DOA Estimation Method in Echo-localization Ability for Human-beings

Yifei Ni

Suzhou Foreign Language School, Suzhou 215000, China

Abstract. The principle of echo-localization processing of human beings is quite similar with animals like bats and dolphins, which has been widely applied in sonar. Direction of Arrival (DOA) methods plays an important role in estimating the radar, sonar and many other echo-localization communication fields. Typical methods in resolving echo-localization problems including classical (Multiple Signal Classification) MUSIC, (Estimation of Signal Parameter via Rotational Invariance Technique), ESPRIT and compressed sensing. The echo-localization methods and the resolving ambiguity methods via uniform array are systematically introduced in this paper. On this basis, the fast and accurate algorithms for human-being echo-localization in parameter (azimuth angle, elevation angle and range) estimation of signal are further introduced from time domain and frequency domain. The accuracy of the angle and distance estimation of the algorithm calculated by computer simulation experiment is over 99.9% and 97% respectively. The effectiveness of DOA method was verified according to MUSIC algorithm, which provides a rapid and accuracy information in echo-localization ability of human being.

Keywords: Echo-localization; Direction of Arrival; Multiple Signal Classification.

1. Introduction

In nature, animals are able to emit ultrasonic waves generated from the larynx through the oral or nasal cavity, applying the returned sound to create a sound "image" of their surrounding environment[1, 2]. So-called ability was found in animals such as bats and dolphins. The process can be expressed as follow: firstly, mouth cavity generates ultrasonic pulses, then ears receive the target echo-signal, and last, brain process the echo-signal and thus obtain target localization and recognition[3]. Inspired by these effects, scholars began to study the echolocation and target recognition ability of human beings.

Compared with normal people, the blind’s auditory system was usually more acute in interpreting echoes, and then form a series of detailed images in their minds, including the distance, even the size and material of objects[4]. The principle of target recognition ability of the blind are quite similar as animals like bats. First, the blind's tongue strikes the upper jaw to produce sound, and when the sound collides with an object in front of it in the process of transmission, an echo signal will be generated. After the ear captures the echo signal, it will be processed by the auditory system and the central meridian system to realize the target location.

To fully understand echolocation ability for human beings, (Tongue Click) TC signal were induced in analyzing the waveforms in time domain and frequency domain of echo-localization [5], which, at the same time, the obtained waveform is compared with the radar waveform and were applied in radar signal processing. Basically, the detection methods of echo-localization signals can be divided into three categories. The first type of methods is algorithms based on frequency domain, which determined by measuring the signal energy of an echo-localization signal that exceeded the frequency threshold in the time period. Such method is available to comparing instantaneous energy with the long-term average value. However, the parameters such as frequency threshold and long-term average value in this method all depend on the observed species and field conditions, and the adaptability of parameter selection is poor[6].The second type is the algorithm based on the time domain, such as Teager-Kaiser energy operator algorithm (referred to as the TK algorithm)[7]. This algorithm detects and obtains the position of the echo-localization signal by calculating the Teager-Kaiser energy peak of time domain signal, which is quite effective for pulse signal type echo-localization signals. But the estimation performance depends on the signal-to-noise ratio (SNR), and it performs poorly at low

SNR condition. The third type of method is based on time-frequency analysis. Applying the fast Fourier transform, the pulse signal higher than the threshold decibel is detected as echo-localization signal[8]. Compared with the frequency domain detection, the information of the time dimension is increased while the overall ability has been increasingly improved.

Direction of arrival (DOA) estimation plays as an important aspect in array signal processing. The main purpose is to identify position of the desired signal in space by processing the sensor array, and the direction angle at which the desired signal arrives array element. In recent years, with the rising development of compressed sensing theory, this technology has also been vigorously developed in array signal processing, mainly using the idea of compressed sensing for DOA estimation. Representative algorithms for DOA estimation methods like L1-SVD, L1-SRACV, JLZA-DOA and Spare-Bayesian-Learning (SBL) etc. The existing low-elevation DOA methods are mainly divided into feature subspace algorithms, maximum likelihood (ML) algorithms and compressed sensing algorithms. The low-elevation feature subspace class algorithm is mainly based on the multiple signal classification (Multiple signal Classification, MUSIC) and the rotation invariant subspace (Estimation of Signal Parameter via Rotational Invariance Technique, ESPRIT) are treated as the framework of solving methods for such problems.

