

THESIS ON INFORMATICS AND SYSTEM ENGINEERING C117

Distributed Signal Processing for Situation Assessment in Cyber-Physical Systems

SERGEI ASTAPOV

TUT
PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Information Technology
Department of Computer Control

Dissertation was accepted for the defence of the degree of Doctor of Philosophy in Informatics and System Engineering on July 5, 2016.

Supervisors: Jürjo-Sören Preden, Ph.D., Senior Research Scientist,
Department of Computer Control, Tallinn University of
Technology

Andri Riid, Ph.D., Senior Research Scientist, De-
partment of Computer Control, Tallinn University of
Technology

Opponents: Pasi Eerikki Pertilä, Ph.D., Senior Research Fellow,
Department of Signal Processing, Tampere University of
Technology, Tampere, Finland

Alvo Aabloo, Ph.D., Professor, Faculty of Science
and Technology, University of Tartu, Tartu, Estonia

Defence of the thesis: August 25, 2016

Declaration:

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for any academic degree.

/Sergei Astapov/



European Union
European Social Fund



Investing in your future

Copyright: Sergei Astapov, 2016
ISSN 1406-4731
ISBN 978-9949-83-007-7 (publication)
ISBN 978-9949-83-008-4 (PDF)

INFORMAATIKA JA SÜSTEEMITEHNIKA C117

Hajutatud signaalitöötlus olukorra hindamiseks küberfüüsikalistes süsteemides

SERGEI ASTAPOV

Contents

List of Publications	9
List of Abbreviations	13
1 Introduction	15
1.1 State of the Art	16
1.2 Motivation and Problem Statement	18
1.3 Author's Contributions	20
1.4 Thesis Outline	21
2 Preliminaries	23
2.1 Smart Sensor Component Structure	23
2.2 Time Domain Signal Analysis	25
2.2.1 Signal Envelope	25
2.2.2 Time Domain Features	26
2.3 Frequency Domain Analysis	27
2.3.1 Discrete Fourier Transform	28
2.3.2 Time-Frequency Domain Analysis	29
2.3.3 Frequency Domain Features	30
2.4 Signal Filtering	34
2.4.1 Static Digital Filters	34
2.4.2 Adaptive Filtering	36
2.5 Pattern Recognition	36
2.5.1 Feature Extraction	37
2.5.2 Classification	38
2.5.3 Fuzzy Logic Based Classification	39
2.6 Acoustic Localization	42
2.6.1 Conventional SRP-PHAT	44
2.6.2 SRP-PHAT with Stochastic Region Contraction	45
2.6.3 Multilateration	50
2.7 Trajectory Estimation and Prediction	50
2.7.1 Kalman Filter	51

2.7.2	Rao-Blackwellized Particle Filter	51
2.7.3	Filter Application to the Process Model	52
2.8	Conclusions	53
3	Single-Sensor Solutions	55
3.1	Pattern Recognition for Stationary Processes	55
3.1.1	Example Process Description	56
3.1.2	Signal Analysis	58
3.1.3	Feature Extraction	61
3.1.4	Classification	64
3.1.5	Feature Selection	68
3.1.6	Presented Approach Discussion	69
3.2	Pattern Recognition for Non-stationary Processes	70
3.2.1	Problem Statement and Process Description	70
3.2.2	The Multistage Procedure of Vehicle Identification	73
3.2.3	Multistage Procedure Test Results	80
3.2.4	Real-Time Operation on Embedded Device	84
3.2.5	Multistage Procedure Discussion	88
3.3	Conclusions	89
4	Multi-Sensor Solutions	91
4.1	Direction of Arrival Estimation with Linear Arrays	91
4.1.1	Reduced Functional for SRP-PHAT	92
4.1.2	Proposed Approach to DOA Estimation	92
4.1.3	Performance at Low Sampling Rates	96
4.1.4	Discussion	98
4.2	Volumetric Direction of Arrival Estimation	99
4.2.1	Conical Array Structure	99
4.2.2	Volumetric Functional for SRP-PHAT	100
4.2.3	Proposed Volumetric DOA Estimation Approach	102
4.2.4	Signal Shifting and Influence of Near Field Error	104
4.2.5	Experimental Evaluation	105
4.3	Acoustic Localization	110
4.3.1	Distributed Localization in WSN	110
4.3.2	Initial Search Region Reduction	112
4.3.3	Application to SRP-PHAT with SRC	116
4.4	Experimental Evaluation	116
4.4.1	Array Regular Configuration	117
4.4.2	Array Irregular Configuration	118
4.4.3	Computational Efficiency Assessment	124
4.5	Implementation on Embedded Hardware	127
4.5.1	Mote Implementation of the Linear Array	127
4.5.2	Implementation on the BeagleBone Platform	132

4.6	Conclusions	134
5	Data Fusion and WSN Communication	137
5.1	Data Fusion Aspects	137
5.1.1	Considered Tasks of Data Fusion	138
5.1.2	Discussed Data Fusion Methods	140
5.2	Trajectory Estimation and Prediction	140
5.2.1	Proposed Trajectory Estimation Approach	140
5.2.2	Experimental Evaluation	142
5.3	WSN Communication Issues	145
5.3.1	Local Time and Communication Delays	145
5.3.2	Data Validation	146
5.3.3	Communication Solution	148
5.4	Conclusions	150
6	Applications	151
6.1	Industrial Machinery Monitoring	151
6.1.1	Problem statement	152
6.1.2	State Identification using Multimodal Information	152
6.1.3	Localization of Machinery Operating Region	154
6.2	Identification and Tracking of Military Vehicles	160
6.2.1	Intelligence Surveillance and Reconnaissance	161
6.2.2	Military Vehicle Experiment Setup	162
6.2.3	Vehicle Localization and Trajectory Estimation	164
6.2.4	Vehicle Identification	170
6.3	Detection and Speed Estimation of Ground Troops	178
6.3.1	PIR Signal Analysis for Movement Speed Estimation	178
6.3.2	Experimental Results	182
6.4	Gunshot Detection and Shooter Localization	184
6.4.1	Shooter Localization Preliminaries	185
6.4.2	Proposed Approach to Shooter Localization	189
6.4.3	Gunshot Acoustic Event Detection and Separation	190
6.4.4	Information Fusion and Shooter Localization	194
6.4.5	Experimental Results	198
6.4.6	Discussion	215
6.5	Conclusions	217
	Conclusions	219
	Bibliography	223
	Abstract	239
	Kokkuvõte	241

Acknowledgments	243
Elulookirjeldus	245
Curriculum Vitae	247
Publications	253

List of Publications

- P1. S. Astapov and A. Riid, “A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 2, pp. 151–160, 2013.
- P2. S. Astapov, J. Berdnikova, and J.-S. Preden, “Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- P3. S. Astapov, J. Ehala, and J.-S. Preden, ”Performing acoustic localization in a network of embedded smart sensors,” *International Journal of Microelectronics and Computer Science (IJMCS)*, vol. 6, no. 3, pp. 86–95, 2015.
- P4. S. Astapov, A. Riid, J.-S. Preden, and T. Aruvali, “Industrial process monitoring by multi-channel acoustic signal analysis,” in *Proc. 14th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2014, pp. 209–212.
- P5. S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, “Object detection for military surveillance using distributed multimodal smart sensors,” in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP)*, Hong Kong, China, Aug 2014, pp. 366–371.
- P6. S. Astapov, J. Berdnikova, and J.-S. Preden, “A two-stage approach to 2D DOA estimation for a compact circular microphone array,” in *Proc. Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, Kitakyushu, Japan, Jun 2015, pp. 1–6.
- P7. S. Astapov, J. Berdnikova, J. Ehala, and J. S. Preden, “Shooter localization by networked multichannel acoustic ground sensors,” in *Proc. Int. Conf. on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, Oct 2015, pp. 332–337.

Author's Contribution to the Publications

All results in [P1]–[P7] were obtained by the author of the thesis under the supervision of Dr. Jürjo-Sören Preden and Dr. Andri Riid. All experimental data was acquired by or with participation of the author of the thesis. The fuzzy classification algorithm used in [P1, P4, P5] was provided by and applied with participation of Dr. Andri Riid.

In [P1] a multistage procedure of vehicle acoustic identification developed by the author of the thesis is presented. The paper discusses the problems of single-sensor acoustic signal processing and pattern recognition for the application of detecting passing vehicles and classifying their types. The experimental data was analyzed and the proposed approach was evaluated in the MATLAB environment. The real-time operation capabilities of the procedure were tested on embedded hardware.

The contribution of the author of the thesis in [P2] lies in the development of an acoustic localization procedure for distributed multichannel acoustic sensors, equipped with linear microphone arrays. The procedure is based on DOA estimation and aims at minimizing the region of search for the acoustic source in order to reduce the number of computations of such localization algorithms, like SRP-PHAT. The impact of reducing the search region on the resulting computational efficiency and localization accuracy was evaluated for the applications of speaker localization and identification of noise-emitting components of industrial machinery. In [P3] the method of search region reduction was considered as an approximate localization procedure for a WSN of low power embedded devices (smart dust motes). The complications of performing acoustic localization at extremely low sampling rates and under strict limitations of processing resources and available memory were addressed. The major part of the engineering work on the motes was performed by Mr. Johannes Ehala.

In [P4] the localization method discussed in [P2, P3] is applied to industrial machinery monitoring. The localization of noise-emitting components of industrial machinery is studied in greater detail along with the possibilities of identifying the operation states of specific machines using acoustic information. Military applications of the distributed acoustic WSN system are discussed in [P5] along with a proposal of using passive infrared sensors to estimate the movement speed of passing objects, made by the author of the thesis. Trajectory estimation and global situation assessment through data fusion are also touched upon.

A computationally effective method of azimuth and elevation angular DOA component estimation for conical arrays is presented in [P6], as apposed to localization in the horizontal plane for linear arrays in [P2]. The proposed method is evaluated in the MATLAB environment and compared with SRP-PHAT in terms of small sound-emitting object DOA estimation.

The microphone arrays and the DOA estimation method discussed in [P6] are applied to the task of shooter position estimation in [P7]. A procedure of acoustic data fusion is proposed, which estimates the position of the shooter, provided the DOA and other parameters of the acoustic events produced by a supersonic projectile, obtained by several distributed smart sensors. The sensor prototypes were designed and implemented on embedded hardware, and the live shooting experiments were performed by the author of the thesis and Mr. Johannes Ehala and Mr. Jaanus Kaugerand. The experimental results were evaluated in the MATLAB environment by the author of the thesis.

List of Abbreviations

ADC	Analog to Digital Converter
ADSR	Attack Decay Sustain Release
AOA	Angle of Arrival
BBB	BeagleBone Black
CNC	Computer Numerical Control
CPS	Cyber-Physical System
DAQ	Data Acquisition Device
DEFS	Differential Evolution Feature Selection
DFT	Discrete Fourier Transform
DOA	Direction of Arrival
FFT	Fast Fourier Transform
FOV	Field of View
GPS	Global Positioning System
ICA	Independent Component Analysis
ISR	Intelligence Surveillance and Reconnaissance
ISRR	Initial Search Region Reduction
KF	Kalman Filter
LSFS	Least-Squares Feature Selection
MB	Muzzle Blast
MCFS	Multi-Cluster Feature Selection

MEMS	MicroElectroMechanical System
MF	Membership Function
MFCC	Mel-Frequency Cepstral Coefficients
MRC	Minimal Rule Classifier
NOI	(Event) Not-of-Interest
PCA	Principal Component Analysis
PF	Particle Filter
PIR	Passive Infrared
RBPF	Rao-Blackwellized Particle Filter
RMS	Root Mean Square
RMSE	Root Mean Square Error
SNR	Signal to Noise Ratio
SRC	Stochastic Region Contraction
SRP-PHAT	Steered Response Power with Phase Transform
SW	Shockwave
TDOA	Time Difference of Arrival
TH	Target Hit Noise
UCA	Uniform Circular Array
UGS	Unmanned Ground Sensor
WSN	Wireless Sensor Network
ZCR	Zero Crossing Rate

Chapter 1

Introduction

Information is the resolution
of uncertainty.

Claude Shannon

A cyber-physical system is a combination of computational, communication and physical processes, integrated into the physical environment of the system's operation domain [6, 137]. The defining characteristic of cyber-physical systems (CPS) lies in their interaction with their environment and objects (and/or subjects) in that environment, as opposed to computer systems, which only provide a service and can only be acted upon [123]. CPS are primarily feedback systems, consisting of many distributed interconnected components, which gather information about the state of the environment and the objects of interaction in order to determine the system's response [64, 88]. The situation assessment is performed based on either the information acquired by the sensor components, or explicit information provided by an operator (if one exists) [88].

Accurate situation assessment is crucial for system state and course of action establishment, and also for satisfying the adaptive and predictive properties of CPS [63]. Real-time operation is another important requirement of CPS [83, 134]. Therefore, situation assessment must be performed in a timely manner in order to ensure high responsiveness of the system [139]. Taking into consideration that the components of modern CPS are typically implemented on embedded hardware with limited computational resources, the methods of information and signal processing must be designed specifically to meet both computation time and complexity restrictions [80].

This thesis summarizes the research experience and the principal results achieved by the author in the field of signal processing and information interpretation methods for situation assessment in CPS. The signal processing methods presented in the thesis are applicable to band-limited signals, such

as acoustic or vibration (acceleration) signals. The majority of the considered methods are applied to acoustic signal analysis. The main considered tasks for situation assessment consist of detection, property identification and trajectory estimation of objects residing in the environment governed by a CPS. In regard to detection and identification, the thesis presents several single-sensor solutions of signal time and frequency domain analysis, feature extraction and pattern recognition [14, 15, 18, 19]. Multi-sensor solutions consist of acoustic localization [10, 12, 21] and object trajectory estimation [11]. Several application specific data fusion methods are also presented [9, 18]. System state and action determination in response to the assessed situation are not discussed in the thesis.

The operation domains of the considered CPS consist of both indoor and open environments, where the independent sensor components are distributed to be able to observe an area, where some process takes place. The sensor components then perform situation assessment of the observed area, produce local assessments independently and transmit their gathered information through a Wireless Sensor Network (WSN) to information aggregation components, where a global situation assessment is made through data fusion [64, 77]. The thesis handles this system structure in a more abstract manner during the examination of the considered signal processing and information analysis methods, and presents specific system architecture and sensor implementation examples for several applications, such as vehicle identification and tracking [18], industrial machinery operation state monitoring [14, 19] and shooter acoustic localization [9, 22].

1.1 State of the Art

The term “cyber-physical system” emerged and became used in the past decade, describing systems that integrate computer systems with the physical processes of the real world [6, 137]. The term is closely situated with the Internet of Things [151], Industry 4.0 [73], Machine-to-Machine (M2M) [162], and several others [63]. All these concepts reflect on a vision of deep connection between our physical world and the information world. However, while the Internet of Things is concerned with interconnection and integration between computer systems and smart components (e.g., appliances with computational and network capabilities), and Industry 4.0 is concerned with industrial applications, CPS describe the foundation of physical-computer world interaction. The concept of CPS utilizes such paradigms as ubiquitous computing and non-deterministic modeling to account for high levels of complexity and uncertainty of the physical world [64, 134, 140].

The history of computer systems integrated with physical processes spans much further back than the inception of the modern vision of CPS. Au-

tomation and process control has long become an integral part of industry and everyday life. Advances made in manufacturing and robotics, energy production and logistics help to increase productivity while reducing process latency and cost [110]. Modern health care instruments with integrated computer components increase the efficiency of patient diagnosis, health state monitoring and surgery [96]. Advanced munitions, like guided missiles and autonomous drones rapidly change modern warfare [131]. Although impressive, computer automation of the past years relies on feedback control loops, where specific information about the controlled process is gathered and the control influence is calculated in a deterministic manner in order to achieve the desired dynamic state of the process. These system usually have a narrow field of operation, greatly depend on the preset bounds and usually require supervision of a human operator. Modern CPS, on the other hand, tend to a higher level of autonomy, where the computer components possess greater intelligence, adaptive and predictive properties, and are able to perform more complex decision making [78, 106, 139]. As a result, the scope of operation of any given modern CPS is far wider, and the intervention of the human operator is less critical, if at all required [88, 140].

Research and development of CPS for various applications has been gaining pace in the past decade [63]. The application range of CPS covers almost all aspects of modern life. CPS for smart industry organize the production process and integrate it with enterprise to lower costs of production, logistics and management [42, 135]. The above mentioned Industry 4.0 concept [73] serves a good example of CPS applications in industry. Emergency response and disaster management systems built on CPS are intended to reduce the response time and optimize resource management for disaster situation handling [54, 139]. Critical infrastructure CPS include Smart Grids [104], which are developed for the Smart City architecture [99], are intended to optimize the urban infrastructures by reducing energy consumption of buildings and in the city streets (e.g., by applying the smart lighting technology) [53], optimizing city traffic, etc. Building energy consumption is reduced by applying building automation CPS with smart utility networks [38, 60], which are integrated with the Smart Grids [53]. Intelligent traffic management comprises adaptive traffic management by traffic light and driveway load control, smart vehicles and the vehicle-to-vehicle, vehicle-to-pedestrian and vehicle-to-infrastructure interaction, built on wireless, cellular and satellite components [63, 122]. CPS incorporated in air transportation systems aid with airspace control and regulation, and air traffic control at airports [31]. The applications of CPS in health and medicine include advanced home care and assisted living possibilities, autonomous health monitoring, smart prescription systems, and smart operating rooms [7, 65, 85]. The military applications of CPS include autonomous Intelligence Surveillance and Reconnaissance (ISR) systems for threat detection, identification of friend or foe and other tasks of

battlefield situation reporting and agile resource management [66, 88, 111].

To support higher levels of intelligence of CPS, they must possess high levels of situation awareness [137]. Situation awareness is provided through advanced sensing, information aggregation and fusion [64]. In contrast to classical computer automation systems, which acquire information of strictly specified modalities and only from certain measurement points, CPS employ multi-modal data acquisition from distributed sensor components in order to assess the state of the observed object, process or environment as fully as possible [88, 94]. Increased situation awareness also implies knowledge accumulation and knowledge base creation to enable predictive and proactive behavior [64, 88]. The rapid development of embedded computer systems has paved the way for ever more compact and computationally powerful smart sensors, which constitute the information acquisition and situation assessment components of modern CPS [80, 84]. The state of the art system architectures are composed of decentralized standalone smart sensors, interconnected through ad-hoc WSN, performing cooperative situation assessment [64, 152]. Multi-modal signals are acquired from many measurement points by sensors with separate or overlapping Fields of View (FOV), signals are aggregated through signal fusion or processed separately to extract information and perform situation assessment through information fusion [88]. Sensor types vary from classical binary sensors (i.e., sensors with two discrete states), acceleration, pressure, magnetic, infrared, chemical, etc., to more sophisticated, e.g., video, acoustic, radio frequency, hyperspectral, etc. Sensors with narrow FOV, e.g., acceleration and magnetic, provide highly localized information, whereas sensors with wide FOV, e.g., video and acoustic, allow for large area coverage [94]. Advanced methods of signal processing and pattern recognition methods allow to expand beyond the typical applications of these kinds of sensors [78, 106]. For example, video and hyperspectral sensors are used in a wide range of applications from face recognition [86], living tissue analysis [43], ground [70] and aerial [109] traffic monitoring, to multi-layer ground mapping [143] and object identification [40, 93] from aerial and satellite images. Applications of acoustic sensors span from speech recognition [154] to engine diagnostics [67], underwater craft detection [141] and sniper positioning [4].

1.2 Motivation and Problem Statement

The main problems situated with situation assessment in CPS arise from their distributed structure [156]. If a large number of components is interconnected through a common fabric of WSN, large data quantities cannot be transmitted between components, as this can overload the network and disrupt real-time operation [107, 152]. Therefore, raw data transmission is sel-

dom possible, and signal processing and initial information extraction must be performed locally. Inter-node synchronization also poses a problem in WSN and is unreliable in a dynamic ad-hoc topology [107]. Furthermore, synchronous transmission creates a bottleneck on the information aggregation nodes, which can also disrupt real-time operation and cause information loss. Thus signal processing and data fusion techniques for distributed CPS should not require precise synchronization in order to increase system reliability. Soft synchronization or node clock drift estimation is still needed for the system to be able to distinguish between observed events and changes in the observed process dynamics, however, strict and continuous synchronicity should not be implied.

Signal processing and data fusion methods also need to be inexpensive in terms of computational resources and power consumption [80]. As several CPS components may be implemented on a single hardware platform, the signal processing algorithms need to be computationally lightweight to allow for real-time processing, while allocating enough resources for other functions, like communication management. In terms of data fusion method implementation, the solutions should not imply fixed size structures, where the number of smart sensors (WSN nodes) is constant [107]. For many applications the standalone components are deemed disposable, as the strength of distributed CPS comes from the number and variety of smart sensors and not from any sensor in particular. Therefore, it can be implied that a sensor will cease to exist or come out of operation, e.g., due to loss of communication, and other sensors may be added to the WSN at any given time.

