}}_K} \right\|^2 \right). \tag{13}$$

In simpler terms, the mean-square error is calculated by comparing the recorded signal ($\boldsymbol{w}_n$) with the synthetic signal ($\boldsymbol{S}_{n,\tilde{\boldsymbol{P}}_K} \hat{\boldsymbol{a}}_{n,\tilde{\boldsymbol{P}}_K}$), that would be observed if the objects were located at $\tilde{\boldsymbol{P}}_K$ coordinates. With zero noise, the mean squared error will be smallest for estimated object coordinates that match the real ones ($\tilde{\boldsymbol{P}}_K = \boldsymbol{P}_K$). The function will have local extrema when the coordinates of one of the objects agree (assuming fewer object's coordinates are estimated than in reality) and on sidelobes (occurring due to some signal correlation).

The baseline version of the algorithm requires searching for a global extremum within $3K$ dimensions, leading to an exponential increase in computational complexity as the number of estimated coordinates increases.

However, if the signals are not highly correlated, it's observed that the vector $\hat{\boldsymbol{a}}_{n,\tilde{\boldsymbol{P}}_K}$ alone indicates the signal amplitude assuming the existence of an object at coordinates $\tilde{\boldsymbol{P}}_K$. If there is indeed an object there, the amplitude should be much higher than in the absence of the object. With this in mind, estimation can be performed by searching for the maximum amplitude. In other words, the vector $\hat{\boldsymbol{a}}_{n,\tilde{\boldsymbol{P}}_K}$ can be interpreted as a weight vector when decomposing the signal $\boldsymbol{w}_n$ using the $\boldsymbol{S}_{n,\tilde{\boldsymbol{P}}_K}$ function. To express Equations (11)-((13)) in a simpler, linear way, the data from N microphones must be combined into a single column vector:

$$\boldsymbol{W}^{\#} = \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_N \end{bmatrix}, \quad \boldsymbol{S}^{\#}_{\tilde{\boldsymbol{P}}_K} = \begin{bmatrix} \boldsymbol{S}_{1,\tilde{\boldsymbol{P}}_K} \\ \boldsymbol{S}_{2,\tilde{\boldsymbol{P}}_K} \\ \vdots \\ \boldsymbol{S}_{N,\tilde{\boldsymbol{P}}_K} \end{bmatrix}. \tag{14}$$

Using the above transformation, the vector $\hat{\boldsymbol{a}}_{\tilde{\boldsymbol{P}}_K}$ can be represented as:

$$\hat{\boldsymbol{a}}_{\tilde{\boldsymbol{P}}_K} = \left(\boldsymbol{S}^{\#T}_{\tilde{\boldsymbol{P}}_K} \boldsymbol{S}^{\#}_{\tilde{\boldsymbol{P}}_K} \right)^{-1} \boldsymbol{S}^{\#T}_{\tilde{\boldsymbol{P}}_K} \boldsymbol{W}^{\#}. \tag{15}$$

Due to the lack of strong correlation between signals from different objects, the estimation can be carried out in an iterative manner, adding more object coordinates in successive iterations. Additionally, Eq. 15 can be simplified. Assuming no correlation, the $\boldsymbol{S}^{\#T}_{\tilde{\boldsymbol{P}}_K} \boldsymbol{S}^{\#}_{\tilde{\boldsymbol{P}}_K}$ matrix is approximately a matrix of the form of the rescaled unit matrix:

$$\boldsymbol{S}^{\#T}_{\tilde{\boldsymbol{P}}_K} \boldsymbol{S}^{\#}_{\tilde{\boldsymbol{P}}_K} \approx \mathbf{1}\sigma_s \rightarrow \hat{\boldsymbol{a}}_{\tilde{\boldsymbol{P}}_K} \approx \sigma_s^{-1} \boldsymbol{S}^{\#T}_{\tilde{\boldsymbol{P}}_K} \boldsymbol{W}^{\#}, \tag{16}$$

where $\sigma_s = \|s_{m,k,n}\|^2$ denotes the identity matrix of appropriate size. This approximation is not recommended in situations of nearby object placement. Additionally, from a practical standpoint, it's beneficial to introduce additional biasing to the matrix in Eq. 14 during the initial iterations. This will expedite the convergence of the algorithms that locate the extrema by disrupting the correlation and, consequently, "blurring" the extremes of the weight vector into neighboring positions.

Once the extremes in the amplitude matrix have been detected, the coordinates can be refined using a realization of Eq. 13. The coordinate matrix $\tilde{\boldsymbol{P}}_K$ can therefore be interpreted as the initial values to be searched "on the grid" before the final coordinate estimation. Determining the number of objects in the observed signal, as well as discerning which extremes are actually from an object (and not false alarms), necessitates dedicated detection algorithms and is beyond the scope of this publication.

III. Implementation and Simulations

The algorithm was implemented using Octave as a single-threaded application.