Favored by developers, the MUSIC algorithm has higher stability and angular resolution accuracy than ESPRIT algorithm[9, 10]. Reference [11] adopted Spatial Smoothing (SS) technology to restore the rank of the covariance matrix to achieve coherence, but will lead to the degradation of the algorithm's estimation performance, making it difficult for this type of algorithm to satisfied the needs in practical applications. Reference [12] combined the alternating projection technique with the MUSIC algorithm, and acquired prior information to achieve low elevation angle estimation. Since the hybrid algorithm’s cost function is a non-convex optimization problem, it cannot constantly guarantee that the algorithm converges to the global minimum optimal solution.ML algorithms can directly process coherent signals with perfect estimation performance under the condition of low signal-to-noise ratio(SNR), but the calculation consumption of the algorithm increased exponentially with the rising number of targets, and the consumption of computational calculation is huge, which makes it difficult to meet the real-time requirements[13]. Compressed sensing algorithms are able to directly estimate the DOA of coherent sources by utilizing the sparse characteristics of the target in the spatial domain, and most sparse reconstruction-based DOA estimation methods have better estimation performance under the condition of low signal-to-noise ratio (SNR)[11]. However, the current sparse reconstruction DOA estimation algorithm has a large amount of calculation, thus how to reduce the calculation amount of the algorithm without reducing the accuracy of the algorithm has always been a research hotspot of this type of algorithm[14].

In this letter, a DOA estimation method based on hybrid-MUSIC algorithm was proposed to study echo-localization ability for human beings. A typical click signal was induced to discuss the effects of hybrid MUSIC in narrow incident angle condition for human-beings in reality. In experiment, the distance between the target plane and the receiver were set to be 80cm, and the emission angle of click signal sound source is 30°. The accuracy of the angle and distance estimation of the algorithm is verified by computer simulation experiment, and result angle is about 29.959° and 29.983° respectively with a accuracy over 99.9%. The estimated distance is 77. 642cm and 78.43 cm respectively while accuracy are greater than 97%. The algorithm shows that it is effective to estimate the localization ability of human being by DOA method.

2. Model of Echolocation in Human-beings

The basic process of human echo-localization was described as follow: sound signals were emitted to the target through one’s mouth or other parts of the body, which bounce back after encountering the targets and thus forming echo-signals (with messages). While echo-signals were transmitted and received by human ears (auditory part), and the brain analyzes the echo to obtain the recognition of the target and positioning information such as distance, pitch angles and azimuth.

Therefore, the system model of human echo-localization can be divided into two modules: target model and auditory system model, which is detailed depicted in figure 1. The auditory system of human has the functions of sensing, transmitting, processing and analyzing sound information. It can transmit a large amount of information into neural channels through numbers of auditory receptor cells, while extract useful information from the rapid changes in the time domain or the frequency shifts. The human auditory system is equivalent to an excellent audio signal processor, and the ingenious physical structure provides it with powerful sound signal processing capabilities.

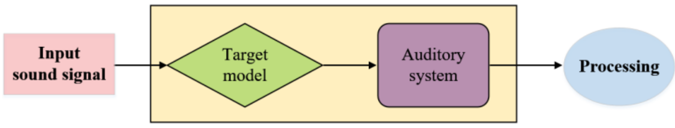

Fig 1. Echolocation system block diagram of human-beings

In the process of echo-localization in human, the human mouth is equivalent to the sound source, human ear is similar to the receiver. The combination of human brain and ear formed into auditory system to process all signals with target recognition. The echolocator continuously emits acoustic signals to the target, and then the properties of the target were judged by analyzing the echoes received by ear. Displayed in figure 2(a), the relationship of echo-localization model was clearly described. In Cartesian coordinate system, assume that the sound source S and the receiver R are located on the same side of the two-dimensional plane, the distance between the left and right receivers and the sound source is equal.

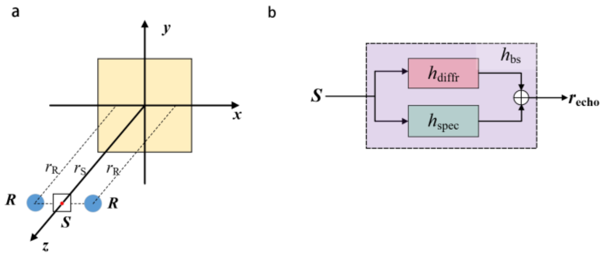

Fig 2. The relationship of echo-localization model(a) v.s. process of echo generation (b)

And the vertical distance from the sound source and the receiver to the two-dimensional plane is equal, were denoted by r_S and r_R respectively. And the position of the sound source and the receiver can move forward towards x axis and y axis, and also lean back and forth on positive side of z axis on the two-dimensional plane.