In state of the art distributed CPS video and image analysis is the most popular choice for wide area coverage [94, 140]. Numerous algorithms for the applications of computer vision and robotics are being adopted for automated monitoring and object recognition [64, 106]. The main topic of this thesis is situated with the development of signal processing methods for acoustic signal analysis, which can increase the robustness of video and other information based situation assessment. Acoustic signal acquisition does not require sufficient luminosity or direct line-of-sight (due to acoustic diffraction) of the observed objects. Furthermore, high definition image processing requires significantly more resources than single dimension signals. Although acoustic signals are more prone to noise pollution, especially in open environments, false detection rate can be reduced if information of other modalities is used in conjunction with acoustic information. In this regard acoustic sensors complement video sensors well in situations, when objects are indistinguishable visually, e.g., due to camouflage, but emit sound. Camouflaged military vehicles serve a good example of this situation [150]. On the other hand, if the FOV of a video sensor is blocked by the observed object, acoustic signals may provide information on the situation behind this object. Several signal analysis methods discussed in this thesis may also be applied to other types of

band-limited signals, e.g., vibration. The thesis handles several approaches of band-limited signal analysis, feature extraction and pattern recognition implementable on embedded hardware of smart sensors.

Multi-sensor acoustic systems provide additional possibilities to detection and identification of acoustic processes performed by single-sensor solutions [28]. Microphone arrays allow for acoustic localization and noise attenuation via beamforming [28, 33]. Both procedures are based on the analysis of phase delays between signal channels to estimate the Time Difference of Arrival (TDOA) and Direction of Arrival (DOA) of the acoustic wavefronts [28]. If a microphone array is implemented on a single physical device, beamforming is performed on the same device and no raw data transmission is required. Acoustic localization, on the other hand, operates better if the microphones are well separated spatially. Thus they cannot be implemented on a single compact device, and acoustic localization thus poses a problem for distributed CPS. No well established methods for distributed acoustic localization currently exist. The thesis proposes an asynchronous method of approximate localization and trajectory estimation in a WSN of multi-channel smart sensors.

1.3 Author's Contributions

The main contribution of the author of the thesis lies in the development of single and multi-channel signal processing procedures of acoustic and other band-limited signals for implementation in distributed systems. The contribution comprises three consecutive parts:

- *Single channel*: Study of band-limited signal processing and feature extraction techniques in the time and frequency domain. Feature set specification for non-harmonic signals with highly spread spectral densities. Application specific pattern recognition techniques for stationary and non-stationary process state estimation and their implementation.
- *Multi-channel*: Development and implementation of microphone array smart sensors, one- and two-dimensional DOA estimation techniques. Proposal of a computationally lightweight distributed localization procedure, which can serve as an initial search region bound for precise synchronous localization or an approximate localization approach for asynchronous distributed sensors. Development of a data fusion technique for localized object trajectory estimation and prediction based on localization data and Kalman filter estimates.
- *Applications*: Industrial machinery operation state monitoring and abnormal state detection by means of acoustic noise emitting region estimation and acoustic pattern analysis. Development and implementa-

tion of a multistage procedure of acoustic signal analysis and pattern recognition for vehicle detection and identification. Object detection and movement speed estimation by means of passive infrared sensor signal analysis. Shooter position estimation using shot detection and shot acoustic component identification techniques combined with a novel data fusion approach of shot trajectory estimation and shooter localization.

All studied and proposed methods are verified on real-life signals acquired during *in situ* experiments by or with participation of the author of the thesis. Smart sensor embedded hardware implementations are proven to be able to operate in real-time.

1.4 Thesis Outline

Each chapter begins with a summary of the research problems discussed and ends with concluding remarks on the theoretical and practical results obtained in the corresponding chapter. The last chapter of the thesis comprises general concluding comments and discusses prospects of future research. The summary of each chapter is provided below.

Chapter 2

This chapter considers the preliminaries of time and frequency domain signal analysis, signal filtering, feature extraction techniques and classification basics for pattern recognition, and discusses the problems of acoustic localization and trajectory estimation of localized acoustic sources. The chapter forms the theoretical basis for the signal processing methods discussed and proposed in the latter chapters of the thesis.

Chapter 3

The chapter is devoted to single-sensor solutions of pattern recognition for stationary and non-stationary process state estimation and object identification. The first part of the chapter introduces the common approach to pattern recognition, which includes the extraction of representative features from a single-sensor signal, feature selection from a set of time and frequency domain features, and classification of processes or objects based on the extracted features. This common approach is presented on an example of stationary process state estimation for the application of industrial machinery operation state identification. The second part of the chapter presents a proposed multistage identification procedure for the non-stationary process of moving vehicle pattern recognition. Experimental evaluation of the considered procedures is provided along the way of discussion.

Chapter 4

The chapter is dedicated to multi-channel signal analysis, and the methods of acoustic sensor array signal processing in particular. The chapter begins with the discussion of DOA estimation methods for the linear and conical geometries of the acoustic arrays, presents the proposed approaches to DOA estimation, and evaluates their accuracy and computational efficiency. Later the proposed approach to distributed acoustic localization is presented and evaluated on experimental data. The last part of the chapter presents the microphone array implementations on embedded hardware and evaluates their ability to operate in real-time based on the implementation test results. The DOA estimation and distributed acoustic localization procedures are tested for the application of indoor speaker localization.

Chapter 5

The topics of this chapter comprise the discussion of problems situated with data fusion and communication issues in the distributed WSN of system components. The considered methods of signal processing are reviewed from the data fusion perspective, and a fusion procedure for trajectory estimation of localized objects is presented. The aspect of data concurrency and the challenges of data temporal and spatial evaluation for data alignment are discussed along with the problems of message communication in a WSN with an ad-hoc topology. An example of a network management and data temporal alignment approach is reviewed.

Chapter 6

The chapter discusses the considered applications of the proposed signal processing and data fusion approaches. The methods of stationary process state estimation along with the method of distributed acoustic localization are applied to the civil application of industrial machinery operation state identification and malfunction detection. The methods of non-stationary process analysis and acoustic localization are applied to the military application of vehicle identification and trajectory estimation for military reconnaissance purposes. A proposed method of object detection and speed estimation based on passive infrared sensor signal analysis is presented later in the chapter. The final application considered in the chapter is situated with gunshot detection and shooter acoustic localization by a system of distributed multi-channel acoustic sensors. Experimental evaluation of the employed procedures is provided for every considered application.

Chapter 2

Preliminaries

This chapter introduces the core concepts and algorithms used in the thesis. The chapter begins with an overview of a generic smart sensor structure. Next, the concepts of time and frequency domain signal analysis are reviewed, and a number of relevant signal features is presented. A short introduction to signal filtering is made. Then, pattern recognition is discussed along with one particular fuzzy logic based classifier mainly applied in this work for classification purposes. Finally, the concepts and tools of acoustic localization and object trajectory estimation are presented.

2.1 Smart Sensor Component Structure

Smart sensors constitute the perceptual part of the CPS [107]. The basic tasks of every smart sensor lie in data acquisition, processing and information communication. Depending on the hardware limitations, a sensor may perform only these basic tasks and transmit certain information, extracted from a signal, or perform local situation assessment and send assessment results for further information fusion. Usually it is a good practice to distribute a large number of low computational power (and thus cheap) sensors with narrow FOV, e.g., proximity sensors, to gather highly localized information and forward low level information (binary state or simple signal parameters) for further fusion [80]. The Smart Dust system [69], for example, takes this notion to the extreme by applying very small (down to a few millimeters) MicroElectroMechanical systems (MEMS), called motes, for extremely localized measurements. More powerful nodes with more complex sensors or multi-sensor solutions with wide FOV are usually distributed more sparingly. These nodes may also perform the tasks of intermediate data aggregation and fusion if sufficient resources remain. Of course, every node in a multi-hop WSN may perform the tasks of data aggregation and forwarding, as discussed in Section 5.3.

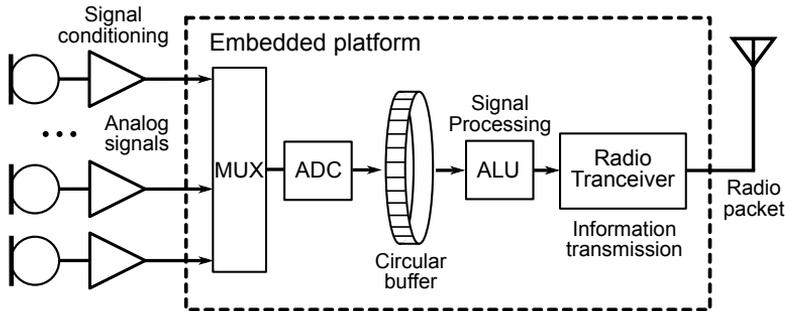


Figure 2.1: Smart sensor data processing principle diagram.

Wireless smart sensors all have a similar principle structure, presented in Figure 2.1. The process starts with signal acquisition, where if analog sensors are concerned, the continuous analog signal is sampled and quantized by an Analog to Digital Converter (ADC) and the discrete time series is then used during processing, or if the sensor produces digital outputs, they are read directly through a certain interface. Analog signal conditioning (i.e., amplification, normalization, bias removal, etc.) is performed by analog circuitry [142] prior to conversion in order not to impose quantization errors onto the signals (e.g., if the signal is amplified in digital form, the error of quantization will be magnified accordingly). If the ADC does not support parallel conversion, consecutive channel selection is performed by a multiplexer, the switching delay of which has to be accounted for if the applied signal processing methods are phase sensitive. After conversion the data is buffered for frame-by-frame processing. For signal buffering a circular buffer is more preferable for real-time implementation to, e.g., classical queue or stack, because it allows to read the latest sample values more efficiently. Specifically, if some delay occurs and the samples are not read in time, the new incoming samples will simply overwrite the older ones without any additional buffer clearing manipulations.

The block denoted as the Arithmetic Logic Unit (ALU) performs all the tasks situated with signal processing, information extraction, etc. The processed data intended for transmission is then handled by the communication module and communicated. Whether the ADC, ALU and the communication module are located on a single processor or microcontroller, or separate entities exist for each task, depends entirely on the implementation of the embedded platform [83, 156]. For example, the signal processing tasks may be assigned to a Digital Signal Processor (DSP) with a Reduced Instruction Set Computing (RISC) architecture optimized for signal processing, or a Field-Programmable Gate Array (FPGA), while process management tasks are run on the Central Processing Unit (CPU).

2.2 Time Domain Signal Analysis

A one-dimensional digital signal is described by a frame of samples $x[k]$, $k = 0, \dots, N-1$, where N is the finite frame length: $[x[0], x[1], \dots, x[N-1]]$. In a multi-channel system consisting of M channels sampled at the same sampling frequency f_s , e.g., a microphone array, the two-dimensional frame will consist of $N \times M$ samples:

$$\mathbf{x} = \begin{bmatrix} x_1[0] & x_2[0] & \cdots & x_M[0] \\ x_1[1] & x_2[1] & \cdots & x_M[1] \\ \vdots & \vdots & \ddots & \vdots \\ x_1[N-1] & x_2[N-1] & \cdots & x_M[N-1] \end{bmatrix}. \quad (2.1)$$

If a multi-channel system acquires multi-modal signals at different sampling rates, the total signal will combine the separate signal frames, where the frame lengths will usually not be equal, $N_1 \neq N_2 \neq \dots \neq N_M$. The signal frame (2.1) is combined of separate measurements, and thus any sample is independent of the neighboring samples, as opposed to, e.g., a two-dimensional video frame [147]. However, in a general case it cannot be guaranteed that the vectors of (2.1) will be linearly independent.

2.2.1 Signal Envelope

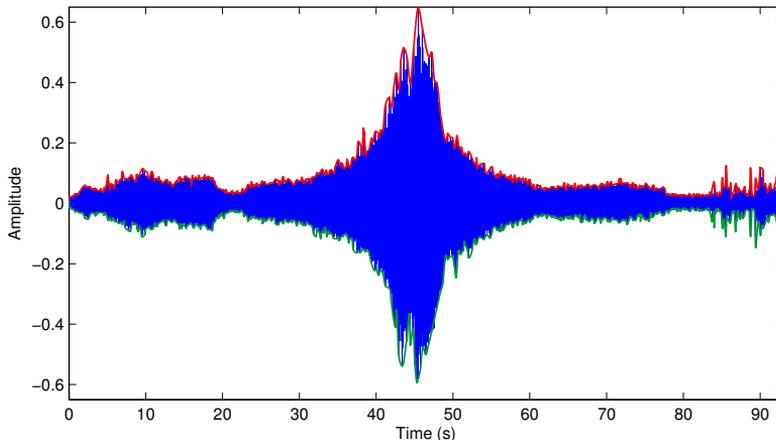


Figure 2.2: Upper and lower envelopes of an acoustic signal of a passing vehicle sampled at $f_s = 20$ kS/s.

Time domain signal analysis focuses on the shape and amplitude of a signal. Signal enveloping is the most popular procedure to estimate the signal shape [133]. Figure 2.2 portrays the upper and lower envelopes of an acoustic signal, calculated by performing positive and negative peak detection

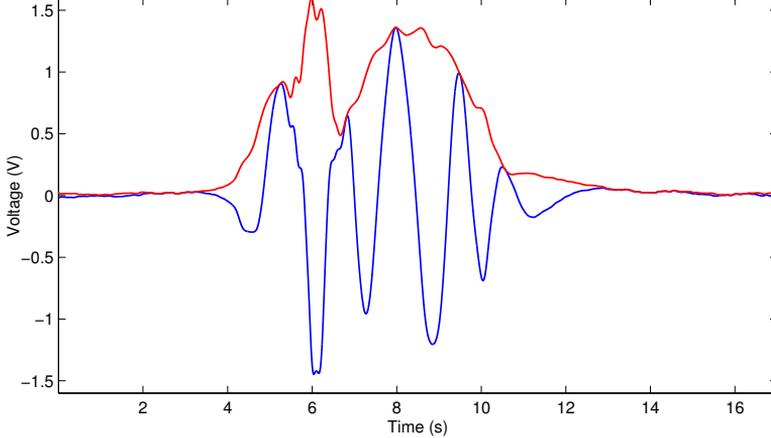


Figure 2.3: Envelope of a passive infrared sensor signal calculated by applying the Hilbert transform.

and applying interpolation. Another approach of calculating the absolute envelope lies in applying the Hilbert transform

$$\mathcal{H}_i\{x\} = \frac{1}{N} \sum_{k=0}^{N-1} x[k] \left[1 - (-1)^{k-i} \right] \coth [(k-i)\pi/N], \quad (2.2)$$

for $i = 0, \dots, N-1$, and taking its absolute value $|\mathcal{H}\{x\}|$. An example of the absolute envelope of a passive infrared (PIR) sensor signal is presented in Figure 2.3. The rise in amplitude then signifies the detection of some event being observed, and the dynamics of amplitude change can be estimated by the Attack Decay Sustain Release (ADSR) portions of the envelope [147].

Implementation-wise the Hilbert transform approach to calculating the signal envelope is computationally expensive and requires longer signal frames to be effective. The peak detection approach is more applicable for frame-by-frame signal analysis, usually used on embedded hardware. In general an envelope gives an adequate estimate of signal shape for highly oscillating and thus more symmetrical signals. For signals with lower dynamics the envelope may misrepresent the signal shape, as shown in Figure 2.3.

2.2.2 Time Domain Features

Besides the signal envelope, other widely used features of time domain signal analysis include instantaneous features [115], calculated for each short signal frame x_t at acquisition time t . These features include, but are not limited to:

- Zero Crossing Rate (ZCR), which is defined as the ratio of zero cross-

ings, i.e., sign changes in a signal frame, and is expressed by

$$\text{ZCR}_t(x_t) = \frac{1}{N-1} \sum_{k=1}^{N-1} [x_t[k] \cdot x_t[k-1] < 0], \quad (2.3)$$

where $[\cdot]$ denotes the Iverson bracket, i.e.,

$$[P] = \begin{cases} 1, & \text{if } P \text{ is true;} \\ 0, & \text{otherwise,} \end{cases} \quad (2.4)$$

where P is a statement, which can be true or false. The interval between successive zero crossings can also be considered a rough representation of the half-period of oscillation for weakly oscillating signals.

- Autocorrelation, which is the cross-correlation of a signal frame with itself. Autocorrelation represents the signal spectral distribution in the time domain and is computed as

$$\text{ACF}_t(x_t, i) = \frac{1}{x_t[0]} \sum_{k=0}^{N-i-1} x_t[k] \cdot x_t[k+i], \quad (2.5)$$

for $i = 0, \dots, N-1$, where $1/x_t[0]$ normalizes the autocorrelation at zero lag to exactly 1.

- The Root Mean Square (RMS) energy:

$$\text{RMS}_t(x_t) = \left(\frac{1}{N} \sum_{k=0}^{N-1} (x_t[k])^2 \right)^{1/2}. \quad (2.6)$$

These features may be used for event detection and, in more trivial cases, event identification, or used in conjunction with spectral features. Some information extraction methods operate strictly in the time domain. Time Encoded Signal Processing and Recognition (TESPAR), for example, counts the so-called real and complex zeroes (i.e., interval duration and the number of inflection points between every two successive zero crossings) and codes them into a single vector of values using a non-linear coder [92]. However, the majority of signal processing techniques relies on time series transforms and analysis of frequency domain features [147].

2.3 Frequency Domain Analysis

Time domain analysis is well applicable to weakly oscillating and harmonic signals, however, if the signal is non-harmonic or highly polluted by noise,

which foremost influences the signal amplitude and shape, time domain features prove to be of little use. Frequency domain analysis offers a much wider range of signal analysis possibilities, e.g., examining harmonic frequencies, analysis of specific frequency bands, evaluation of frequency shifts due to the Doppler Effect, etc.

2.3.1 Discrete Fourier Transform

The discrete time domain signal is converted to the frequency domain by the Discrete Fourier Transform (DFT). For a finite duration discrete signal $x[n]$ of length N , $n = 0, \dots, N - 1$, the DFT function is given as

$$X[k] = \mathcal{F}\{x[n]\} = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}kn}, \quad (2.7)$$

where $k = 0, \dots, N - 1$. The transform is performed along two integer dimensions n (discrete time indices) and k (discrete frequency indices), i.e., it can be presented as a linear system transformation, which requires $N(N - 1)$ multiplications and a similar number of additions. Algorithms that reduce the computational complexity of DFT are known as Fast Fourier Transform (FFT) methods. They utilize the periodic and symmetric properties of the Fourier basis function $e^{-j2\pi/N}$ to avoid redundant calculations. Numerous implementations of the FFT exist; one of the most widely used and applied in MATLAB and also in the implementations considered in this thesis is FFTW, developed by Frigo and Johnson [57].

The complex frequency spectrum $[X[0], X[1], \dots, X[N - 1]]$ contains both amplitude and phase information and is symmetrically divided into complex conjugate “positive” and “negative” frequencies, the positive ones residing in the interval $[X[0], \dots, X[N/2 + 1]]$ with $X[0]$ being the signal DC component, which is ignored in this thesis. In order to obtain the absolute magnitude spectrum, the absolute values of this portion of the spectrum are calculated as

$$|X[k]| = |\mathcal{F}\{x[n]\}| / N, \quad k = 1, \dots, N/2 + 1. \quad (2.8)$$

Thus, abiding the Nyquist-Shannon sampling theorem [133], the magnitude frequency spectrum of a signal frame of length N consists of $N/2$ frequency components, each of which is multiple to the frequency resolution given by $\Delta f = f_s/N$ Hz, where f_s is the sampling rate. An example of the absolute magnitude of a periodic vibration signal is presented in Figure 2.4. The frequency resolution here is equal to $\Delta f = 1$ Hz, thus magnitudes of the frequency components are represented correctly. However, if the resolution is low, the magnitude of frequency components not multiple to the resolution will get distributed between the consecutive discrete frequency bands. To

solve the problem of magnitude disproportionate representation due to low resolution, a windowing operation, acting as a low-pass filter, is applied. Implementation-wise the division by N in (2.8) performs scaling, but does not add to the quality of further processing. Thus, it can be omitted, which will result in unscaled magnitude values (see Figure 2.4).