A. Parameters of simulation

Operating parameters assumed during the simulation were as follows: estimation was conducted in 2D space (X, Y coordinates mapped into range, azimuth system, the plane $Z = 0$), with a single speaker $M = 1$ located at coordinates $P_{1,n} = [-0.05\ 0\ 0]$, and $N = 8$ microphones arranged in a uniform linear array with coordinates:

$$\boldsymbol{P}_N = \begin{bmatrix} 0 & 0 & 0 \\ d & 0 & 0 \\ \vdots & \vdots & \vdots \\ 7d & 0 & 0 \end{bmatrix}, \tag{17}$$

where $d = 0.01524$. All coordinates shown are expressed in meters. The length of the probing signal is $T_L = 30\ ms$ with the rest of its parameters: $f_s = 93.75\ kHz$, $f_c = 3.5\ kHz$, $\Delta f = 2.5\ kHz$. With such parameters, the signal covers the $5\ kHz$ frequency from $1\ kHz$ to $6\ kHz$ and can therefore be considered broadband. The speed of sound in the air was set at $c \approx 331.29\ m/s$. The simulation assumes $K = 3$ objects with coordinates:

$$\boldsymbol{P}_K = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.45\sqrt{2} & 0.45\sqrt{2} & 0 \\ -0.5 & 0.6 & 0 \end{bmatrix}. \tag{18}$$

Amplitude of each path was set as $a_{m,k,n} = 10$ for each signal path, and the noise variance was set as $\sigma_n^2 = 1$.

B. Simulation results

The concise determination of the amplitude based on Eq. 15 and Eq. 16 was done in distance-azimuth coordinates due to its better suitability for this type of equipment. Each coordinate was converted to Cartesian form. The coordinate matrix $\boldsymbol{P}_K$ consisted of XYZ coordinates obtained from the transformation of a set of angular coordinates with values $r = [0.2\ 0.22\ \cdots\ 1.2]\,[m]$, $\alpha = [-75\ -72.5\ \cdots\ 75]\,[^\circ]$ (combined on a one-to-one correspondence). In total, this translated into 3111 coordinates. Matrix of amplitudes is shown in the figures below:

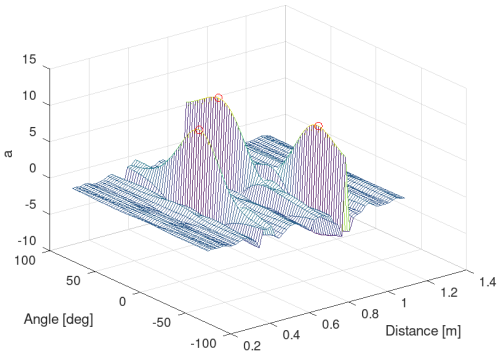

Fig. 2. Matrix of amplitudes obtained based on Eq. 15

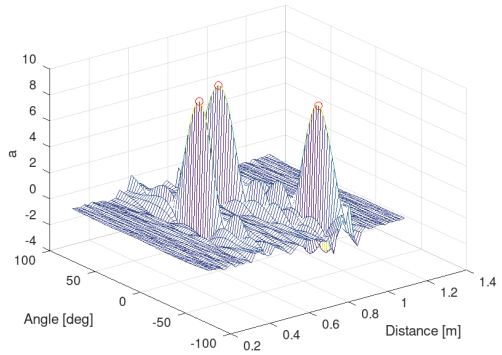

Fig. 3. Matrix of amplitudes obtained based on Eq. 16

In the Fig. 2 and Fig. 3 there are 3, clearly visible peaks. At each of one, with the circle is marked true object location. Comparing both of them, the version based on Eq. 15 has narrower mainlobe. In both cases there is clearly visible that mainlobe in the range dimension is much narrower in that estimate of the amplitude in that domain is changing rapidly. In the azimuth domain – there are much broaded mainlobes which is due to the small number of samples on which the estimate was based.

IV. Tests based on recorded data

A. Device description

To conduct tests using real signals, a device was built with 8 microphones (INMP441) and one speaker (JBL Go). The signal from the microphones was captured in digital form (I2S) using a logic state analyzer and stored in computer memory. A photo of the device is shown in Fig. 4.

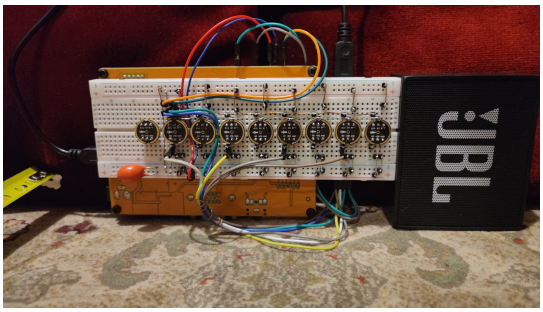

Fig. 4. Photo of the constructed device

B. Testing enviornment

The parameters of the signal as well as the device were the same as in the simulation from Section 3. The only difference was the presence of one object, which initially used a standard 0.5L can. However, it turned out that the echo of the reflected signal was so weak that the microphones were unable to record it. Therefore, two simulations were performed - in the first, there was no object from which the signal could bounce, only the direct sound from the speaker was recorded. In the second simulation, the speaker was moved to a certain distance from the microphones, and the sound coming out of it simulated reflection from an object. Finally, the two signals were summed. This problem could be avoided by using microphones with better sensitivity and by generating sound of higher intensity. For the above reasons, the exact coordinates of the simulated object are not known. It was placed at an angle 45° around $1m$ away - so it should be detected at half of that distance.