Sound waves were emitted towards target by the sound source S, and the sound waves will be scattered when meet the target. The backscattered sound waves signals were received by the receiver, referred to the echo signals, which carried attribute information such as the distance, orientation and size of the target. The process of echo generation can be regarded as a linear time-invariant system. Assumed that, for a two-dimensional plane object with smooth surface, the impulse response consists of edge diffraction impulse response h_{diffr} and specular reflection impulse response h_{spec}, h_{bs} represents backscatter of the system, S is a sound generated by a human echolocator, then the echo signals were reflected by the target, which is displayed in figure 2(a). The process could be expressed as equation (1):

$$r_{echo}(t) = s(t) \otimes h_{bs}(t) = s(t) \otimes \left(h_{diffr}(t) + h_{spec}(t)\right) \tag{1}$$

When a signal was emitted by sound source to a two-dimensional plane with smooth surface, specular reflection will occur, as shown in the figure 3. Specular reflection path generated by the sound source and receiver at a specific position, which was separated by yOz plane into two parts.

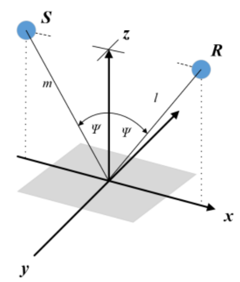

Fig 3. The specular reflection path of echo-signal

The impulse response function of specular reflection can be expressed as equation (2):

$$h_{\text{spec}}(t) = \delta(t - R/c)/R \tag{2}$$

In the upper equation, c denotes the sound transmit velocity in the air. R is the path length of specular reflection of sound waves, where $R = m + l$ (meter). When $t = R/c$, δ (t - R/c) = 1, otherwise δ (t-R/c) = 0, 1/R is the attenuation of the signal. Impulse response function is related to the position of target, sound source and receiver.

In simulation, by changing the position of the target, sound source and receiver, different impulse response functions can be obtained, and then different echo signals can be acquired. Typical click tongue signals were adopted from universal database, two circumstances were set to prove echo-localization model. Vertical distance from the sound source and the receiver to the two-dimensional plane were equal, distance between the sound source S and the left and right receivers R were set to be 10 cm respectively. The vertical distance to the two-dimensional plane (maintain $r_R = r_S$), r_R were set to be 40 cm in condition 1, while parameters r_R were set to be 80 cm in condition 2. Adopted echo-localization model and equation (1) and (2), simulations were proceeded under above conditions, the calculated signal waveforms received by the right receiver were displayed in figure 4 (a) and (b) respectively.

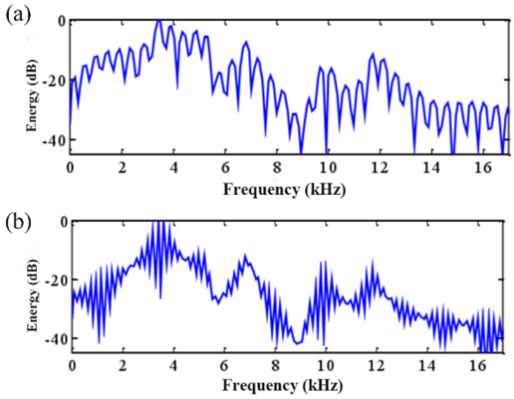

Fig 4. Simulation results of signal waveforms received by right receiver with r_R = 40 cm(a) v.s. r_R = 80 cm(b)

Comparing to the simulation results displayed in figure 4 (a) and (b), it is clearly found that, when the distance between the sound source and receiver to the target were increased slightly (from 40 cm to 80 cm), the weaker the echo signal intensity was, the later the signal received linearly. On the contrary, the receiver receives the signal earlier as the echo signal grow stronger. This result is consistent with the real situation, which proves that the model is effective at these conditions.

2.1 DOA Algorithm

Direction of arrival (DOA) estimation is an important research direction of array signal processing. Generally, the received signals are subjected to spatial Fourier transform, and then the square of modulus is taken to obtain spatial spectrum, finally the arrival direction of the signal is estimated.