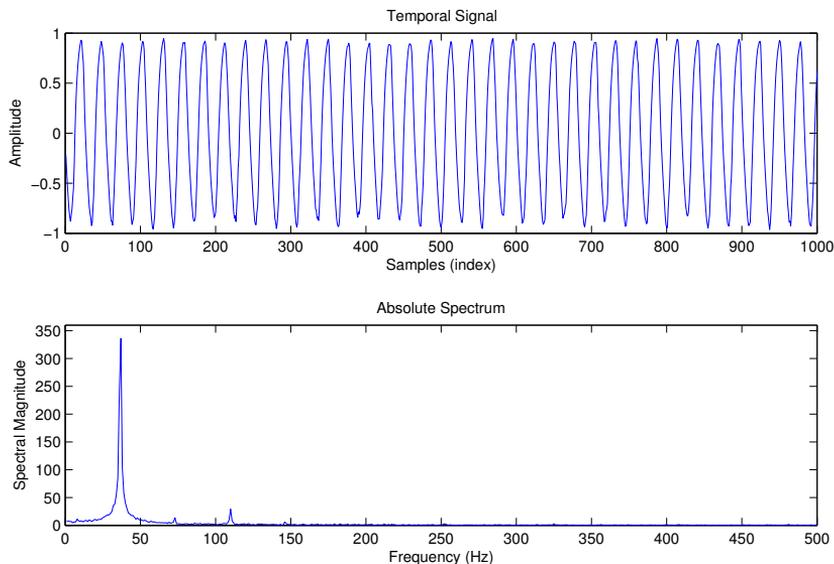


Figure 2.4: Vibration signal frame acquired at $f_s = 1$ kS/s (top), its DFT (bottom). Signal frame length $N = 1000$, spectrum length $N/2 = 500$ samples, the frequency resolution is equal to $\Delta f = 1$ Hz.

2.3.2 Time-Frequency Domain Analysis

Frequency domain analysis on embedded hardware is performed in a frame-by-frame manner by converting every acquired time domain signal frame through DFT and analyzing the frames separately. To ensure high refresh rates (less than one second) for monitoring highly dynamic processes and reduce the amounts of used memory, the frame length N is chosen accordingly to meet the functional and hardware limitations. If precise spectral distributions are required, and the frequency resolution Δf does not allow them to be achieved, a larger N may be chosen and the frames processed in a sliding window fashion (i.e., with non-zero frame overlap).

For manual signal analysis during system development it is important to establish the frequency domain signal analysis parameters. For this purpose spectrograms are used, which portray the changes in the frequency components of the signal in time. A spectrogram is obtained by applying the Short Time Fourier Transform (STFT) to a series of consecutive frames with the following parameters:

- window length in samples,
- window function (e.g., rectangular, Hamming, Hann, etc.),
- window overlap in samples.

The window function softens the disproportionate representation of frequency components due to low resolution Δf and reduces energy leakage between the short window frames [133, 147]. An example of a logarithmic scale spectrogram is presented in Figure 2.5. In this simple example the dynamical properties of the vibration process are visible in both the time and frequency domains with the signal amplitude rising according to the rise of the vibration fundamental frequency. The spectrum presented in Figure 2.4 corresponds to the 70th second of this spectrogram. In both figures the fundamental (at approximately 40 Hz) and one of its harmonics (at approximately 120 Hz) are well distinguishable.

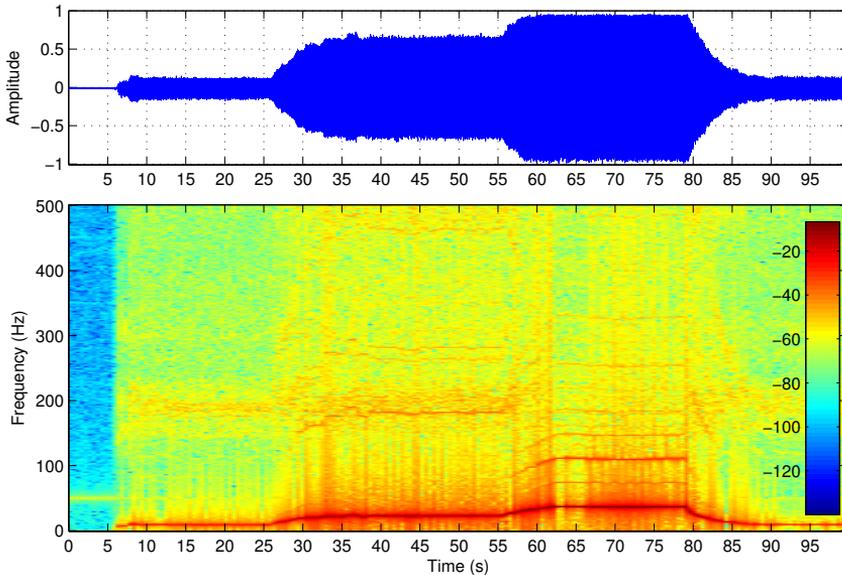


Figure 2.5: Vibration signal acquired at $f_s = 1$ kS/s (top), and its spectrogram calculated by applying STFT with the window length of $w = 512$ samples and using a rectangular windowing function (bottom).

2.3.3 Frequency Domain Features

If the observed signal is harmonic, event identification is reduced to the trivial detection of the fundamental frequency corresponding to any of the observed process states. On the other hand, if the signal is non-harmonic with a widely spread spectral distribution, other methods of analysis are applied.

What follows is a short list of some of the available spectral features, by which the observed process states can be distinguished.

Mel-Frequency Cepstral Coefficients (MFCC) is a popular method of acoustic signal feature representation mainly used in speech signal coding and recognition [159]. The process of MFCC feature extraction is presented in Figure 2.6. Mel-frequency scaling mimics human perception of sound, where lower frequencies are more distinguishable than higher frequencies. An alternative to mel scaling is Bark scaling, which also may be used in MFCC and on its own. The absolute power spectrum is mel-scaled according to

$$\text{Mel}(f) = 2595 \cdot \log_{10} (1 + f/700), \quad (2.9)$$

where f is the linear frequency, using a number of triangular filters, which produces an average scaled power value per filter (see Figure 2.7). Cepstral coefficients (describing harmonic properties of the spectrum) are calculated from these scaled power values by applying the Discrete Cosine Transform. The dynamic and transitional properties of consecutive spectra are described by delta coefficients, i.e., derivatives obtained by polynomial approximation over a finite succession of cepstral coefficient vectors.

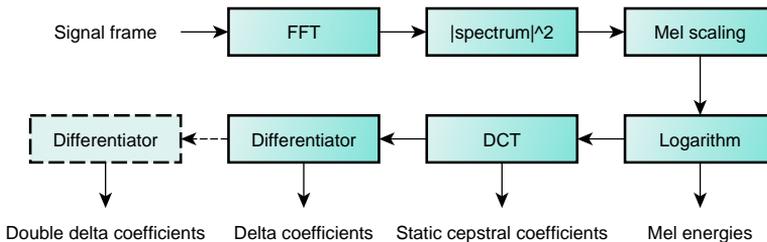


Figure 2.6: Mel-Frequency Cepstral Coefficients procedure steps.

Instantaneous frequency domain features [115] for the absolute magnitude spectrum $|X_t[k]|$, $k = 1, \dots, K$, of length $K = N/2$ acquired at time t include, but are not limited to:

- Spectral centroid, which is the first central moment in respect to frequency in a magnitude spectrum. It is calculated as the frequency averaged over the absolute magnitude spectrum:

$$\text{SC}_t(|X_t|) = \frac{\sum_{k=1}^K k \cdot |X_t[k]|}{\sum_{k=1}^K |X_t[k]|}. \quad (2.10)$$

- Spectral skewness, which is the third central moment in respect to frequency:

$$\text{SS}_t(|X_t|) = \frac{1}{K} \sum_{k=1}^K \left(\frac{|X_t[k]| - \overline{|X_t|}}{\sigma_{|X_t|}} \right)^3, \quad (2.11)$$

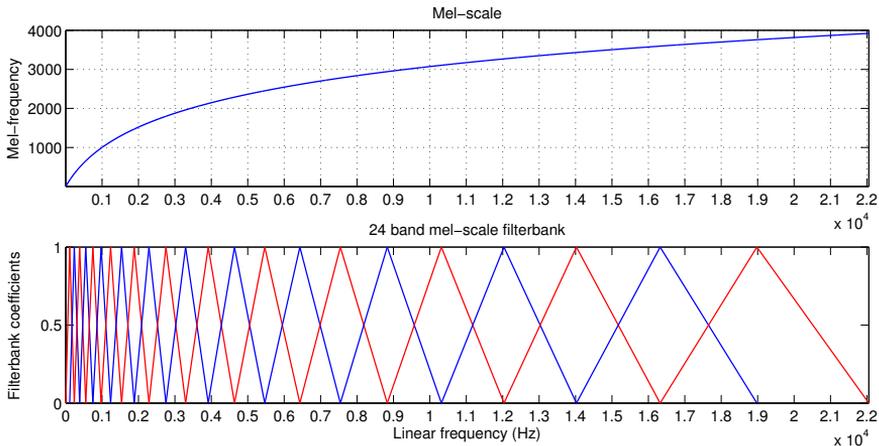


Figure 2.7: Mel-scaling of the linear frequency by a 24 band MFCC filterbank.

where $\sigma_{|X_t|}$ is the standard deviation of spectrum $|X_t|$, and $\overline{|X_t|}$ is its mean value. Skewness is a measure of asymmetry of a distribution around its mean value.

- Spectral kurtosis, which is the fourth central moment in respect to frequency:

$$\text{SK}_t(|X_t|) = \frac{1}{K} \sum_{k=1}^K \left(\frac{|X_t[k]| - \overline{|X_t|}}{\sigma_{|X_t|}} \right)^4. \quad (2.12)$$

Kurtosis gives the measure of flatness of a distribution around its mean value.

- Spectral decrease, which is a measure of decrease of the spectral magnitude:

$$\text{SD}_t(|X_t|) = \frac{1}{\sum_{k=2}^K |X_t[k]|} \cdot \sum_{k=2}^K \frac{|X_t[k]| - |X_t[1]|}{k-1}. \quad (2.13)$$

It represents the slope of the spectrum from left to right.

- Spectral roll-off specifies the frequency band below which a η amount of spectral energy resides, where $\eta \in \mathbb{R} \mid 0 < \eta < 1$ is the ratio threshold. Spectral roll-off is another measure of skewness of the spectrum and is computed as

$$\text{SR}_t(|X_t|, \eta) = \arg \max_p \left[\sum_{l=1}^p |X_t[l]|^2 \leq \eta \cdot \sum_{k=1}^K |X_t[k]|^2 \right]. \quad (2.14)$$

- Spectrum spread, also known as effective bandwidth, measures the width of the spectrum portion where the effective quantity of energy resides. It is calculated by employing the spectral centroid (2.10) as

$$\text{BW}_t(|X_t|) = \left(\frac{\sum_{k=1}^K [k - \text{SC}_t(|X_t|)]^2 \cdot |X_t[k]|^2}{\sum_{k=1}^K |X_t[k]|^2} \right)^{1/2}. \quad (2.15)$$

- Band energies measure the energy of the power spectrum at the defined bands. The band energy for the i -th band is computed as

$$\text{BE}_t(|X_t|, i) = \frac{\sum_{l \in S_i} |X_t[l]|^2}{\sum_{k=1}^K |X_t[k]|^2}, \quad (2.16)$$

where S_i is the set of power spectrum samples belonging to the i -th band.

- Spectral slope is a measure of spectral energy decrease in the direction of higher frequencies. It is determined by the gradient and y -intersect parameters of a straight line calculated applying linear regression to the magnitude spectrum. For a set of K data points $(k, |X_t[k]|)$, $k = 1, \dots, K$, the gradient of the best fitted straight line is denoted as

$$a_t(|X_t|) = \frac{K \sum_{k=1}^K k \cdot |X_t[k]| - \sum_{k=1}^K k \sum_{k=1}^K |X_t[k]|}{K \sum_{k=1}^K k^2 - \left(\sum_{k=1}^K k \right)^2}, \quad (2.17)$$

and the y -intersect is denoted as

$$b_t(|X_t|) = \frac{\sum_{k=1}^K |X_t[k]| \sum_{k=1}^K k^2 - \sum_{k=1}^K k \sum_{k=1}^K k \cdot |X_t[k]|}{K \sum_{k=1}^K k^2 - \left(\sum_{k=1}^K k \right)^2}. \quad (2.18)$$

- Spectral flux, which determines the rate of change of successive magnitude spectra:

$$\text{SF}_t(|X_t|, |X_{t-1}|) = \sum_{k=1}^K ||X_t[k]| - |X_{t-1}[k]||. \quad (2.19)$$

- Spectral variation, or sometimes also referred to as spectral flux, is another way to determine the rate of change between successive spectra. It is calculated as a normalized zero lag cross-correlation between two successive amplitude spectra as

$$\text{SV}_t(|X_t|, |X_{t-1}|) = 1 - \frac{\sum_{k=1}^K |X_{t-1}[k]| \cdot |X_t[k]|}{\sum_{k=1}^K |X_{t-1}[k]| \sum_{k=1}^K |X_t[k]|}. \quad (2.20)$$

It tends to 0 in case of similar successive spectra and to 1 otherwise.

2.4 Signal Filtering

Analog filtering is performed during analog signal conditioning prior to AD conversion (see Figure 2.1). Basically any operation on the signal (e.g., amplification) can be considered as a filter with a specific transfer function [133]. Analog filters are implemented on analog circuitry and are thus hard-wired [142]. Digital filters are applied to the signals after conversion and can be redesigned for specific applications, even at runtime.

2.4.1 Static Digital Filters

Filters are generally divided into two categories of either recursive Infinite Impulse Response (IIR), or non-recursive Finite Impulse Response (FIR). IIR filters require lower filter orders than FIR filters to achieve similar frequency and phase responses. Consequently IIR filters require less operations, compared to equivalent FIR filters. On the other hand, IIR filters have varying phase delays (constant phase delay design is difficult and not always achievable), can be unstable, and can produce limit cycles (i.e., impose fluctuations on steady-state signals) due to their recursive structure [133]. Constant phase delay is essential to phase-based signal analysis, so for phase-sensitive operations (e.g., phase shift estimation between channels) FIR filters are preferred. Filters of even order may be designed to have a zero phase delay. A more straightforward, but more computationally expensive approach lies in running the filter on a signal frame both forward and backward. This will result in the phase delay of the forward pass being canceled out during the backward pass. This approach, however, is not applicable to signal streams.

For other applications the choice between IIR and FIR filters depends on the desired frequency roll-off, amount of ripple in the pass band and the quality of attenuation in the stop band. IIR filters in general require a lower filter order to achieve better performance, however, any IIR filter can be approximated by a FIR filter of greater order. For low pass filtering a windowing function applied during DFT may be sufficient for filtering low-noise signals. For frame-by-frame processing of individual short time frames the amount of phase lag and residual fluctuations must be reduced in order not to distort frame to frame signal transitions. While applying both time and frequency domain signal analysis, the filter should be designed not to disrupt time domain signal features. For example, if the observed event produces a signal with a wide spectral distribution and a low pass filter is applied, it may reduce the time domain signal amplitude and make the event indistinguishable in the time domain, e.g., by envelope.

An example of low pass filtering of the vibration signal presented in Figure 2.5 is portrayed in Figure 2.8. Noise attenuation of 100 dB is achieved by applying a 12th order equiripple FIR filter with the cutoff frequency of

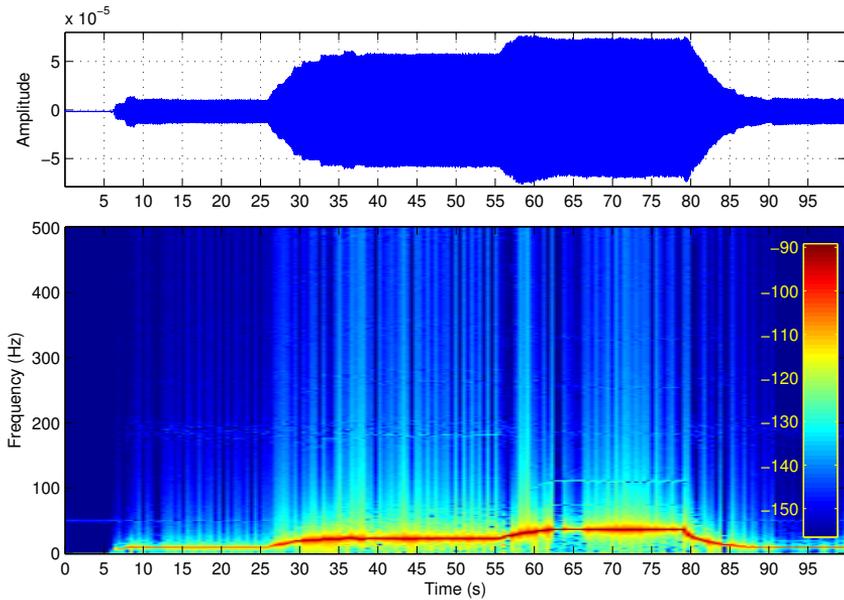


Figure 2.8: Result of low pass filtering of the signal presented in Figure 2.5.

50 Hz, transition band of 50–80 Hz and constant phase delay of 6 samples. The frequency and phase response (in normalized frequency) of the applied filter are presented in Figure 2.9. Although the vibration fundamental frequency on the logarithmic scale is more distinguishable in Figure 2.8 than in Figure 2.5, such a large attenuation combined with low filter order result in attenuation of 80 dB in the pass band and large ripples in the stop band. As a result the signal loses a great amount of energy, and its amplitude is reduced by several orders of magnitude. This trade-off is completely permissible if event identification is performed by only analyzing the fundamental.

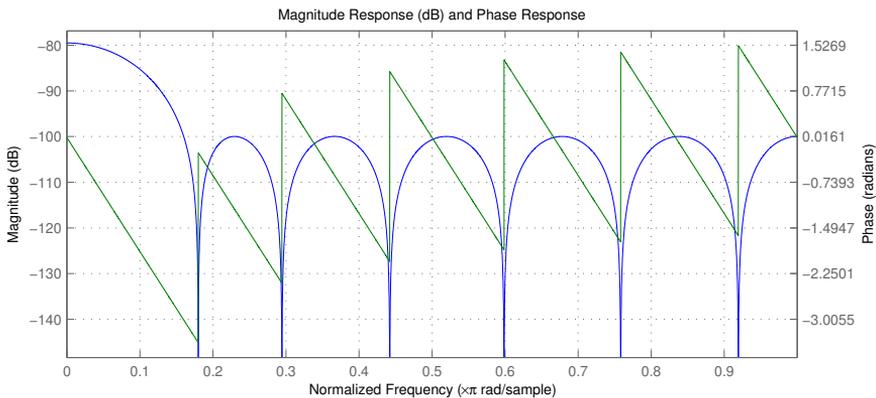


Figure 2.9: Frequency and phase response of the filter used in Figure 2.8.

Other signal features, on the other hand, can be corrupted to the extent of being unusable.

2.4.2 Adaptive Filtering

Applying pre-designed filters for the duration of system on-line operation is considered static filtering. It is useful if the noise band is known in advance and does not change dramatically during system operation. Adaptive filters, on the other hand, have the ability to tune their parameters at system runtime by applying an optimization algorithm with a specifically designed cost function in a feedback loop [1]. Highly correlated noise (e.g., echo or channel cross-talk) is reduced by applying adaptive feedback cancellation [28]. Another approach to adaptive filtering lies in adaptive noise cancellation, which requires knowledge of the signal and noise properties. The properties may be established by acquiring the signal from several process observation points with noise being significantly more powerful at one of these points [28].

However, in complex and unpredictable environments the signal and noise parameters may change rapidly, and thus reliable continuous adaptive filtering poses a problem. In an ever-changing environment the optimization options (e.g., the cost function) may become obsolete, which will result in optimization divergence and disrupt further signal processing. On the other hand, adaptive noise cancellation would be a viable option, as it uses the physical and not approximated noise reference, however, it is applicable only when the measurement points can be covered by one smart sensor in order not to transmit raw signals between WSN nodes.

Distributed systems with a large number of nodes have an advantage in regard to handling unexpected noise. While signal analysis in one group of sensors may get corrupted by unexpected noise types and signal masking, another group may not acquire the same noise and produce reliable results. Furthermore, in multi-modal systems a noise source that disrupts operation of sensors of one modality, may be totally unnoticed by sensors of other modalities. Therefore, in distributed CPS the negative influence of unexpected noise is overcome during the stages of data fusion, rather than during filtering and noise suppression on individual nodes [156].

2.5 Pattern Recognition

Local situation assessment in terms of identifying an event or a state of the observed process is performed via pattern recognition [106]. The general steps of pattern recognition performed at system runtime and the information required at every step are presented in Figure 2.10. Depending on whether the signal is analyzed in the time, frequency, or both domains, parameters for signal filtering, FFT, or any other transformation, e.g., signal fusion, are

provided to the smart sensor. Signal representations ready for analysis are then put through feature extraction, which computes a compact set of signal parameters (i.e., features) that represents the signal patterns corresponding to the events of interest, while being influenced by noise as little as possible. The extracted feature vector is then classified and a decision, usually represented by a class label, is made for the time instance at hand.