C. Results

The effects of both algorithms are shown in the figures below.

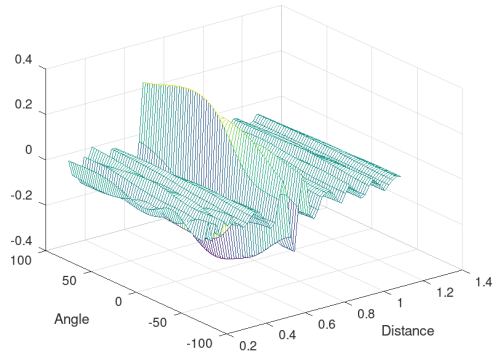

Fig. 5. Matrix of amplitudes obtained based on Eq. 15

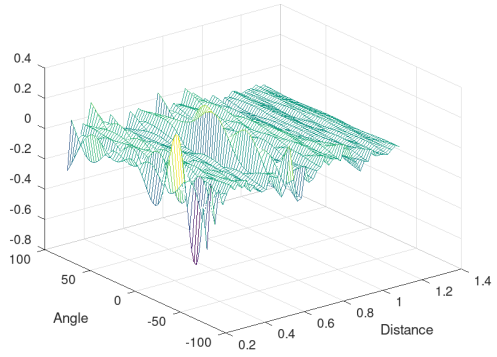

Fig. 6. Matrix of amplitudes obtained based on Eq. 16

In both cases there is a visible peak placed around position $0.5m$ and $45°$. In the full version of the algorithm (Fig. 6), there are also peaks at low distances - these come from the signal directly from the speaker.

V. Computational complexity

Assuming the parameters presented in Section 3, the data is recorded at a rate of $f_s N$ real numbers per second. This is synonymous with the rate at which the data must be processed. Generating synthetic data requires calculating KMN propagation paths. Each requires about 3 multiplications, 1 summation and 1 trigonometric function value calculation per sample, which requires about 400 CPU cycles. In total, the generation of synthetic data alone requires about $400KMNf_s$ clock cycles per second. The data for the coarse estimation (Eq. 16) can be generated before the running stage of the algorithm. This will leave the execution of N multiplications and additions which will require about $20Nf_s$ CPU cycles per path. Assuming the operating frequency of modern processors at 3 GHz, it should not be a challenge to realize the processing of data from several hundred microphones.

VI. Summary

This paper presents and tests two methods of coarse coordinate estimation with different computational complexity and estimation quality. Simulations and tests using real signals were performed.

The presented algorithms enable the device to operate in real time, even with a significant number of recorded signals.

The discussed methods still have potential for improvement. The method presented was based on a single scan of the space. By making a series of them, it is possible to extend the whole by another dimension. This would make it possible to estimate the Doppler shift and even the velocity vector of the observed object.

References

[1] M. A. A. Al-qaness, M. Elsayed Abd Elaziz, S. Kim, A. Ewees, A. Abbasi, Y. Alhaj, and A. Hawbani, "Channel state information from pure communication to sense and track human motion: A survey," *Sensors*, vol. 19, p. 3329, 07 2019. [Online]. Available: https://doi.org/10.3390/s19153329

[2] J. Durán, E. J. Abril, F. J. Rodríguez, J. F. García, V. Matellán, J. J. Villacorta, A. Izquierdo, S. Cerezal, and M. Calvo, "Vigicop: Autonomous surveillance robots with sodar detection and autonomous navigation," in *44th Annual 2010 IEEE International Carnahan Conference on Security Technology*, 2010, pp. 165–169. [Online]. Available: https://doi.org/10.1109/CCST.2010.5678728

[3] E. A. Krajny, L. Ośródka, and M. J. Wojtylak, "Application of doppler sodar in short-term forecasting of pm_{10} concentration in the air in krakow (poland)," *Atmospheric Measurement Techniques*, vol. 17, no. 8, pp. 2451–2464, 2024.

[4] F. Nathanson, J. Reilly, and M. Cohen, *Radar Design Principles: Signal Processing and the Environment*, ser. The SciTech radar and defense series. SciTech Publishing, Incorporated, 1999. [Online]. Available: https://books.google.pl/books?id=9IKI0egWkPgC

[5] A. Kebo, I. Konstantinidis, J. J. Benedetto, M. R. Dellomo, and J. M. Sieracki, "Ambiguity and sidelobe behavior of cazac coded waveforms," in *2007 IEEE Radar Conference*, 2007, pp. 99–103. [Online]. Available: https://doi.org/10.1109/RADAR.2007.374198

[6] A. Hero, "A cramer-rao type lower bound for essentially unbiased parameter estimation," p. 69, 01 1992.

[7] *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. Pearson Education. [Online]. Available: https://books.google.pl/books?id=pDnV5qf1f6IC

[8] R. Penrose, "A generalized inverse for matrices," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, no. 3, p. 406–413, 1955. [Online]. Available: https://doi.org/10.1017/S0305004100030401