Assuming that there is a receiver array in space, which is composed of L array elements and distributed linearly and uniformly. A narrow-band signal source is incident on a linear uniform receiver array at an incident angle θ, s(t) represents the transmitted signal at time T, and the signal received by the 1st array element is

In space, a uniform linear array with only a single signal source, and the received signal can be expressed as equation (3):

$$x_l(t) = \delta_l s(t-\tau_1) + n_l(t) \qquad (l = 1,2,...,L) \tag{3}$$

In equation, where δ_l is the gain of the signal of the first array element, τ_l denotes the time delay when the signal reaches the first array, and $n_l(t)$ represents the noise signal received by the first array at time t. Then the signal at time t was transformed into a signal vector $\boldsymbol{X}_l(t)$with length M, which were displayed as:

$$X_l(t) = \left[x_l(t), x_l(t-1), ..., x_l(t-M+1)\right]^{\mathrm{T}} \tag{4}$$

In equation, $(\bullet)^{\mathrm{T}}$ represents the transpose matrix, from which the signal received by the l-th array can be obtained:

$$X_l(f) = \sum_{d=0}^{M-1} x_l(t-d)\, e^{\frac{-2\pi jfd}{M}} \qquad f = 0,1,....,M-1 \tag{5}$$

Where j is the imaginary unit, the frequency domain broadband signal model can be further obtained as follows:

$$\overline{\boldsymbol{X}}(f) = \left[X_1(f), X_2(f), ..., X_L(f)\right]^{\mathrm{T}} \tag{6}$$

The covariance matrix $\boldsymbol{R}(f)$ of the wideband signal received by the array element can be expressed as follow:

$$\boldsymbol{R}(f) = E\left[\overline{\boldsymbol{X}}(f)\overline{\boldsymbol{X}}(f)^{\mathrm{H}}\right] = AE\left[\boldsymbol{S}(f)\boldsymbol{S}(f)^{\mathrm{H}}\right]A^{\mathrm{H}} + \sigma^2 \boldsymbol{I} \tag{7}$$

$$\boldsymbol{R}(f) = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1L} \\ R_{21} & R_{22} & \cdots & R_{2L} \\ \vdots & \vdots & \cdots & \vdots \\ R_{L1} & R_{L2} & \cdots & R_{LL} \end{pmatrix} \tag{8}$$

In equation (8), the upper triangular matrix of covariance matrix $\boldsymbol{R}(f)$ were considered as the input of neural network, it is necessary to split the real part and imaginary units of the elements in the upper triangular matrix, and then combine the elements into new sequence to obtain the input features. In the equation, *real* (•)indicates the real part of the element, and *imag* (•) denotes the imaginary part of the element.

$$\boldsymbol{y} = \left[real(R_{11}), imag(R_{22}), real(R_{11}), \ldots, imag(R_{LL})\right] \tag{9}$$

2.2 Hybrid-MUSIC Algorithm

The MUSIC algorithm and its improved algorithm have shown strong vitality of the algorithm and its characteristics of high-precision, high resolution and stability in direction measurement. The basic principle was correlated to the array output vector. The eigenvalues and their corresponding eigenvectors are obtained by feature decomposition, and then the beam direction angle is obtained by processing these vectors matrix.

In order to facilitate the derivation of the formula, conditions are assumed as follows:

(1) The antenna arrays are all in the far-field area, which mean the received signals are regarded as plane waves.

(2) Signal sources are all regarded as zero-mean stationary processes related to narrow band and random, and their polarization modes are the same and they are not related to each other.

(3) The received noises of the array all conform to the characteristics of white Gaussian noise, and are irrelevant with other array.

(4) The element parameters of each receiving array are equal.

Supposed that there are d signal sources with incident angles of θ_1, θ_2, …θ_D respectively, the receiving vector of array signal Can be expressed as:

$$X(t)=A(\theta)S(t)+N(t) \tag{10}$$

In equation,

$X(t)=\left[x_1(t),x_2(t),\cdots,x_M(t)\right]^T$ is M×1 order data vector quantity;

$S(t)=\left[s_1(t),s_2(t),\cdots,s_D(t)\right]^T$ is D×1 order signal vector;

$N(t)=\left[n_1(t),n_2(t),\cdots,n_D(t)\right]^T$ is M×1 order additive white Gaussian noise;