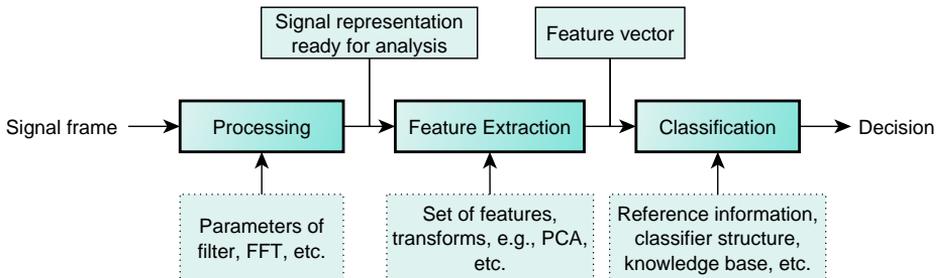


Figure 2.10: General steps of signal processing for pattern recognition.

2.5.1 Feature Extraction

During feature extraction a feature vector $\chi = [\chi_1, \chi_2, \dots, \chi_L]$ of length L (which is considered the dimensionality of the feature space) is generated for every incoming signal frame (or set of consecutive frames). It may contain features derived from both time and frequency domains; at this stage all features can be viewed as source invariant information. The feature vector may thus contain any set of time domain features, e.g., (2.3)–(2.6), frequency domain features, e.g., (2.10)–(2.20), MFCC features, or any other signal parameters that describe the classified pattern.

Combining features of different origin instead of using, e.g., spectral distribution described by MFCC or band energies (2.16), can lead to unbalanced feature sets with significant variance differences between features and highly correlated features. As a result features with greater variance will influence the classification decision more, and the correlated features will most likely be redundant. The problem is solved by either rigorously analyzing and choosing the best fit features, or applying feature selection algorithms, e.g., Differential Evolution Feature Selection (DEFS) [79], Least-Squares Feature Selection (LSFS) [74], or Multi-Cluster Feature Selection (MCFS) [35], and orthogonal transformations, e.g., Principal Component Analysis (PCA), or Independent Component Analysis (ICA) [147].

Feature selection methods determine the best fit and least correlated features from a set of features by, e.g., applying a population based method for confining an optimal subset of features (DEFS), or minimizing the least-squares mutual information between features (LSFS). PCA transforms the

feature vectors to new sets of orthogonal dimensions with a diagonal covariance matrix, reducing the dimensionality L of the new vectors if required, and leaving only the most important components. However, PCA assumes normal distribution and de-correlates only the second-order feature statistics. ICA, on the other hand, de-correlates the feature vectors in higher-order statistics (usually the fourth order, i.e., kurtosis) of the feature set and assumes non-normal feature distribution. Both transforms also bring the features to unit variance.

Feature selection and derivation of the transformation matrices for PCA or ICA are performed offline during system design on a pre-acquired set of features. During online system operation the transformation matrix is then applied to each feature vector $\boldsymbol{\chi}$ at every discrete time instance. Online feature selection and transformation matrix derivation is complicated due to a large amount of features needed to be gathered on the embedded device for statistical analysis correctness and the risk of procedure divergence from an optimal solution due to unexpected noise sources and unknown events arising in the monitored environment. Furthermore, changes made to the features must be accounted for during classification, thus the classification procedure needs to be appended to account for these changes.

2.5.2 Classification

Classification can be performed in various ways, from calculating the similarity metric (e.g., correlation, maximum likelihood) between the reference and the incoming feature vectors, to applying model-based classifiers (e.g., Artificial Neural Networks [113]) and multistage classification trees (e.g., CART [127]), or probabilistic classifiers and predictors (e.g., naive Bayes [81], Hidden Markov Models [55]). The simplest approach to classification lies in calculating the correlation between the feature vector $\boldsymbol{\chi} = [\chi_1, \chi_2, \dots, \chi_L]$ and C reference vectors $\mathbf{r}_i = [r_i(1), r_i(2), \dots, r_i(L)]$, $i = 1, \dots, C$, as

$$\rho_i = \frac{L \sum_{l=1}^L \chi_l \cdot r_i(l) - \sum_{l=1}^L \chi_l \sum_{l=1}^L r_i(l)}{\sqrt{L \sum_{l=1}^L \chi_l^2 - \left(\sum_{l=1}^L \chi_l\right)^2} \sqrt{L \sum_{l=1}^L (r_i(l))^2 - \left(\sum_{l=1}^L r_i(l)\right)^2}}, \quad (2.21)$$

where the winning class y is defined as

$$y = \arg \max_{1 \leq i \leq C} (\rho_i). \quad (2.22)$$

Model-based classifiers divide the feature sub-spaces occupied by features of different classes and during classification estimate the degree of belonging of the L -dimensional feature (i.e., feature vector of length L) to each of these L -dimensional sub-spaces [106]. Classification trees separate the feature sub-spaces by bounds, one feature component at a time, forming a decision tree.

Probabilistic classifiers use conditional probability properties to estimate the probability of the class label at any time instance, given the history of previous classifications. The application of Bayesian estimators, like Hidden Markov Models (HMM), to classification tasks allows to identify finite state non-stationary processes, e.g. spoken phrases [145].

Classifier structure is derived during the training stage using training datasets containing P feature vectors

$$\mathbf{X} = \left[\begin{array}{cccc|c} \chi_{1,1} & \chi_{1,2} & \cdots & \chi_{1,L} & c_1 \\ \chi_{2,1} & \chi_{2,2} & \cdots & \chi_{2,L} & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \chi_{P,1} & \chi_{P,2} & \cdots & \chi_{P,L} & c_P \end{array} \right], \quad (2.23)$$

where a reference class label c_i may be added to each feature vector for supervised learning procedures. If the reference class labels are not *a priori* known, the feature space is divided using unsupervised learning procedures, e.g., clustering [2]. Initial classifier training is performed during system design and can be updated during online operation by applying online machine learning procedures. However, self-learning tends to reduce the classification accuracy over time in situations where high uncertainty is involved [106]. In multi-component systems, where the decision is made based on the assessments from many observation points, the nodes with higher certainty of correct classification can provide the reference class labels to nodes with uncertain classifier outputs in order to maintain overall classification reliability.

2.5.3 Fuzzy Logic Based Classification

The main classification algorithm applied in this thesis is based on fuzzy logic [125]. It is similar to model-based classifiers, in that it separates the feature sub-spaces in the L -dimensional feature space, but applies fuzzy inference for class label estimation. The inference is performed by assigning one of C different discrete valued labels to the feature vector $\boldsymbol{\chi} = [\chi_1, \chi_2, \dots, \chi_L]$ of length L . The initial classifier consists of R rules of the following structure:

$$\begin{array}{l} \text{IF } \chi_1 \text{ is } A_{1r} \text{ AND } \chi_2 \text{ is } A_{2r} \text{ AND } \dots \text{ AND } \chi_L \text{ is } A_{Lr} \\ \text{THEN } y \text{ belongs to class } c_r, \end{array} \quad (2.24)$$

where A_{ir} is the linguistic term of the i -th input (i.e., feature vector element, $i = 1, \dots, L$) associated with the r -th rule, and $c_r \in \{1, \dots, C\}$ is the class label assigned by the r -th rule. The form that contains only one rule per class is called the Minimal Rule Classifier (MRC).

The training of the classifier is performed in a supervised manner on datasets of structure (2.23) in the following steps:

1. Initialization: derivation of the MRC;
2. Rule base expansion: misclassification rate reduction by iterative splitting of rules with most erroneous results in two;
3. Consolidation: linkage of small rules resulting from consolidation to rules covering more samples, while sustaining classification accuracy;
4. Compression: removal of redundant conditions in rules;
5. Conversion: changing the membership function type (optional).

MRC training consists of estimating the parameters of Membership Functions (MF) that quantify the linguistic terms in rules (2.24). For classifier initialization triangle-shaped MF are used:

$$\mu_{ir}(\chi_i) = \begin{cases} \frac{\chi_i - a_{ir}}{b_{ir} - a_{ir}}, & a_{ir} \leq \chi_i \leq b_{ir}; \\ \frac{c_{ir} - \chi_i}{c_{ir} - b_{ir}}, & b_{ir} < \chi_i \leq c_{ir}; \\ 0, & (\chi_i < a_{ir}) \vee (\chi_i > c_{ir}), \end{cases} \quad (2.25)$$

where a_{ir} and c_{ir} locate the base of the triangle and b_{ir} locates its peak. The training is performed using a set of reference feature vectors (2.23), for which a class label is provided manually. This is done during system offline tuning. The procedure consists of the following steps:

1. The set of vectors is partitioned into R subsets S_j , $j = 1, \dots, R$, each consisting of P_j vectors of the same class.
2. MF parameters are calculated as $a_{ir} = \min_{k \in S_j} (\chi_i(k))$, $c_{ir} = \max_{k \in S_j} (\chi_i(k))$,
 $b_{ir} = \frac{1}{P_j} \sum_{k \in S_j} \chi_i(k)$, $i = 1, \dots, L$.
3. The base of each MF is slightly enlarged (e.g., by 1%) to give non-zero membership values to the training samples located at the edges of multidimensional clusters [125].
4. The established MF are added to the classifier rule-base defined by (2.24).

During the steps of rule base expansion through to compression the structure of the MRC is changed in order to ensure maximal training dataset coverage, while minimizing the number of misclassifications [124, 126]. The last step of conversion is performed if the classifier is required to operate on samples that fall beyond the rule borders specified by the triangular MF base parameters. In this case smooth MF, e.g., the double Gaussian MF, which have smooth transitions to zero membership, can be applied instead of triangular MF,

which have distinct points of zero membership. In this manner triangular MF are replaced with near-equivalent Gaussian curves defined as

$$\mu_{ir}(\chi_i) = \begin{cases} \exp \left\{ -\frac{(\chi_i - b_{ir})^2}{2 \cdot (0.4247 \cdot (b_{ir} - a_{ir})^2)} \right\}, & \chi_i < b_{ir}, \\ \exp \left\{ -\frac{(\chi_i - b_{ir})^2}{2 \cdot (0.4247 \cdot (c_{ir} - b_{ir})^2)} \right\}, & \chi_i \geq b_{ir}. \end{cases} \quad (2.26)$$

The inference is performed (i.e., the class label is assigned) in a winner-takes-all manner by specifying the rule with the highest degree of activation

$$y = c_r, \arg \max_{1 \leq r \leq R} (\tau_r), \quad (2.27)$$

where τ_r is the activation degree of the r -th rule:

$$\tau_r = \bigcap_{i=1}^L \mu_{ir}(\chi_i), \quad (2.28)$$

where μ_{ir} is the MF corresponding to the linguistic term A_{ir} , and the intersection (linguistic AND) is defined either by the minimum, or product operation.

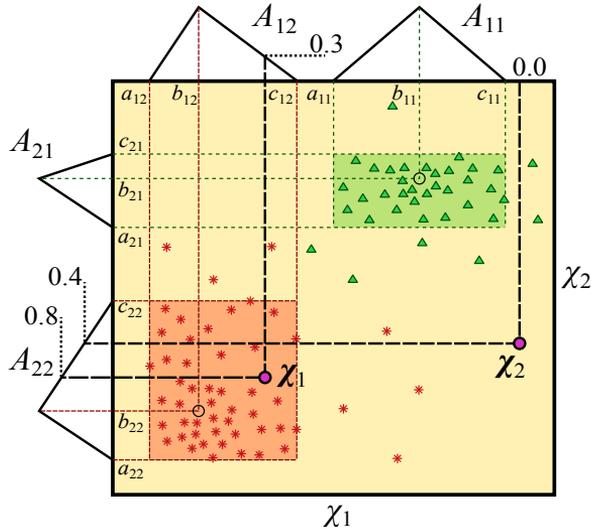


Figure 2.11: Example of classification by fuzzy inference.

An example of fuzzy inference for a two-dimensional feature space and two incoming feature vectors χ_1 , χ_2 , is presented in Figure 2.11. The features are separated in two classes by the linguistic terms A_{11} and A_{21} for rule 1, governing class 1, and by A_{12} and A_{22} for rule 2, governing class 2. The vector χ_1 belongs to class 2 and χ_2 does not belong to any of

the two classes. According to (2.27) and (2.28), and applying the minimum operation for intersection, the activation degrees and class labels are assigned in the following manner. For \mathbf{x}_1 the activation degree of rule 1 is $\tau_1(\mathbf{x}_1) = \min \{\mu_{11}(\chi_1), \mu_{21}(\chi_2)\} = \min \{0, 0\} = 0$, and the activation degree of rule 2 is $\tau_2(\mathbf{x}_1) = \min \{\mu_{12}(\chi_1), \mu_{22}(\chi_2)\} = \min \{0.3, 0.8\} = 0.3$. The class label of \mathbf{x}_1 is then $y(\mathbf{x}_1) = \arg \max \{\tau_1(\mathbf{x}_1), \tau_2(\mathbf{x}_1)\} = \arg \max \{0, 0.3\} = 2$. Following this logic, for \mathbf{x}_2 the activation degrees of both rules 1 and 2 are $\tau_1(\mathbf{x}_2) = 0$, $\tau_2(\mathbf{x}_2) = 0$, and the class label assigned to \mathbf{x}_2 is thus 0, which means that \mathbf{x}_2 does not belong to any of the two classes. Figure 2.11 also portrays how the classifier reduces misclassification and class sub-space overlap by discarding outlying training samples.

2.6 Acoustic Localization

Acoustic localization is performed by estimating the bearing and distance to the localized object. Planar localization consists of estimating the azimuth ϕ , or the horizontal Angle of Arrival (AOA), and the distance r . Volumetric bearing estimation is performed by estimating the Direction of Arrival (DOA) of acoustic wavefronts, which consists of the azimuth ϕ and elevation θ AOA. Both approaches with two estimated parameters are considered to be 2D localization in spherical coordinates. Total 3D localization in spherical coordinates consists of estimating all three parameters (r, θ, ϕ) , as portrayed in Figure 2.12.

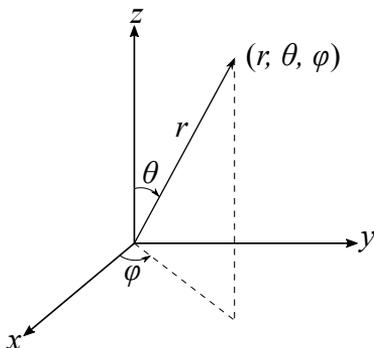


Figure 2.12: Volumetric localization parameters.

Acoustic localization is performed by multi-sensor systems, consisting of microphone arrays, by applying phase shift analysis between signals and estimating the Time Difference of Arrival (TDOA) of acoustic wavefronts [28]. To calculate the geometry of wavefront arrival, either a near-field, or a far-field signal source disposition assumption is made. Acoustic wavefronts are concentric by nature, and the near-field assumption implies that the geometric distance between wavefronts at separate measurement points is calcu-

lated as the difference between the radii of these wavefronts (see Figure 2.13, right). The far-field assumption, on the other hand, implies that the waves are coming from a greater distance and their fronts are spread enough to be considered linear (see Figure 2.13, left). The far-field assumption is met for a linear microphone array if the inequality

$$|r| > \frac{2(Ml)^2}{\lambda_{\min}} \quad (2.29)$$

holds, where M is the number of microphones, l is the distance between consecutive microphones m_i and m_{i+1} , λ_{\min} is the minimal wave length of the wide-band acoustic signal, and r is the radial distance from the array center to the source. For wide aperture arrays, where the microphones are significantly spaced from one another, the area of operation has to be known in advance to make the correct assumption of source disposition. If the microphones are compactly placed on a single smart sensor, the far-field assumption is made for the majority of applications.

The calculation of DOA relies on the speed of sound c , which depends on the ambient temperature. The speed of sound in air is calculated as

$$c = 331.45\sqrt{1 + t^\circ/273}, \quad (2.30)$$

where t° is the temperature in degrees Celsius. If the temperature of the environment of system operation is not constant (which is usually the case in outdoor environments), the speed of sound needs to be regularly checked to ensure correct DOA estimation.

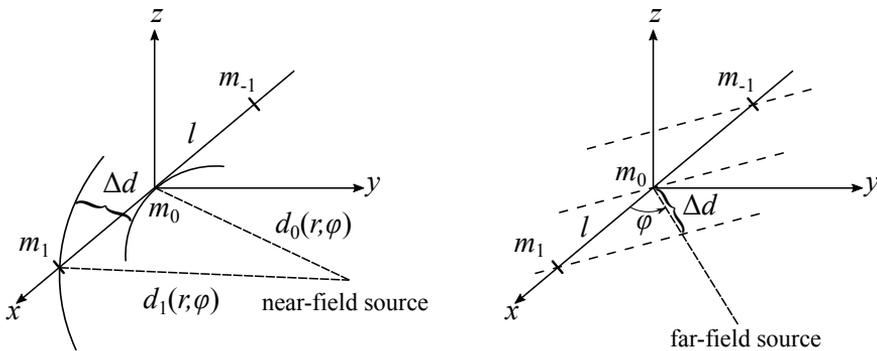


Figure 2.13: Near-field and far-field disposition of the acoustic source.

There exists a variety of methods for acoustic localization, most of which employ TDOA as a basic principal [28,33]. The methods utilize sensor array structures, in which a number of microphones is arranged in some specific manner (e.g., linear, tetrahedron, spherical, etc.). The TDOA and consequently DOA is generally estimated using some measure of correlation between the signals. For example, a popular method of Steered Response Power

with Phase Transform (SRP-PHAT) computes cross-correlation across all pairs of microphones at the theoretical time delays associated with all possible DOA [47]. Multiple Signal Classification (MUSIC) applies eigenspace analysis to the signal correlation matrix in order to get the largest eigenvalues corresponding to the most probable DOA [71]. Multilateration methods calculate distances from every sensor to the source using TDOA and estimate the position of that source by solving systems of non-linear equations [91].

Typical acoustic localization methods utilize information from every sensor. This does not pose a problem for wired systems with a single powerful computational hub. In distributed WSN, however, collecting raw signals from nodes is a real challenge, especially if the number of nodes is large and signal frames are long. Recent developments in distributed localization combine individual sensor estimates for source positioning by applying, for example, maximum likelihood iterative search [30], fuzzy clustering [100, 101], and various others [39, 46, 68, 90, 97, 146]. This thesis proposes a distributed localization approach with SRP-PHAT and multilateration used for comparison where possible.

2.6.1 Conventional SRP-PHAT

Steered Response Power with Phase Transform (SRP-PHAT) is one of the most effective acoustic DOA estimation methods for reverberant environments, proposed in [47]. The SRP $P(\mathbf{a})$ is a real-valued functional of a spatial vector \mathbf{a} , which is defined by the FOV of a specific array. The maxima of $P(\mathbf{a})$ indicate the direction to the sound source. $P(\mathbf{a})$ is computed as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) across all pairs of sensors at the theoretical time delays, associated with the chosen direction. Consider a pair of signals $x_k(t)$, $x_l(t)$ of an array consisting of M microphones. The time instances of sound arrival from some point $a \in \mathbf{a}$ for the two microphones are $\tau(a, k)$ and $\tau(a, l)$, respectively. Hence the time delay between the signals is $\tau_{kl}(a) = \tau(a, k) - \tau(a, l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega\tau_{kl}(a)} d\omega, \quad (2.31)$$

where $X_i(\omega)$ is the spectrum (i.e., the Fourier Transform) of signal $x_i(t)$, $X_i^*(\omega)$ is the conjugate of that spectrum and Ψ_{kl} is the PHAT weight, defined as

$$\Psi_{kl} = (|X_k(\omega) X_l^*(\omega)|)^{-1}. \quad (2.32)$$

The PHAT is an effective weighting of a GCC for finding TDOA from signals in a highly reverberant environment, though, it can over-sharpen the SRP.

A more flexible approach lies in applying the β -PHAT weight, defined as

$$\Psi_{kl}(\beta) = (|X_k(\omega)X_l^*(\omega)|)^{-\beta}. \quad (2.33)$$

SRP-PHAT is a more general tool for acoustic localization and DOA estimation [47]. Computing the SRP for every point in the area or volume \mathbf{a} results in a SRP image of the whole observable FOV. This enables localization of multiple sound sources, while also reducing reverberation and masking effects. In this approach the FOV area or volume is discretized into a spatial grid, with a SRP value computed for every point. Though the approach is most comprehensive, the immense number of computations it requires makes the process extremely slow and resource demanding.