$A(\theta)=\left[a(\theta_1),a(\theta_2),\cdots,a(\theta_D)\right]$ is the M×D order steering vector matrix of array pair signals. The noise signal is white Gaussian noise and independent of the incident signal. The covariance matrix of the array output $\boldsymbol{R}_x$ can be obtained by correlation processing:

$$\boldsymbol{R}_x=E\left[XX^H\right] \tag{11}$$

However, in reality, different methods are proposed to solve the output covariance matrix $\boldsymbol{R}_x$ of the array. Therefore, the signal source position can only be estimated approximately. For the covariance matrix $\boldsymbol{R}_x$ of the sample signal, we can obtain:

$$\hat{\boldsymbol{R}}_x=\frac{1}{N}\sum_{i=1}^{N}X(i)X^H(i) \tag{12}$$

R is the maximum likelihood estimate of $\boldsymbol{R}_x$, and if $\boldsymbol{N}\to\infty$, they are equal. But in reality, $\boldsymbol{N}$ can't be infinite, which leads to inevitable errors. Since the noise signal is complex and unpredictable in reality, $a^H(\theta)$ and $\boldsymbol{E}_n$ are not orthogonal, so the corresponding DOA estimation are needed to calculate by correlation optimization algorithm:

$$\theta_{MUSIC}=\arg_o \min a^H(\theta)E_nE_n^Ha(\theta) \tag{13}$$

Therefore, the final signal DOA estimation formula of MUSIC algorithm is obtained:

$$P_{mu}(\theta)=\frac{1}{a^H(\theta)E_nE_n^Ha(\theta)}=\frac{1}{\left\|E_n^Ha(\theta)\right\|^2} \tag{14}$$

The orthogonality between $a(\theta)$ and $\boldsymbol{E}_n$ means that the denominator of this equation is required to take the minimum value and $P_{mu}(\theta)$ contains a peak. Therefore, the DOA can be estimated by the peak value of MUSIC spatial spectrum when incident signals reached.

2.3 Experimental Results

For two incoherent signals with slightly difference of incident angle, it can be distinguished and calculated by the MUSIC algorithm rapidly and accurately, while high estimation accuracy for incoherent signals are obtained.

For the above discussion, computer simulation experiments were adopted for human-echo signal detection. The basic simulation of MUSIC algorithm assumes that the receiving array is a 2-element uniform linear array, narrow-band signals which are irrelevant to each other. The sampling number is 8000, and the signal-to-noise ratio (SNR) is 20 dB. Typical click signals were served as narrow-band signals emitted by sound source, and the two receivers are placed according to Figure 2. Among them, the coplanar distance between the sound source and the receiver is 40cm.

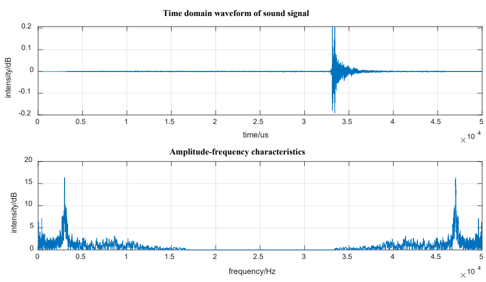

Fig 5. Waveform of sound signal in time domain and frequency domain

The sound source generates sound signals to the plane, which is 80cm away from the two-dimensional reflection plane, and the corresponding reflection angle is about 30°respectively. The time domain waveform and amplitude frequency spectrum diagram of the sound signals generated by the sound source are shown in Figure 5.

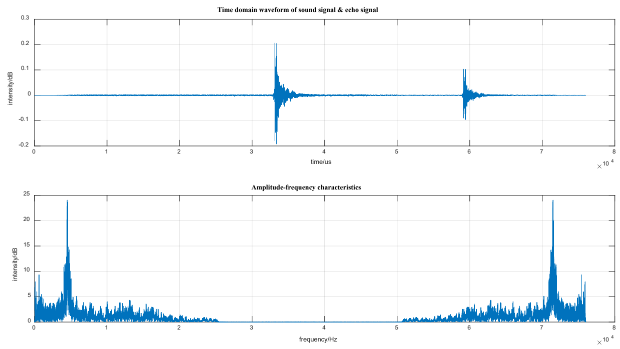

Fig 6. Waveform of received signal in time domain and frequency domain

The signal collected by the receiver is shown in Figure 6, in which the front part of the signal is the directly received signal, and the second part is the echo signal corresponding to this signal, as shown in Figure 6. According to the signal filtering method corresponding to the above algorithm, the received signal was recovered. According to the time domain waveform and spectrum characteristics of the filtered signal, correlation analysis methods were adopted to estimate the distance and reflection angle of the reflector from the sound signal with echo.