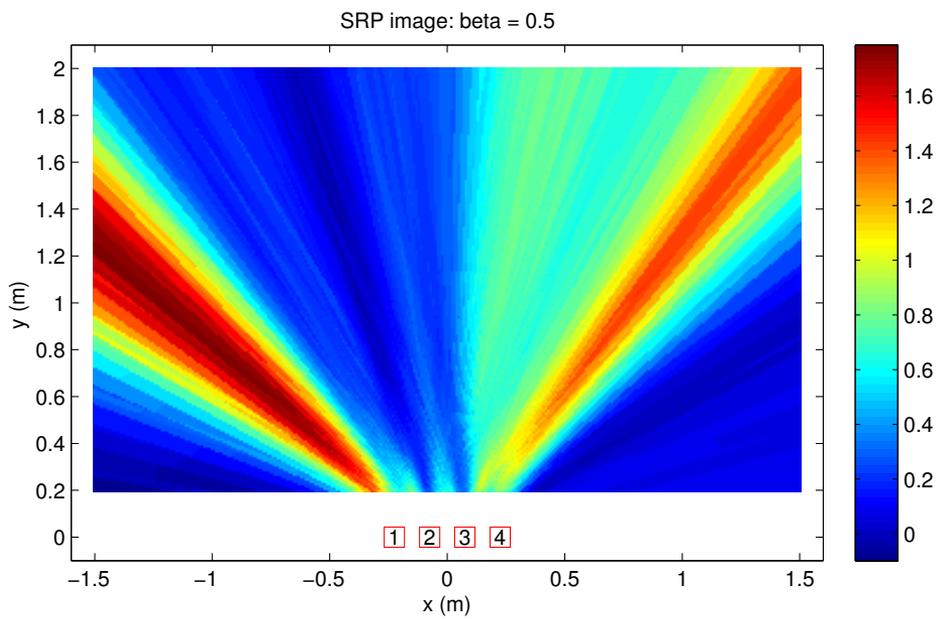
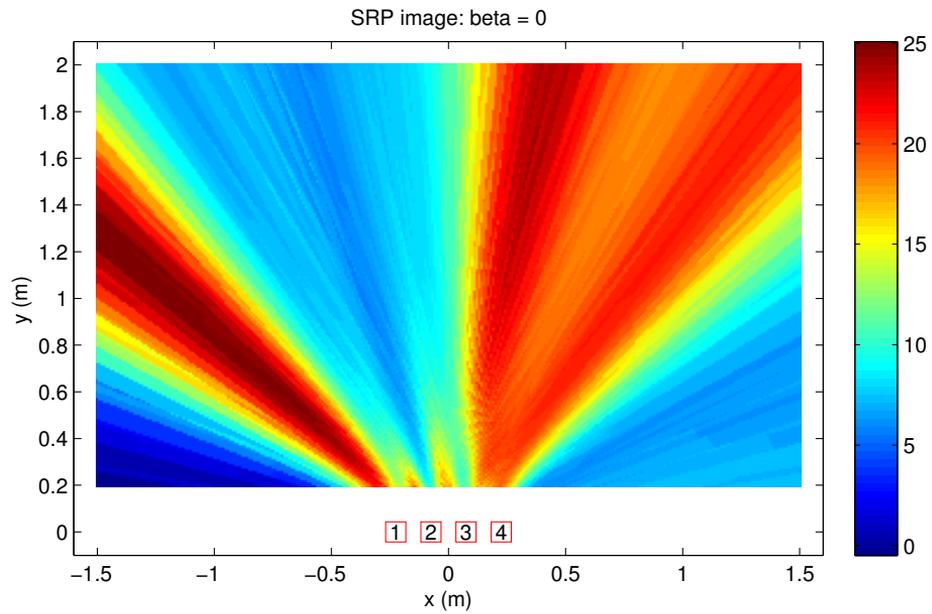
While using a narrow-aperture microphone array for a wide FOV, the DOA of multiple acoustic sources are well distinguishable with SRP images computed on the entire FOV. Figure 2.14 portrays an SRP image of one signal frame of a man and woman speaking simultaneously at -50 and 30 degrees relative to the y -axis, respectively, acquired by an array of microphones denoted by red numbered squares. The effect of PHAT whitening is evident from the figure, with reverberation effects being reduced with the increase in the β coefficient of (2.33). The image corresponding to $\beta = 1$, equivalent to applying (2.32), shows excessive sharpening. Generally $\beta \simeq 0.8$ produces smoother images, while retaining more cumulative energy, which is reduced with the increase in β .

The FOV used in Figure 2.14 is 3×1.8 m, discretized into a planar grid with a distance step of 0.04 m along every axis. Thus the planar vector \mathbf{a} consists of $(3 \cdot 1.8)/0.04^2 = 3375$ discrete directions, and so the evaluation (2.31) must be performed the according number of times. Such images cannot be by any means generated in real-time by embedded sensors, while also being inapplicable for autonomous analysis (i.e., the grid search for elevated SRP values and DOA estimation from them is a needlessly complicated task).

By using a wide-aperture array a more precise localization is achievable, as portrayed in Figure 2.15. The PHAT weight again is shown to successfully reduce reverberation. The number of evaluations (2.31) here is $(4.6 \cdot 3.8)/0.02^2 = 43700$ with the chosen FOV and discretization step of 0.02 m. It is evident that conventional SRP-PHAT is not applicable for real-time evaluation of even the smallest (several square meters) FOV. Several propositions have been made for SRP optimization, e.g., [41, 49, 158]. For this work the method of locating high maxima of SRP energy by applying Stochastic Region Contraction (SRC) [48, 49] is chosen.

2.6.2 SRP-PHAT with Stochastic Region Contraction

The number of SRP evaluations (2.31) is significantly reduced by applying Stochastic Region Contraction (SRC), which iteratively reduces the search



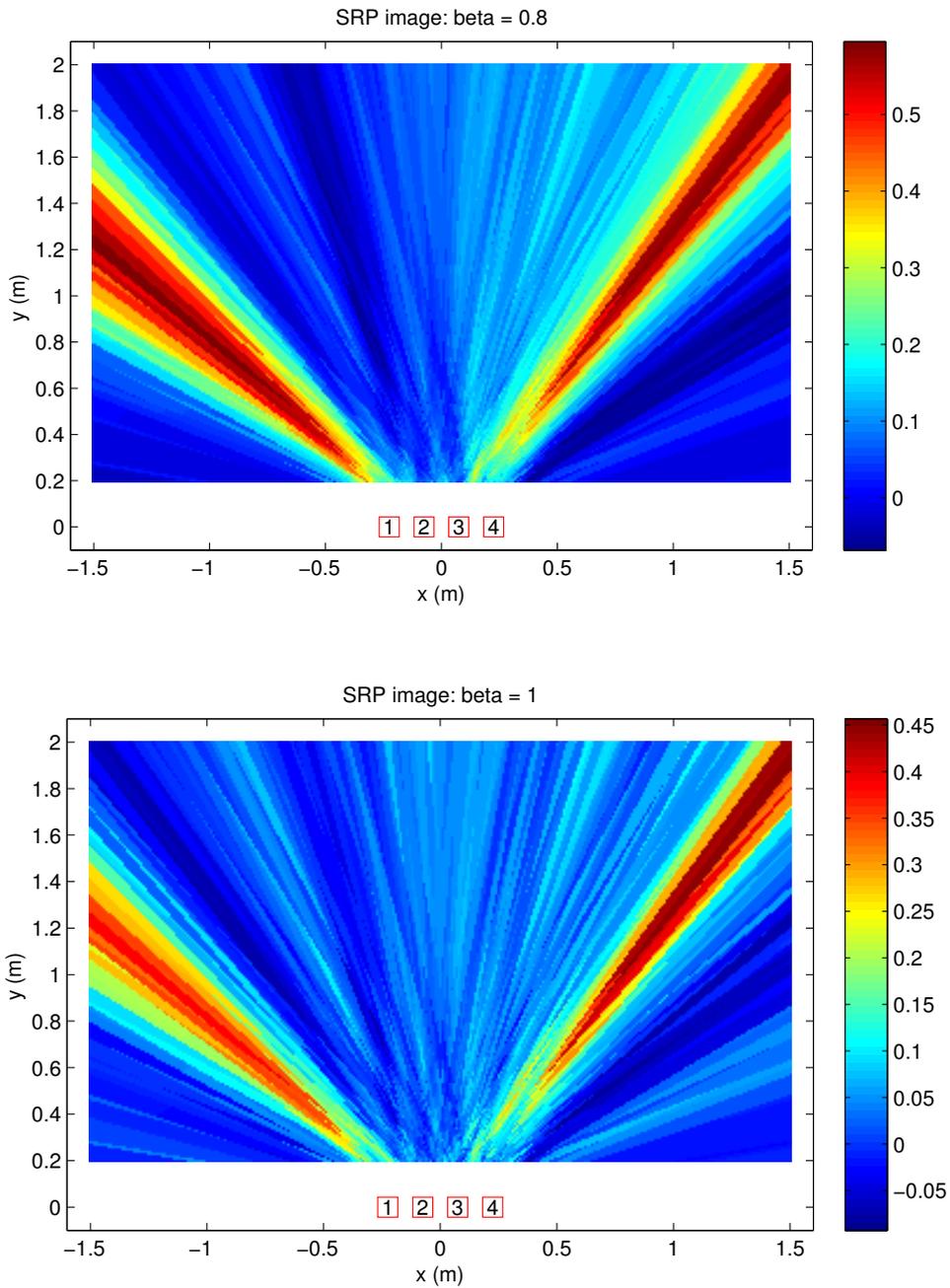


Figure 2.14: SRP image of two acoustic sources: a man and woman speaking simultaneously at -50 and 30 degrees relative to the y -axis, respectively. Different β values for (2.33) are used for the same signal frame.

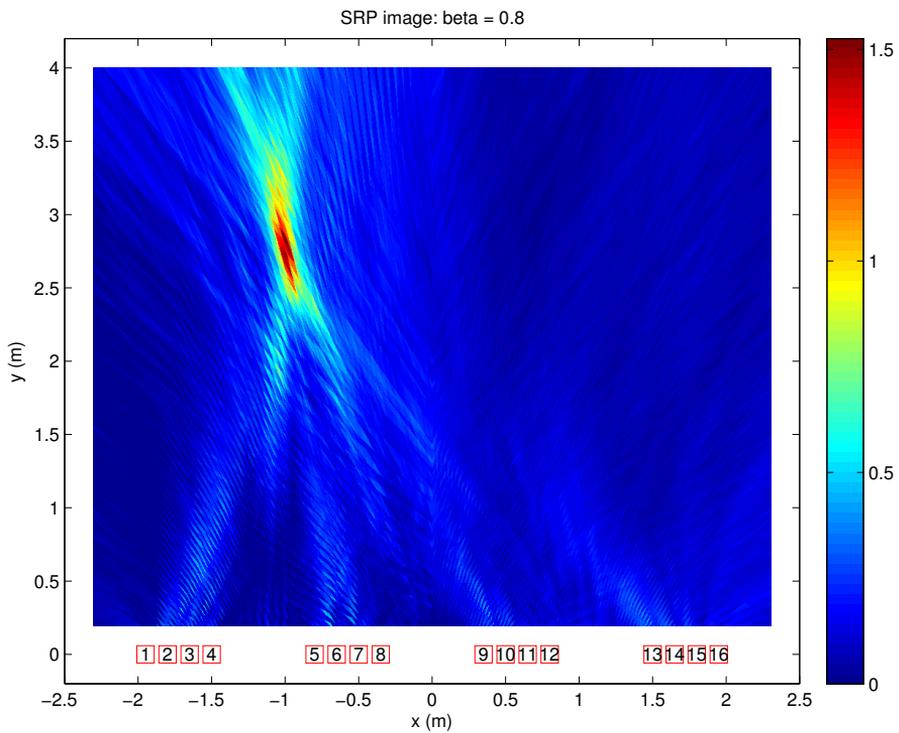
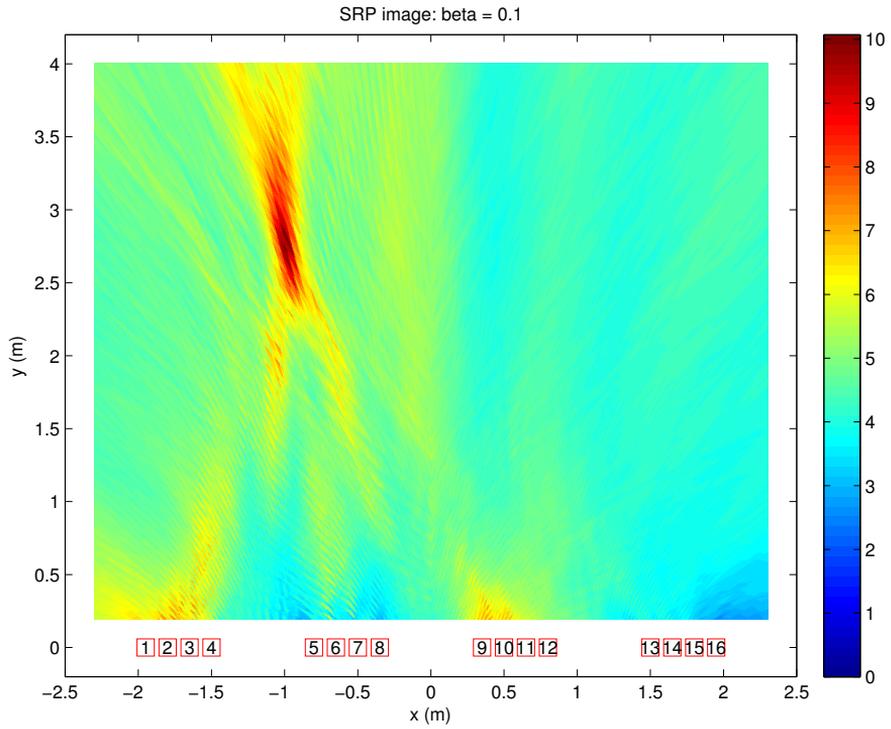


Figure 2.15: SRP image of a single acoustic source, generated by a wide-aperture microphone array.

volume for the global maximum. SRC starts with the initial search volume (i.e. the whole FOV), stochastically explores the functional of that volume by randomly picking a specific number of points, and then contracts the volume into the sub-volume containing the desired global optimum and proceeds iteratively until the global maximum can be located with a finite precision [49]. The procedure may be described in pseudo-code as:

1. Initialize iteration $i = 0$.
2. Set initial parameters: $V_0 = V_{\text{FOV}}$ — initial volume; J_0 — the number of random points that need to be evaluated to ensure, that one or more is likely to reside in the sub-volume of higher values, surrounding the global maximum; N_0 — number of points used to define the new sub-volume V_{i+1} .
3. Calculate $P(\mathbf{a})$ for J_i points.
4. Sort out the best (highest) $N_i \ll J_i$ points.
5. Contract the search volume to the smaller volume V_{i+1} , defined by a rectangular boundary vector $B_{i+1} = [x_{\min}(i+1), x_{\max}(i+1), y_{\min}(i+1), y_{\max}(i+1), z_{\min}(i+1), z_{\max}(i+1)]$, that contains these N_i points.
6. **IF** $V_{i+1} < V_u$ (a sufficiently small sub-volume, in which the global optimum is contained) **AND** $\text{FE}_i < \Phi$ (the total number of evaluations (2.31) for iteration i is less than the maximum number of allowed evaluations), **THEN** determine the global maximum, **STOP**.
7. **ELSE IF** $\text{FE}_i \geq \Phi$, **STOP**, discard results.
8. **ELSE** Among the N_i points keep a subset G_i of points, which have values greater than the mean μ_i of the N_i points.
9. Evaluate J_{i+1} new random points in V_{i+1} .
10. Form the set of N_{i+1} as the union of G_i and the best $N_{i+1} - G_i$ points from the J_{i+1} just evaluated. This gives N_{i+1} high points for iteration $i + 1$.
11. $i = i + 1$, **GO TO** Step 5.

There are several proposed ways of selecting N_i and J_i depending on the specific FOV and on the condition of monotonic or non-monotonic increase of the mean μ_i . The one emphasized in [49] consists of fixing N_i and adjusting J_i incrementally in the following fashion: N_i is chosen as $N_i \equiv N = 100$; J_i is the number of evaluations (2.31) to find $N - G_i$ points greater than μ_i . In this thesis a different approach is proposed.

2.6.3 Multilateration

Multilateration is a technique of estimating object position coordinates based on TDOA information. The distance between the sensor i with coordinates (x_i, y_i, z_i) and the signal source can be defined as a vector length

$$d = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}, \quad (2.34)$$

where (x, y, z) are the unknown source coordinates. Thus, knowing the TDOA τ_{ij} between two sensors i and j , the distance difference between sensor i and the source, and sensor j and the source is calculated as

$$d_{ij} = c \cdot \tau_{ij} = c(\tau_i - \tau_j) = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} - \sqrt{(x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2}, \quad (2.35)$$

where c is the speed of sound in air, (x, y, z) are source coordinates, and (x_i, y_i, z_i) and (x_j, y_j, z_j) are the coordinates of sensors i and j , respectively [91]. For any group consisting of G sensors the acoustic source is localizable by the following system of $G - 1$ nonlinear equations:

$$\begin{cases} d_{1,2} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2} - \sqrt{(x_2 - x)^2 + (y_2 - y)^2 + (z_2 - z)^2} \\ d_{1,3} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2} - \sqrt{(x_3 - x)^2 + (y_3 - y)^2 + (z_3 - z)^2} \\ \dots \\ d_{1,G} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2} - \sqrt{(x_G - x)^2 + (y_G - y)^2 + (z_G - z)^2} \end{cases}$$

To estimate the solution to this system of nonlinear equations at least four sensors are needed; this yields three TDOA values $\tau_{1,2}$, $\tau_{1,3}$, $\tau_{1,4}$, and the system is solved by applying a least squares method [108], e.g., Levenberg-Marquardt. Various practical approaches exist, e.g., [25, 34]. For ground applications the solution can be simplified by specifying constant z dimension and denoting the unknown source location as (x, y) . Then the minimal number of sensors needed is reduced to three.

2.7 Trajectory Estimation and Prediction

Trajectory estimation can be considered as a procedure of filtering sequential position estimates in order to achieve a steady representation of object movement. Both spatial and temporal errors are introduced to the position estimates due to localization errors and imprecision of WSN synchronization, respectively. Treating the sequence of position estimates as a time-invariant data series and applying curve fitting and smoothing will reduce spatial jitter, however, the result will not be an accurate representation of a time-variant trajectory. Recursive Bayesian estimators, like the Kalman filter or Particle filter, handle the noise imposed by the inaccuracies of the dynamic process.

2.7.1 Kalman Filter

The discrete time Kalman filter (KF) is a linear quadratic estimator [1], which provides the closed form recursive solution for a linear discrete-time dynamic system of the form:

$$\begin{aligned}\mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_{k-1}\mathbf{x}_k + \mathbf{r}_{k-1}\end{aligned}, \quad (2.36)$$

where \mathbf{x}_k is the system state vector at time step k , \mathbf{y}_k is the measurement vector at k , \mathbf{A}_{k-1} is the transition matrix of the dynamic model, \mathbf{H}_{k-1} is the measurement matrix, $\mathbf{q}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ is the process noise with covariance \mathbf{Q}_{k-1} and $\mathbf{r}_{k-1} \sim \mathcal{N}(0, \mathbf{R}_{k-1})$ is the measurement noise with covariance \mathbf{R}_{k-1} . Kalman filtering consists of a prediction step, where the next state of the system is predicted given the previous measurements, and an update step, where the current state is estimated given the measurement at that time instance. The prediction step is characterized by the following equations:

$$\begin{aligned}\hat{\mathbf{x}}_k^- &= \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{A}_{k-1}\mathbf{P}_k\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}\end{aligned}, \quad (2.37)$$

where $\hat{\mathbf{x}}_k^-$ and \mathbf{P}_k^- are the system *a priori* (i.e., before observing the measurement at time k) state and covariance estimates, and $\hat{\mathbf{x}}_k$, \mathbf{P}_k are *a posteriori* (i.e., after observing the measurement) estimates. The update step is performed as:

$$\begin{aligned}\mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ \hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-) \\ \mathbf{P}_k &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-\end{aligned}, \quad (2.38)$$

where \mathbf{K}_k is the Kalman gain of prediction correction at time instance k .

KF is optimal for a linear system with Gaussian measurement and process noise [1, 147]. In case of a non-linear system or a linear system with non-Gaussian noise the unscented Kalman filter or Particle filters generally perform better at the price of additional computational effort.

2.7.2 Rao-Blackwellized Particle Filter

Particle filters (PF) recursively compute the posterior distribution of the states $\mathbf{x}_{1:k}$ given the measurements $\mathbf{y}_{1:k}$. This is done using a set of particles $\mathbf{x}_k^{(i)}$ and importance weights $w_k^{(i)}$ for $i = 1, 2, \dots, N_p$:

$$p(\mathbf{x}_{1:k} | \mathbf{y}_{1:k}) \approx \sum_{i=1}^{N_p} w_k^{(i)} \delta(\mathbf{x}_{1:k} - \mathbf{x}_{1:k}^{(i)}), \quad (2.39)$$

where N_p is the number of particles and $\delta(\cdot)$ is the Dirac delta function. On each time step the particle weights are recomputed depending on how well they fit the sequence of past measurements.