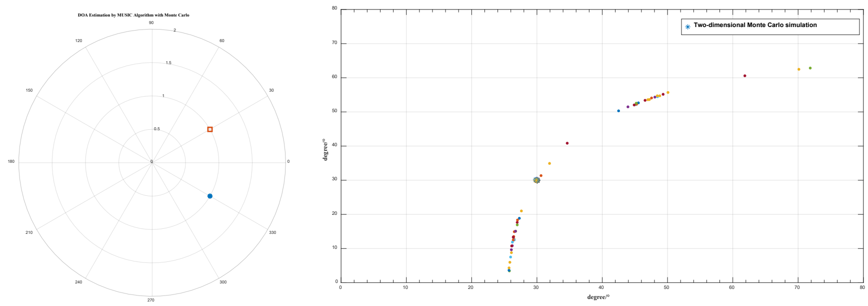

Fig 7. Angle estimation by MUSIC algorithm with Monte Carlo simulation

After 1000 times Monte Carlo simulations, the calculated incident angles were 29.959° and 29.983° respectively, shown in Figure 7, which were calculated to be over 99.9% compared with the initial incident angle set. The estimated distance between the receiver and the reflector is 77. 642cm and 78.43 cm respectively, the accuracy is greater than 97% compared with 80cm set in experiment setup.

3. Conclusion

In this paper, the principle and development status of DOA estimation method were introduced, as well as application and characteristics in echo-localization field. The hybrid-MUSIC algorithm is adopted to simulate the environmental conditions of human ears. The distance between the target plane and the receiver is 80cm, and the emission angle of click signal sound source is 30°. The accuracy of the angle and distance estimation of the algorithm is verified by computer simulation experiment, and result angle is about 29.959° and 29.983°, with a estimation accuracy over 99.9%. For estimated distance between coplanar and receiver, calculation results show that are 77. 642cm and 78.43 cm while the accuracy is greater than 97%. This result shows its effectiveness to estimate the localization ability of human being by DOA method.

References

[1] T.A. Stroffregen, J.B. Pittenger, Human echolocation as a basic form of perception and action, Ecological psychology, 7 (1995) 181-216.

[2] S. Teng, A. Puri, D. Whitney, Ultrafine spatial acuity of blind expert human echolocators, Experimental Brain Research, 216 (2012) 483-488.

[3] F.Q. Niu, Y.M. Yang, Z.M. Zhou, X.Y. Wang, S. Monanunsap, C. Junchompoo, Echolocation clicks of free-ranging Irrawaddy dolphins (Orcaella brevirostris) in Trat Bay, the eastern Gulf of Thailand, The Journal of the Acoustical Society of America, 145 (2019) 3031-3037.

[4] S. Teng, D. Whitney, The acuity of echolocation: Spatial resolution in sighted persons compared to the performance of an expert who is blind, Journal of Visual Impairment & Blindness, 105 (2011) 20-32.

[5] G.E. Smith, C.J. Baker, Human echolocation waveform analysis, IET International Conference on Radar Systems (Radar 2012), IET, 2012, pp. 1-5.

[6] T. Akamatsu, D. Wang, K. Wang, Z. Wei, Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China, The Journal of the Acoustical Society of America, 109 (2001) 1723-1727.

[7] V. Kandia, Y. Stylianou, Detection of sperm whale clicks based on the Teager–Kaiser energy operator, Applied Acoustics, 67 (2006) 1144-1163.

[8] R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, D. Moretti, Passive acoustic detection and localization of sperm whales (Physeter macrocephalus) in the tongue of the ocean, Applied acoustics, 67 (2006) 1091-1105.

[9] T.B. Lavate, V. Kokate, A. Sapkal, Performance analysis of MUSIC and ESPRIT DOA estimation algorithms for adaptive array smart antenna in mobile communication, 2010 Second International Conference on Computer and Network Technology, IEEE, 2010, pp. 308-311.

[10] V. Hryhorenko, D. Klyushin, S. Lyashko, Multiblock ADMM in machine learning, 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), IEEE, 2019, pp. 461-464.

[11] W. Zhu, B.-X. Chen, Novel methods of DOA estimation based on compressed sensing, Multidimensional Systems and Signal Processing, 26 (2015) 113-123.

[12] Y. Liu, H. Liu, X.-G. Xia, L. Zhang, B. Jiu, Projection techniques for altitude estimation over complex multipath condition-based VHF radar, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11 (2018) 2362-2375.

[13] C. Li, B. Chen, Y. Zheng, M. Yang, Altitude measurement of low elevation target in complex terrain based on orthogonal matching pursuit, IET Radar, Sonar & Navigation, 11 (2017) 745-751.

[14] X. Wang, E. Aboutanios, Adaptive Reduced-Dimensional Beamspace Beamformer Design by Analogue Beam Selection, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 4350-4354.