The Rao-Blackwellized Particle Filter (RBPF) improves the performance of PF when the state model has a linear-Gaussian substructure [1]. RBPF divides the state vector \mathbf{x}_k into two parts: \mathbf{x}_k^{P} , which is estimated using the PF, and \mathbf{x}_k^{K} , estimated by KF. The joint probability density function (PDF) is then given using the Bayes rule as

$$p(\mathbf{x}_k^{\text{K}}, \mathbf{x}_{1:k}^{\text{P}} | \mathbf{y}_{1:k}) = p(\mathbf{x}_k^{\text{K}} | \mathbf{x}_{1:k}^{\text{P}}, \mathbf{y}_{1:k}) p(\mathbf{x}_{1:k}^{\text{P}} | \mathbf{y}_{1:k}). \quad (2.40)$$

If the term $p(\mathbf{x}_k^{\text{K}} | \mathbf{x}_{1:k}^{\text{P}}, \mathbf{y}_{1:k})$ is linear-Gaussian, it can be optimally estimated by KF. The second factor is estimated using PF. The RBPF is performed in several steps, which include PF and KF updates and particle resamples [105].

2.7.3 Filter Application to the Process Model

This work concentrates on the KF and rather uses RBPF for comparison, as a more sophisticated approach, which builds on the KF, but handles process non-linearity. This way, judging by the performance of both filters, the applicability of a more simple and lightweight KF, compared to RBPF, can be established.

Abstract object movement is described as a discrete Wiener process velocity model [26] with the state vector defined as $\mathbf{x}_k = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]^T$, where (x_k, y_k) denotes object position and (\dot{x}_k, \dot{y}_k) — the velocity in a two-dimensional Cartesian space. The transition and measurement matrices for model (2.36) are then defined as:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad (2.41)$$

where Δt is the time interval between consecutive estimates in seconds. The process and measurement noise variance is specified by matrices

$$\mathbf{Q} = q \begin{bmatrix} \frac{1}{3}\Delta t^3 & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & \frac{1}{3}\Delta t^3 & 0 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & 0 & \Delta t & 0 \\ 0 & \frac{1}{2}\Delta t^2 & 0 & \Delta t \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}, \quad (2.42)$$

where q and r are the power spectral densities of process and measurement noise, respectively. These parameters can be specified statically, depending on the estimated spatial and temporal errors of position values, or dynamically reassigned at runtime by, e.g., a fuzzy inference mechanism, as it is done in the fuzzy Kalman filter.

2.8 Conclusions

The range of conventional signal processing techniques far exceeds the scope of this chapter. The presented concepts and methods serve as a theoretical basis for the main body of the thesis. The discussed time and frequency domain features are used for signal analysis and pattern recognition in later chapters devoted to process state and object identification. The correlation and fuzzy logic based classification procedures are applied to classification tasks throughout the main body of the thesis. Although a great variety of other classification methods exists, the main topic of the thesis is not devoted primarily to the discussion of classification procedures, and thus only the considered two classification approaches are adopted. The presented localization method of SRP-PHAT is used for comparison with the proposed DOA estimation and distributed localization procedures. An approach to reducing the number of computations required for SRP-PHAT and SRC is also presented later in the thesis. Multilateration is considered as an alternative approach to the proposed localization procedure, applicable only in wired sensor systems. KF and RBPF are applied to the tasks of trajectory estimation and search region prediction in the proposed approach of fusing consecutive localization estimates with filter estimates. The Wiener process velocity model is applied for object movement modeling throughout the main body of the thesis.

Chapter 3

Single-Sensor Solutions

The chapter presents methods of stationary and non-stationary process analysis by single-sensor devices for the task of pattern recognition. The first part of the chapter is dedicated to the typical process flow of pattern recognition for stationary process state identification. The stages of signal analysis, feature set selection and classification are discussed on an example of industrial machinery monitoring application [14, 19]. The second part of the chapter discusses non-stationary process analysis for the application of moving vehicle identification [17]. A multistage signal analysis procedure of hierarchical decision-making is proposed, its embedded implementation [16] is reviewed and field test results [18] are presented.

3.1 Pattern Recognition for Stationary Processes

Stationary processes assume time-invariant statistical properties. From the point of view of signal analysis, a stationary process possesses time-invariant properties either in the time, or the frequency domain. Therefore, the typical pattern recognition scheme, presented in Figure 2.10, is commonly applied for recognition of process states. Regarding a non-stationary finite state process, where each of the process states can be considered stationary, state separation is performed by applying specific feature extraction methods that reveal the differences in signal properties between the process states. If the non-stationary finite state process possesses stationary transition probabilities (e.g., precisely defined state duration periods), it can be modeled as a Markov chain [55], and the probability of each state estimated by applying, e.g., a Hidden Markov Model (HMM). This section considers pattern recognition of stationary process states of a finite state non-stationary process, the transition probabilities of which are undefined.

3.1.1 Example Process Description

A process of industrial machinery operation is considered as an example for the discussed signal analysis and state classification procedures. The application of these methods to machinery operation state monitoring and fault detection are discussed in Section 6.1.

The considered process consists of several industrial machine operation states, observed by means of acoustic and vibration information. Figure 3.1 presents a principal diagram of sensor placement and signal acquisition. Acceleration sensors are mounted on different parts of the machine (a router in case of Figure 3.1) to monitor the utilization of different components, and the acoustic noise emitted by the machine is remotely acquired by acoustic sensors. The machine is put through a typical working cycle and the operation states are identified using multi-modal information.

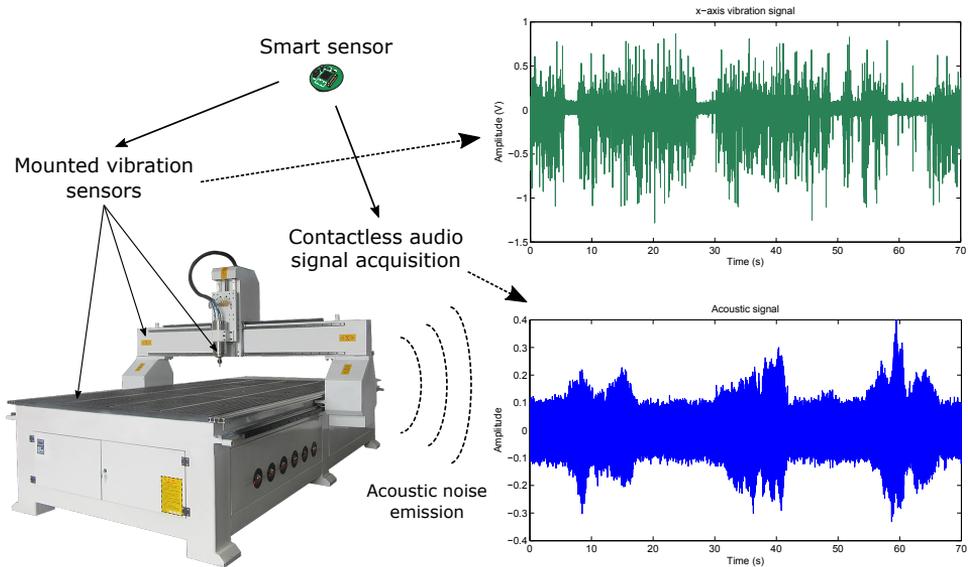


Figure 3.1: Multi-modal signal acquisition by smart sensors for industrial machinery operation state identification.

Experimental Setup

The experiment took place at the shop floor of a small size manufacturing facility during common operational conditions, i.e., full staff and standard machinery operation cycles. Two pieces of manufacturing equipment were chosen for testing: a manually operated circular saw bench, and a three degree of freedom Computer Numerical Control (CNC) router XYZ 6020.

During the experiment the manual saw bench was cutting planks of 16 mm medium-density fibreboard. The experimental signal includes six

cutting cycles. The saw passes the following operation states:

1. Saw is idle;
2. Motor enabled, saw blade spinning;
3. In addition to 2: dust collector enabled;
4. In addition to 3: cutting process.

The CNC router was routing on a sheet of 21 mm plywood. During the whole process of routing out several shapes from the plywood the spindle did not cease its rotation, thus the procedure is regarded as a continuous working cycle. The router passes the following operation states:

1. Router is idle;
2. Compressed air supply enabled;
3. In addition to 2: vacuum pump enabled;
4. In addition to 3: dust collector enabled;
5. In addition to 4: routing process.

Signal Acquisition

Acoustic signals were acquired by Shure SM58 microphones using a Roland Edirol UA-25EX audio signal processor at 44.1 kS/s sampling rate with the bit depth of 16 bits. For both considered machines the microphone was placed beside and directed towards the machine approximately 1.2 m above the floor. Thus no direct contact was made between the sensor and the machine.

Vibration measurements were made with an analog dual-axis accelerometer ADXL311 with a sensitivity of ± 2 g, 0 g bias of 1.5 V and sensitivity of 174 mV/g at the operating voltage of $V_{DD} = 3$ V. The signals were acquired at a sampling frequency of 1 kS/s using an Agilent U2354A data acquisition device (DAQ). For the saw experiment the accelerometer was firmly attached to the bench rip fence with both axes being parallel to the ground. The x -axis was pointed parallel to the saw cutting surface. During the routing experiment the sensor was attached to the spindle parallel to the Earth surface with x - and y -axis pointed along the first two degrees of freedom of the carriage. As in both cases the sensor is placed parallel to the ground, the gravitational component is not present in the readings and the 0 g bias is easily subtracted from the signal. For signal analysis one channel that displays greater responsiveness to the different process stages is chosen from the dual-axis signal.

3.1.2 Signal Analysis

Signal analysis is performed in the time and frequency domains to determine the process states detectable by sensors of different modalities. For the generation of spectrograms a rectangular windowing function is applied and zero overlap is used in order to simulate frame-by-frame operation of the sensors at runtime. Window length is set according to the signal frame length, which is equal to $2^{14} = 16384$ samples for acoustic sensors and $2^8 = 256$ samples for vibration sensors. This constitutes 0.372 s and 0.256 s frame duration, respectively. The resulting spectra consist of 8192 and 128 frequency components, accordingly.

Saw Signals

The saw acoustic signal is presented in Figure 3.2. All four operation states are distinguishable with the motor being started (state 2) approximately at the 15th second, dust collector enabled (state 3) at approximately the 25th second, and six cutting instances (state 4) from the 40th to the 160th second. The motor produces a low energy signal with a uniformly spread spectral density in the band of frequencies below approximately 10 kHz. The dust collector pattern is determined by the band below 1 kHz, and the sawing cycles are defined by the multiple high energy frequency components in the band of 1.5–10 kHz. Several instances of heavy background noise, e.g., at the 140th second, are also present in the signal. Each process state possesses a spectral distribution with low variance in frequency component energy, and is thus considered stationary.

The saw vibration signal is presented in Figure 3.3. As the sensor is mounted on the rip fence, it does not sense the vibration of the dust collector, which is not mounted on the bench itself. The bench motor is identified by the 50 Hz fundamental frequency and its harmonics. The vibration produced by the contact between the cut piece and the rip fence during sawing is more evident in the higher frequencies, starting from 250 Hz. Vibration energy seems to be well dampened by the rip fence, which results in an overall low energy signal. The interval between 160 and 180 seconds in Figure 3.3 signifies motor rotation decrease after turnoff. This slow transition between states 2 and 1 is not evident from acoustic information of Figure 3.2.

CNC Router Signals

The spectrogram of the CNC router is presented in Figure 3.4. The transitions between states occur at the following moments: state 1 to state 2 at the 40th second, state 2 to state 3 at the 50th second, state 3 to state 4 at approximately the 330th second, state 4 to state 5 at approximately the 520th second, state 4 back to state 3 at approximately the 910th second.

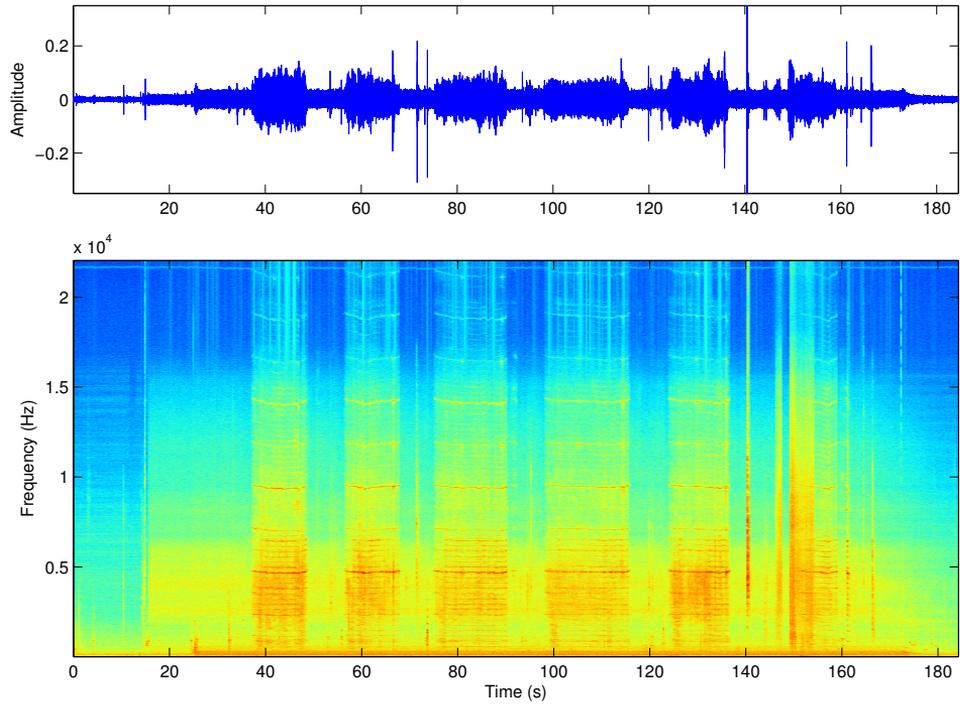


Figure 3.2: Saw acoustic signal (top) and its spectrogram (bottom).

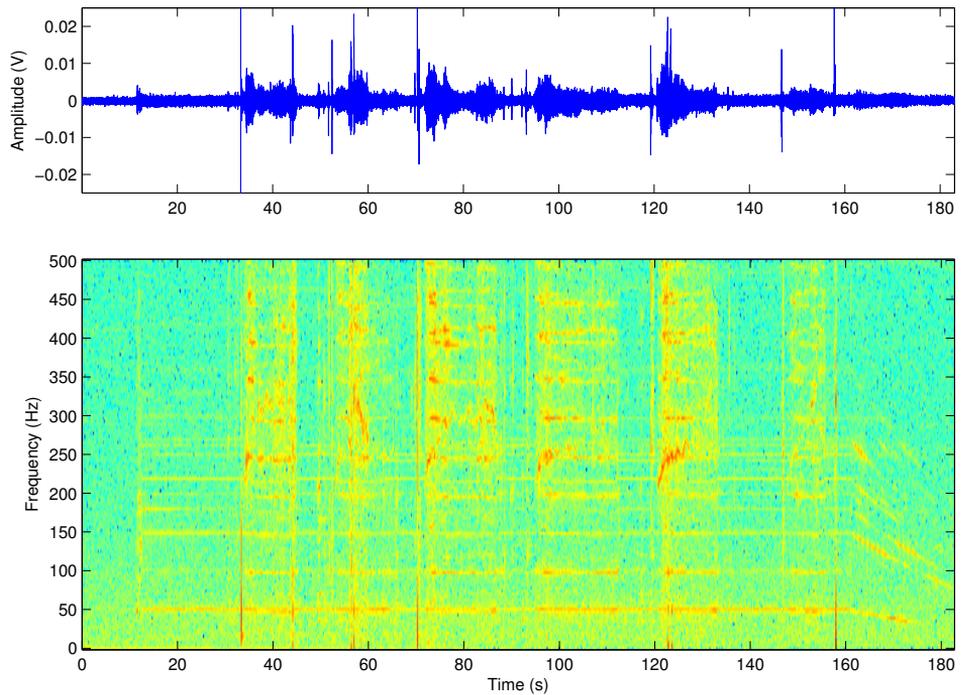


Figure 3.3: Saw vibration signal (top) and its spectrogram (bottom).

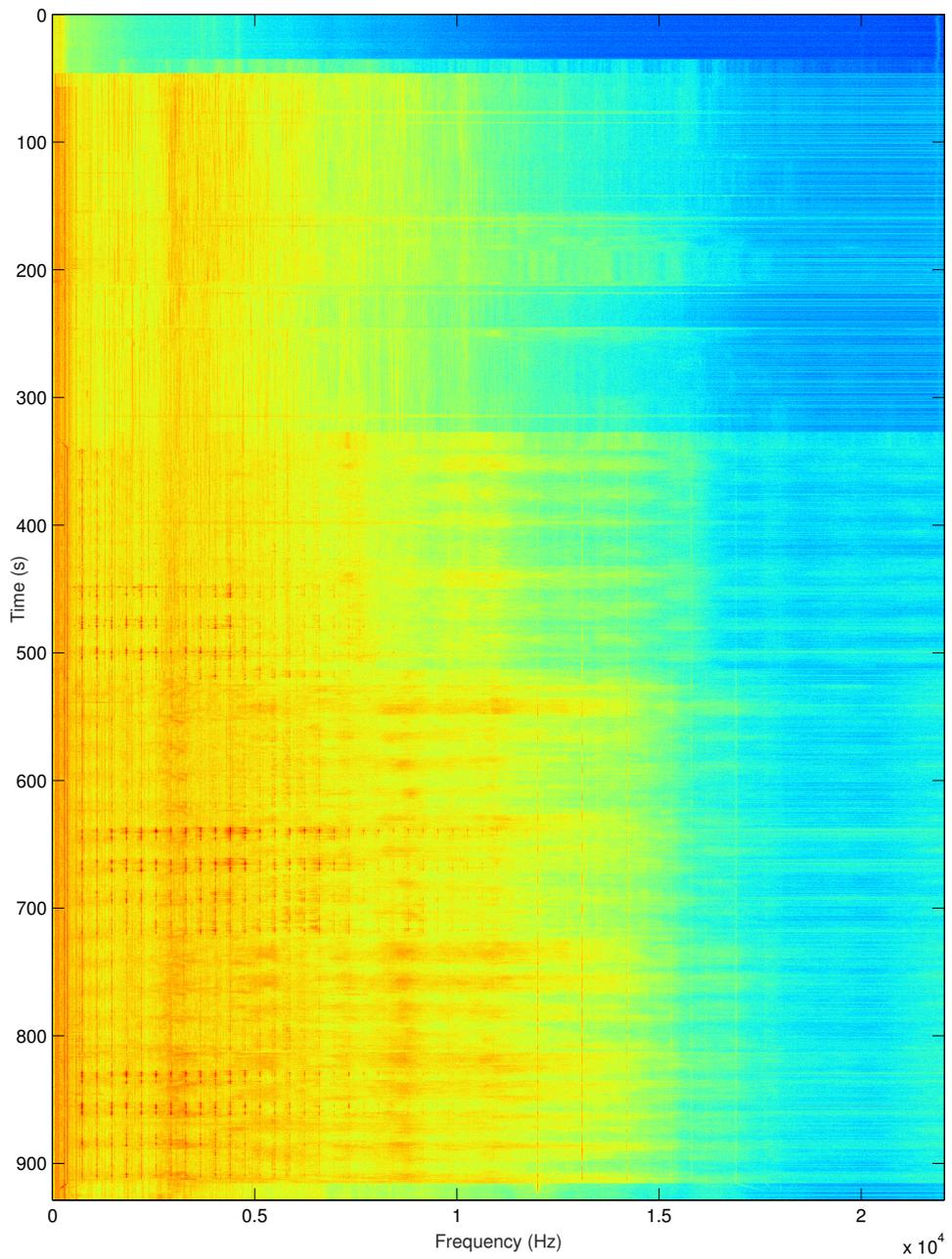


Figure 3.4: Router acoustic signal spectrogram.

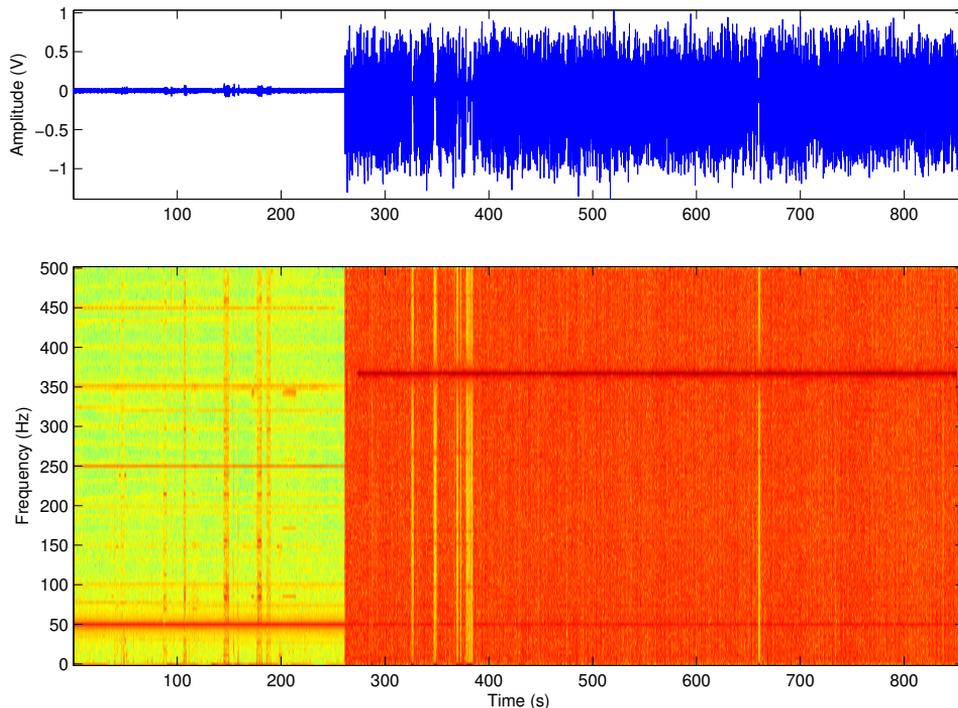


Figure 3.5: Router vibration signal (top) and its spectrogram (bottom).

State 3 (vacuum pump) produces high energy noise, which partially masks states 4 and 5. Therefore, pattern separation for intermittent states poses a more serious problem than in the case of saw acoustic signal. State 5 (routing) is emphasized by the band of frequencies higher than 8 kHz.

The router vibration signal is presented in Figure 3.5. Using the acceleration sensor mounted on the spindle it is possible to distinguish between state 5 (spindle rotation during cutting) and states 1–4 (spindle idle). During idle spindle states the defining frequency of faint noise is equal to 50 Hz (AC power current residual). Spindle rotation produces a highly uniform spectral density with a distinct peak at 370 Hz, which corresponds to measured spindle rotation of 22200 revolutions per minute (actual rotation setting was equal to 22000 rpm).

3.1.3 Feature Extraction

Feature extraction method choice for stationary process signals is relatively straightforward. Different types of features are computed from a test signal and matrices of form (2.23) are created. If the state labels per each feature vector (corresponding to a signal frame) are known, the fitness of the feature set is tested by training a classifier in a supervised manner and estimating the quality of dataset cover, the extent of feature sub-space separation (see

Section 2.5.2) and the impact that each feature has on the final decision. If the reference state labels are not known, the feature sub-spaces are separated in an unsupervised manner (e.g., by clustering), and the extent of separation and individual feature impact are measured by applying statistical tools to the resulting clusters [2].

In this example three types of features are extracted and three test datasets are generated per each test signal. These types of features are:

1. Band energies (2.16) with manually specified frequency bands;
2. Instantaneous features — various features from the list (2.3)–(2.6), (2.10)–(2.20);
3. Mel-Frequency Cepstral Coefficients (MFCC), discussed in Section 2.3.3.

Band energies represent the distinctive bands, by which the process states are separable. They are manually chosen during signal analysis stages and presented in Table 3.1. The various features, referred to as instantaneous (i.e., each calculated from a single signal frame), are chosen from the time and frequency domain feature lists of Chapter 2. For all test signals the following features are chosen: signal frame autocorrelation (2.5), signal RMS energy (2.6), spectral centroid (2.10), spectral skewness (2.11), spectral kurtosis (2.12), spectral decrease (2.13), spectral roll-off (2.14) with $\eta = 0.93$, spectrum spread (2.15), and both features of spectral slope (2.17), (2.18). Instantaneous features describe the overall shape and energy distribution in the signal, rather than energy of specific frequency bands. MFCC features consist of 24 mel energies, computed by applying the filterbank presented in Figure 2.7, and 12 cepstral coefficients.

Table 3.1: Frequency bands in Hz chosen for band energy features

Band number	Saw acoustic	Saw vibration	Router acoustic	Router vibration
1	50–450	40–60	70–650	10–490
2	1000–7000	95–105	725–745	40–60
3	4700–4800	230–270	2500–5600	365–375
4	9300–9600	280–310	5700–7800	
5	14000–14400	390–420	8000–16000	
6		440–470	17500–22050	

Visual representations of the saw acoustic signal and the router acoustic signal MFCC features are presented in Figure 3.6 and Figure 3.7, respectively. Figure 3.6 is the representation of the test signal presented in Figure 3.2 with the same state transition times. Figure 3.7, in turn, is the representation of the signal in Figure 3.4. The MFCC mel energy coding makes the process

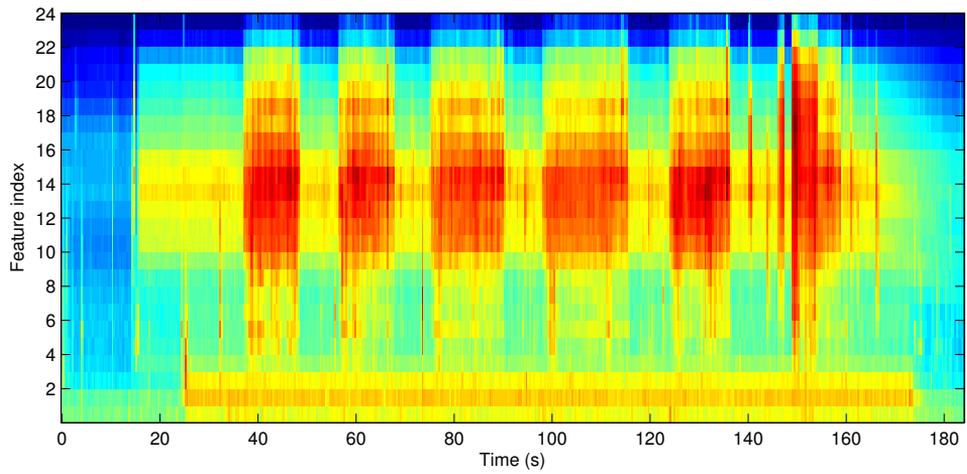


Figure 3.6: Saw acoustic signal MFCC features: 24 mel energies.

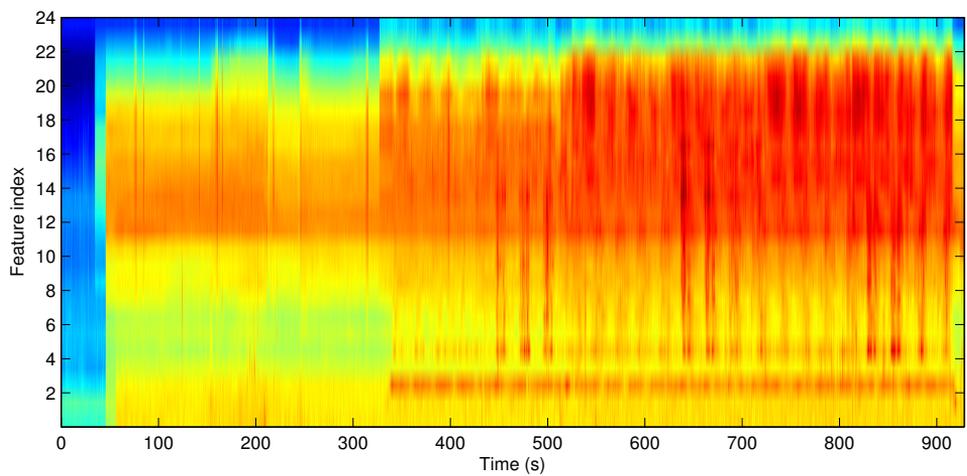


Figure 3.7: Router acoustic signal MFCC features: 24 mel energies.

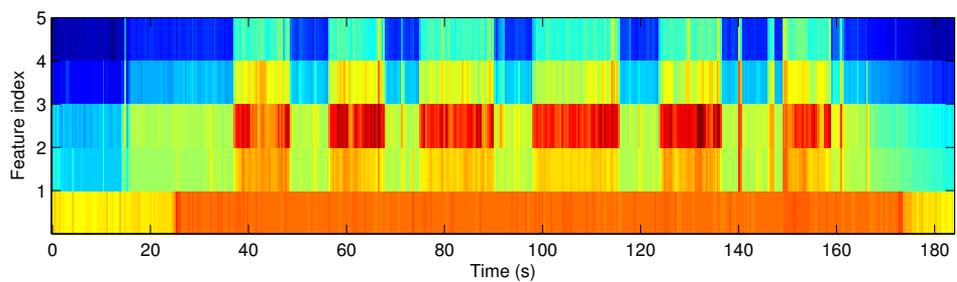


Figure 3.8: Saw acoustic signal 5 band energy features defined in Table 3.1.

states visually more distinguishable. On the other hand, MFCC produces a large number of redundant, heavily correlated features. Figure 3.8 presents the band energies dataset of the saw acoustic signal. All four states are clearly distinguishable by only five features. Further feature refinement can be performed by applying orthogonal transforms [147], e.g., PCA and ICA, or by performing feature selection.

3.1.4 Classification

Classification algorithm knowledge base generation is performed on training signals, which were acquired at the same experimental conditions alongside with the test signals. The training datasets generated from these signals are roughly equal or greater in length and contain information on all the classified process states. Two classification methods are chosen for the example at hand: the correlation metric (2.21) and the fuzzy logic based classifier, discussed in Section 2.5.3. These procedures serve as examples of the simplest and an advanced approach to frame-by-frame classification, respectively. The class labels are defined as integer values of the enumerated process states described in Section 3.1.1. These integer values are presented in Table 3.2.

Table 3.2: Process class labels with reference to the state lists of Section 3.1.1

Process state	Saw acoustic	Saw vibration	Router acoustic	Router vibration
1	1	1	1	1
2	2	2	2	1
3	3	2	3	1
4	4	3	4	1
5	-	-	5	2

Rule Base Generation and Class Label Estimation

The knowledge base of the correlation based classifier consists of C reference feature vectors \mathbf{r}_i , $i = 1, \dots, C$, where C is the number of classified states. Each reference vector \mathbf{r}_i is calculated by averaging several chosen feature vectors (20 in this case) from the training dataset, belonging to class i . The correlation is then calculated C times using (2.21), and the class label is defined as (2.22). The number of reference vectors per class label may be larger, and this tends to improve classification accuracy in more problematic cases, however, the correlation calculation then must be performed an according number of times, which is not always possible on embedded hardware, especially if the feature vectors are significantly long.

The fuzzy classifier is trained using the whole training dataset of the form (2.23) by applying the supervised learning procedure discussed in Section 2.5.3. For this task only the Minimal Rule Classifier (MRC) is generated [125]. The class label is calculated using (2.27), (2.28), applying product for the intersection operation.

Classification Results

The classification results are presented in Figures (3.9)–(3.12) by means of confusion matrices (with correctly classified frame quantities situated on the main diagonal) and series of estimated labels (denoted by blue crosses) plotted over reference labels (red circles). These four figures are chosen to illustrate the main conclusions on process state classification for this example.

- Misclassifications mainly occur between neighboring classes. This is expected, because for this example the neighbouring states have common traits.
- Classification burst errors happen during state transitions. As state transitions do not occur instantly, the exact point of class change cannot be strictly defined. Figure 3.10 portrays this situation best.
- Short-term abnormalities are naturally occurring in physical processes, and thus continuous classification will produce errors due to outliers. An example of this is presented in Figure 3.4, 325–380 s interval, as incorrectly classified in Figure 3.11, frame 800–1400 interval.

The correlation based classification results for all combinations of test signals and applied feature extraction methods are presented in Table 3.3. Correlation proves to be applicable to the more trivial case of router vibration signal classification, where two distinct states have significantly different properties. Worst performance is observed for the saw vibration signal classification, where state separation is complicated by a low signal to noise ratio (SNR) and significant state transition periods. Generally correlation performs better in case of long feature vectors, e.g., MFCC (36 compared to less than 10 of band energies). This is not unusual, as cross-correlation benefits from large information quantities and produces sharper correlation values over larger cross-correlated entities (longer feature vectors).

The fuzzy logic based classification method provides higher classification accuracy for all test signals, as shown in Table 3.4. By applying an advanced feature sub-space confinement and separation technique, it is able to manage variation in the features, which increases classification robustness. Therefore, model-based classifiers are more applicable to both stationary and non-stationary process classification.

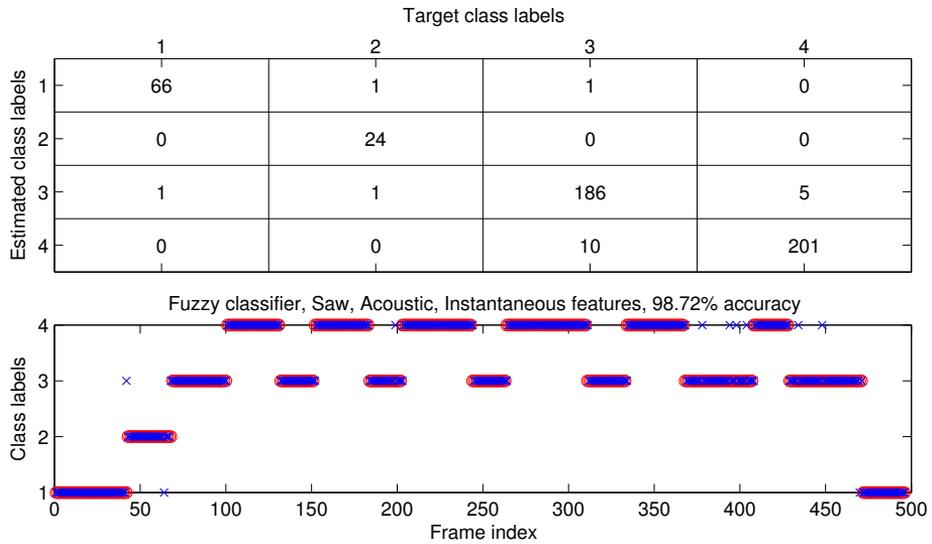


Figure 3.9: Classification results of the saw acoustic test signal.

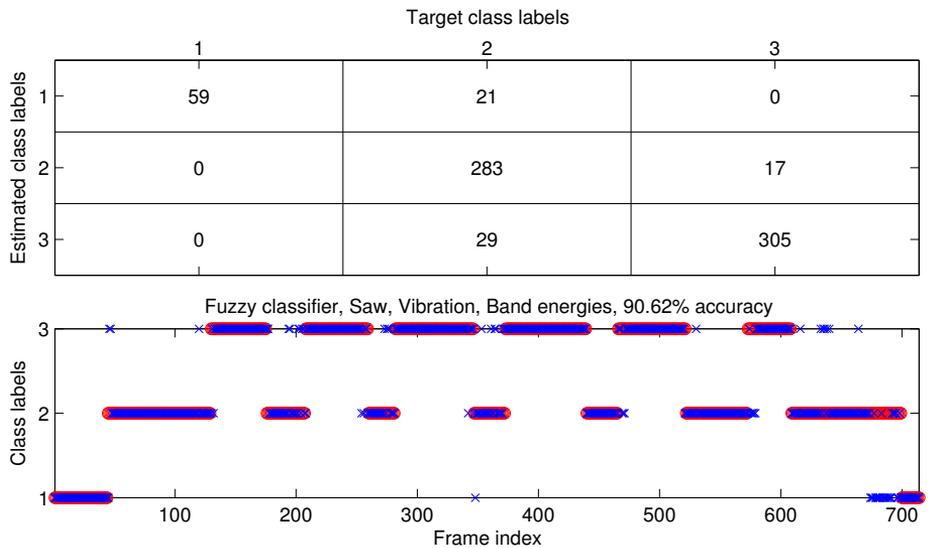


Figure 3.10: Classification results of the saw vibration test signal.

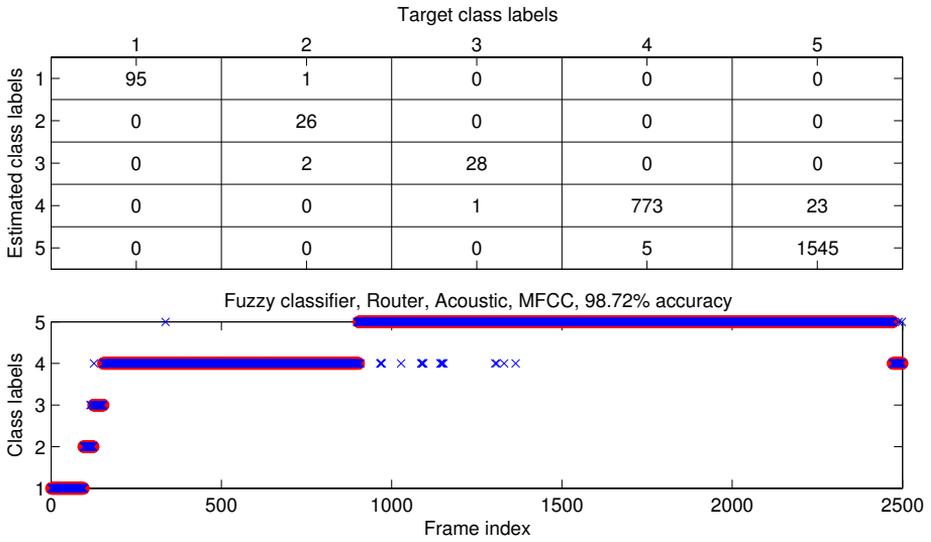


Figure 3.11: Classification results of the router acoustic test signal.

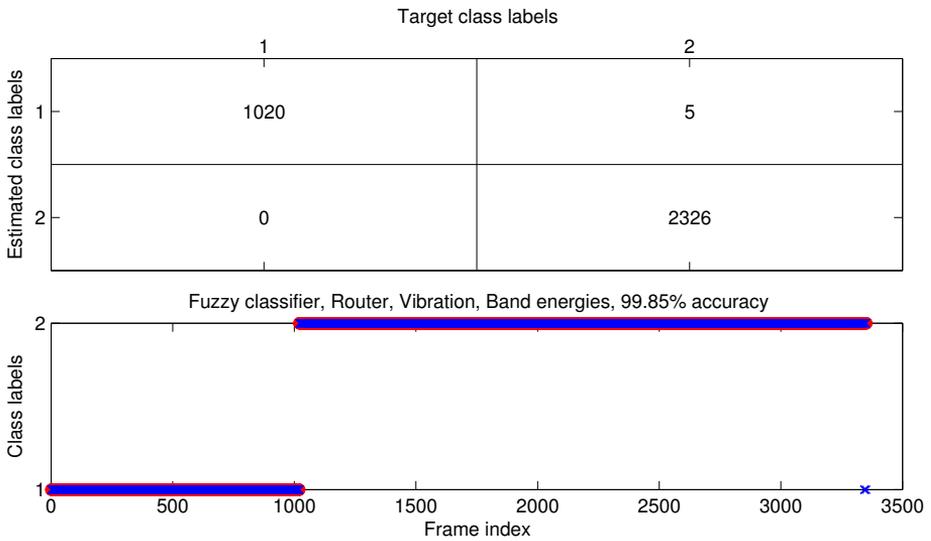


Figure 3.12: Classification results of the router vibration test signal.

Table 3.3: Correlation based classification results in percent

Object	Signal	Band energies	Instantaneous	MFCC
Saw	Acoustic	81.25	74.19	87.90
Saw	Vibration	54.06	68.91	65.69
Router	Acoustic	91.95	53.82	95.72
Router	Vibration	98.39	96.44	99.85

Table 3.4: Fuzzy classifier based classification results in percent

Object	Signal	Band energies	Instantaneous	MFCC
Saw	Acoustic	94.27	98.72	96.49
Saw	Vibration	90.62	83.75	81.65
Router	Acoustic	98.67	97.83	98.72
Router	Vibration	99.85	99.28	99.49

3.1.5 Feature Selection

Feature selection is used to reveal the best fit features that determine the classified states. It can be applied to the manually chosen feature sets in order to reduce the number of used features. Here feature selection is applied to automatic frequency band selection, as opposed to manual selection by frequency domain analysis.

The number of frequency bands equal to 8192 for acoustic signals is reduced to 1024 by averaging every 8 successive frequency components into one. The same procedure is performed over every 4 successive components of the vibration signals to reduce the number of bands from 128 to 32. This is done to soften the energy distribution between successive bands and to reduce feature selection computational complexity. Three feature selection algorithms are used: Differential Evolution Feature Selection (DEFS) [79], Least-Squares Feature Selection (LSFS) [74], and Multi Cluster Feature Selection (MCFS) [35]. An example of feature evaluation performed by LSFS is presented in Figure 3.13. The elevated squared-loss mutual information values indicate the minimized least-squares mutual information between the features of the corresponding band index to all other bands. Comparing the results of LSFS with spectral distributions of Figure 3.2 and Figure 3.4 shows that the bands with high squared-loss mutual information are in fact distinctive properties of process states.

The fitness of the features chosen by the three feature selection methods is tested by classifying the feature datasets with the fuzzy logic based classifier (after training it on the selected feature training datasets). The results of test signal classification are presented in Table 3.5. The results show that feature extraction based on automatically selected features is a feasible solution with

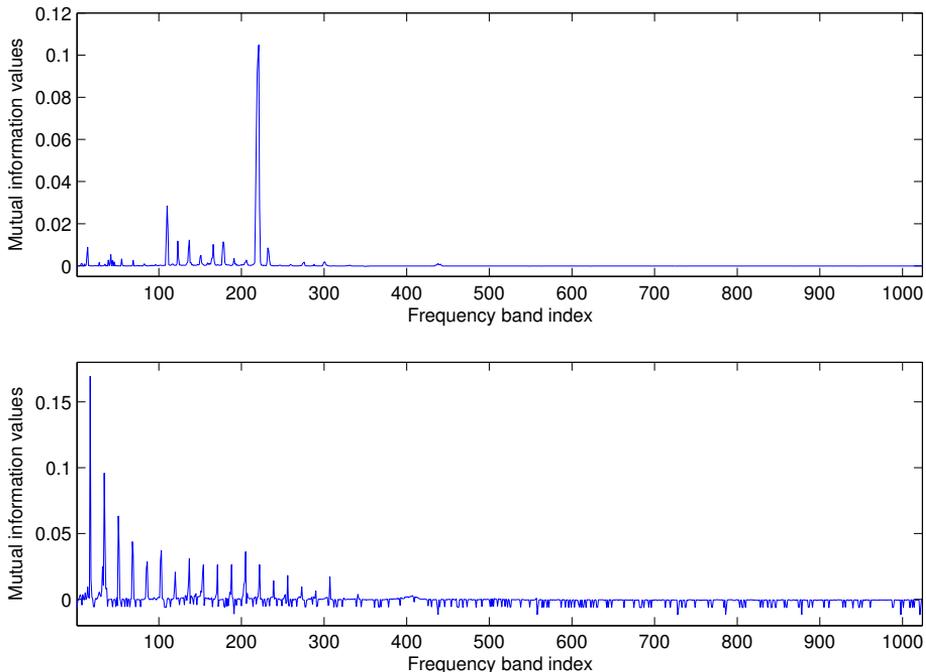


Figure 3.13: Evaluation of features by LSFS. Squared-loss mutual information of the saw (top) and router (bottom) acoustic signal frequency bands.

the classification quality being comparable to that of the manually specified feature sets. However, in low SNR conditions of the saw vibration signal all feature selection methods fail to provide adequate features, which results in poor classification accuracy.

Table 3.5: Fuzzy classification results for selected features in percent

Object	Signal	DEFS	LSFS	MCFS
Saw	Acoustic	94.96	95.16	95.56
Saw	Vibration	70.31	58.12	65.27
Router	Acoustic	99.08	98.80	98.68
Router	Vibration	99.88	99.76	77.73

3.1.6 Presented Approach Discussion

The presented approach to stationary process state identification is the simplest form of situation assessment for smart sensors. Its straightforward operation flow of the pattern recognition steps presented in Figure 2.10 and the ability to process only one signal frame at a time make it well implementable on low power hardware. As no sequential information of process dynamics, e.g., (2.19), (2.20), or MFCC delta coefficients, is analyzed, the amount of

required memory and signal buffer size are reduced. For stationary processes the dynamical properties are only relevant during state transitions.

To achieve full cover of the observed process, all possible process states must be handled during initial testing. For the example of industrial machinery monitoring this would mean covering all possible working parameters (saw blade and spindle rpm rates, treated material types, etc.) for every sensor type and placement. Abnormal behavior can be then identified by thresholding the classification label confidence level (e.g., correlation values or rule activation degrees for fuzzy classifiers) and during data fusion of information incoming from several sensors.

The main problem of the presented approach is situated with reduced noise tolerance. The problem is more relevant in case of acoustic signal analysis because acoustic sensors are susceptible to noise incoming from all surrounding potential acoustic sources. In an unconfined environment the properties of arising noise cannot be predicted and sufficiently estimated by single-sensor systems. The problem can be tackled either by applying multi-sensor solutions, or through multi-modal data fusion.

3.2 Pattern Recognition for Non-stationary Processes

The majority of real life dynamic processes are non-stationary with varying state transition probabilities. For bounded or finite conditions non-stationary processes can be divided into short-term semi-stationary components that succumb to statistical analysis and thus can be described by probabilistic models. Another approach lies in the development of application specific procedures that describe certain process properties under a number of assumptions. This section considers an application specific pattern recognition procedure for moving vehicle identification.

3.2.1 Problem Statement and Process Description

For this example of non-stationary process pattern recognition a single-sensor procedure of passing vehicle type identification is considered. The procedure is based on acoustic noise analysis, which provides the possibility to distinguish between well separable classes of motor vehicles, such as passenger cars and large trucks. The acoustic noise patterns of moving vehicles consist of multiple components, including noise produced by the engine, exhaust system and tires [37]. The harmonic nature of the engine noise is, however, seldom present in the civil vehicle sound pattern due to the fact that engine sounds are well dampened in modern cars. This fact, complemented by the Doppler Effect [29], renders the spectral analysis based on fundamental fre-

quency detection, e.g., [163], ineffective. Instead, parameters of the overall spectral shape and energy distribution describing the vehicle noise patterns is adopted. Signal amplitude envelope is also adopted as a time domain feature.

Experimental Setup

For the experiments a single-sensor device is placed 2–15 meters from the vehicle path and directed perpendicular to it. Two experiments are performed, which are from this point denoted as Experiment 1 and Experiment 2.

For Experiment 1 a microphone was placed at an empty parking lot and two cars (Mercedes S320 and Mazda MX-5) were in turn passing the microphone stand at a speed of 35–45 km/h at the passing point, starting to accelerate from a distance of approximately 40 meters. Each car passed the microphone three times: the Mercedes first three times and the Mazda three times afterwards. The sounds were acquired during summer time in mild weather conditions, thus ambient noise levels were relatively low. The signal spectrogram is presented in Figure 3.14. Six passing car acoustic patterns are clearly visible.

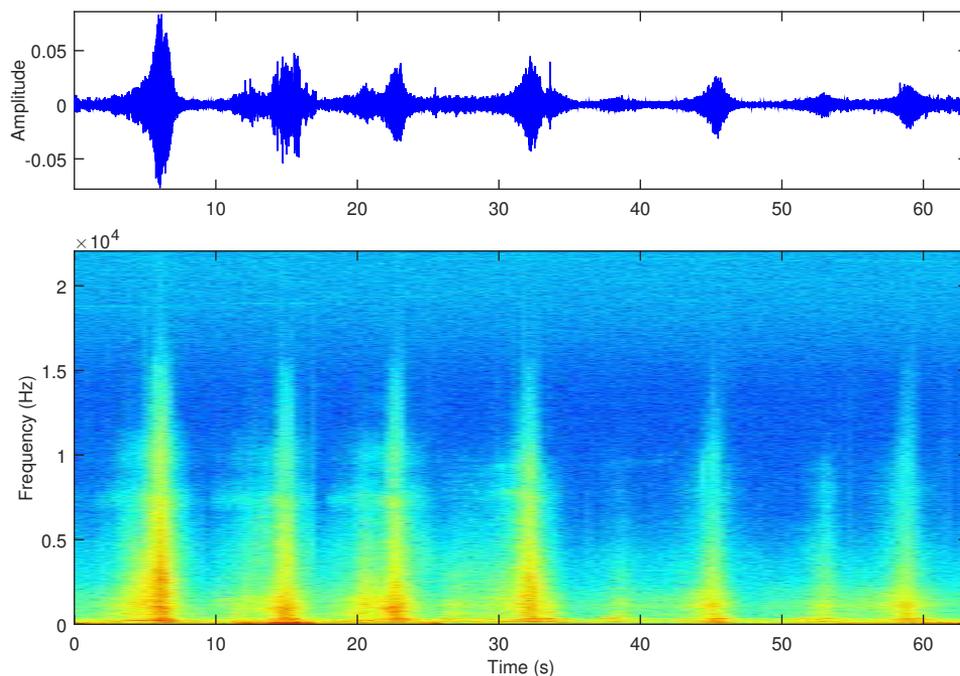


Figure 3.14: Acoustic signal of six passing passenger car instances of Experiment 1 (top) and its spectrogram (bottom).

Experiment 2 was conducted at a lively three-lane highway during dense traffic in late fall under heavy wind and light rain conditions. Consequently

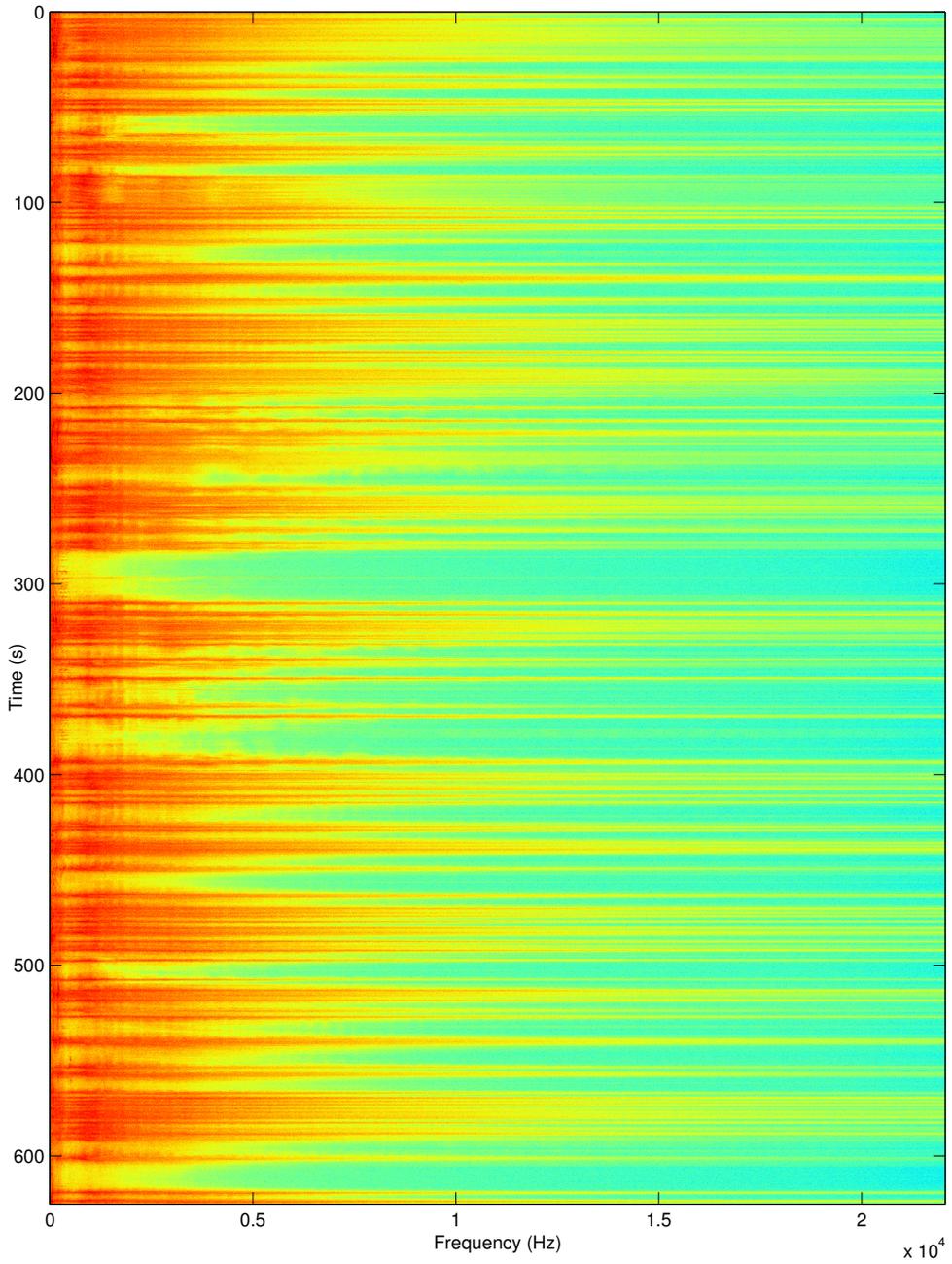


Figure 3.15: Experiment 2 signal spectrogram of numerous passing vehicles.

the noise levels in this signal are significantly high, as can be seen on the signal spectrogram, presented in Figure 3.15. Ambient noise from wind and rain influences the whole frequency band, unlike in Figure 3.14. Dense traffic results in vehicle passing instances being much less visually separable. During the experiment vehicles of different types, including passenger cars, minibuses, trucks, long buses, etc., have passed the sensor at different distances at a speed of 50–60 km/h.

Figure 3.14 and Figure 3.15 serve as good examples of truly non-stationary processes, which cannot be split into short-term stationary components. Therefore, the application of a more complex procedure compared to a typical pattern recognition scheme of Section 3.1 is in order.

Signal Acquisition

The acoustic signal of Experiment 1 was acquired using a Shure SM58 microphone and a Roland Edirol UA-25EX audio signal processor at 44.1 kHz sampling rate with the bit depth of 16 bits. The signal of Experiment 2 was acquired using a condenser microphone Sennheiser KE 4-211-2 and an embedded computing device Gumstix Overo Water. The signal was also sampled at 44.1 kHz mono channel with the bit depth of 16 bits. The use of two different microphones with different frequency responses and dynamic ranges also adds to the difference in SNR and signal quality induced by the difference in weather conditions between the two experiments.

For signal analysis the signal frame length is set to $2^{14} = 16384$ samples, which constitutes 0.372 s frame duration. The resulting spectrum thus consists of 8192 frequency components.

3.2.2 The Multistage Procedure of Vehicle Identification

The proposed multistage algorithm, presented in Figure 3.16, consists of two independent stages. The hierarchical decision-making scheme (on the left) distinguishes between relatively loud sounds and mild background noise, then it separates vehicle-produced acoustic patterns from heavy background noise and finally estimates the vehicle type from a set of predefined types. This part of the algorithm operates in a frame-by-frame manner, computing a single class label per signal frame. The Attack Sustain Release (ASR) envelope estimation procedure, on the other hand, runs parallel to the decision-making procedure and complements the past frame classification results with reassurance of positive vehicle passing event detection. The hierarchical structure of the algorithm reflects the superiority of vehicle detection over correct vehicle type classification. In other words, distinction between vehicle-produced sound and other types of noise is more important than correct vehicle type estimation.

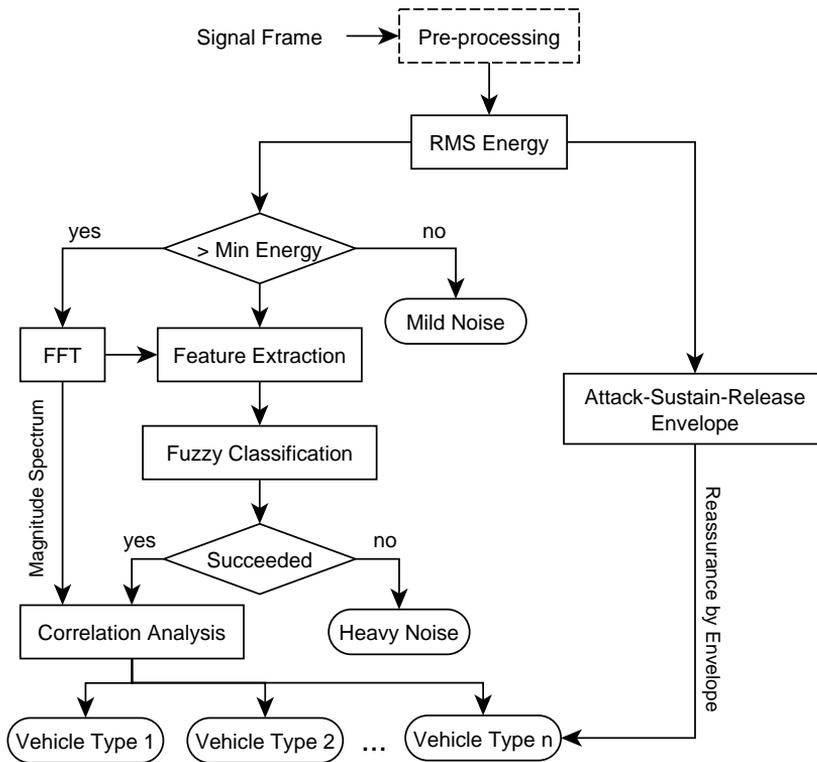


Figure 3.16: Block diagram of the proposed hierarchical algorithm for vehicle detection and classification.

Signal Pre-processing

The proposed procedure in Figure 3.16 starts by conditioning (filtering, etc.) the incoming acoustic signal frame, which is optional and depends on the specific environment and identified vehicle type properties. Static filtering is effective only in cases where the frequency band containing the signal is known. In case of Experiment 2, for example, band pass filtering with the pass band of approximately 500–15000 Hz would be appropriate (see Figure 3.15), however, larger and heavier vehicles have distinguishing spectral properties below 500 Hz, therefore, their acoustic patterns may be corrupted. For both experiments a low pass filter with the pass band of 15 kHz is applied.

RMS Energy and Thresholding

After signal conditioning the frame RMS energy (2.6) is calculated and compared to the lower energy threshold. If the threshold is not exceeded, the hierarchical procedure terminates (ASR calculation proceeds) and the frame is marked with a label of mild noise. The estimation of the lower energy threshold occurs during algorithm parameter estimation by means of test

signal analysis. The initial threshold is chosen as the minimal value of RMS energy of all the frames that correspond to vehicle passing instances.

Feature Extraction

If the signal frame x_t passes energy thresholding, the absolute magnitude spectrum $|X_t|$ is calculated by applying the FFT, and a set of frequency domain features is extracted. What follows is the list of the used features.

1. Band energies (2.16), which is a vector consisting of n_b frequency bands. The bands are chosen similar to the distribution of the mel-scale (2.9).
2. Spectral centroid (2.10).
3. Spectral roll-off (2.14) with $\eta = 0.9$.
4. Spectral slope with both parameters $a_t(|X_t|)$, defined by (2.17), and $b_t(|X_t|)$, defined by (2.18).

An example of spectral slope for a magnitude spectrum of length $K = 8192$ is presented in Figure 3.17. The overall decline of spectral energy towards higher frequencies defines the parameters of the straight line and not the precise energy distribution in bands.

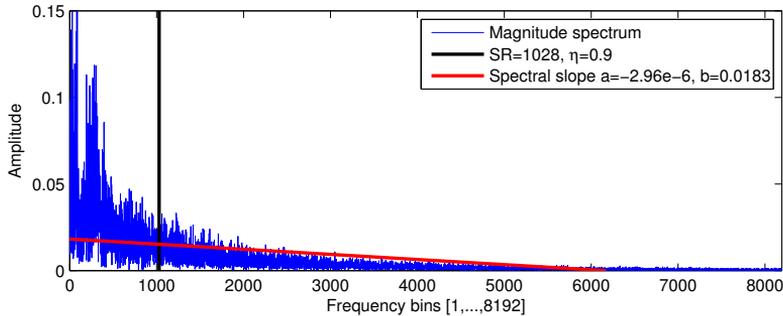


Figure 3.17: Spectral roll-off and spectral slope of an acoustic signal frame.

The features are concatenated into a feature vector

$$\chi_t = [\text{BE}_t(|X_t|, 1), \dots, \text{BE}_t(|X_t|, n_b), \text{SC}_t(|X_t|), \text{SR}_t(|X_t|, \eta), a_t(|X_t|), b_t(|X_t|)], \quad (3.1)$$

which is analyzed during the later stages of classification. The presented set of features is unbalanced (feature variance may differ significantly), and band energy features may become correlated for specific vehicle types. The method of fuzzy classification, however, overcomes this problem by identifying the most responsive and definitive features through the training procedure, discussed in Section 2.5.3.

Vehicle Pattern Detection by Fuzzy Classification

The sound patterns of passing vehicles are not consistent due to there being several noise components among these patterns. The influence of the components, such as engine, exhaust and tire produced noise, depends on the distance, trajectory and type of a vehicle. For example, the pattern of an approaching vehicle may be composed mainly of tire noise, while passing the sensor the engine noise can be dominant, and while the vehicle is moving away from the measurement point the exhaust noise may intensify. Coupled with the Doppler Effect, which induces frequency shifts on the spectral components of fast moving vehicles [29], these circumstances introduce significant variance to the spectral features of the signal. An example of three clusters residing in a three-dimensional feature space (of three band energies) is presented in Figure 3.18. Each point in the figure represents a feature vector of an Experiment 2 test signal frame, corresponding to the following labels: strong background noise (blue dots), light vehicles (red stars), and heavy vehicles (green triangles). The quality of separation of these overlapping clusters directly influences the quality of vehicle detection.

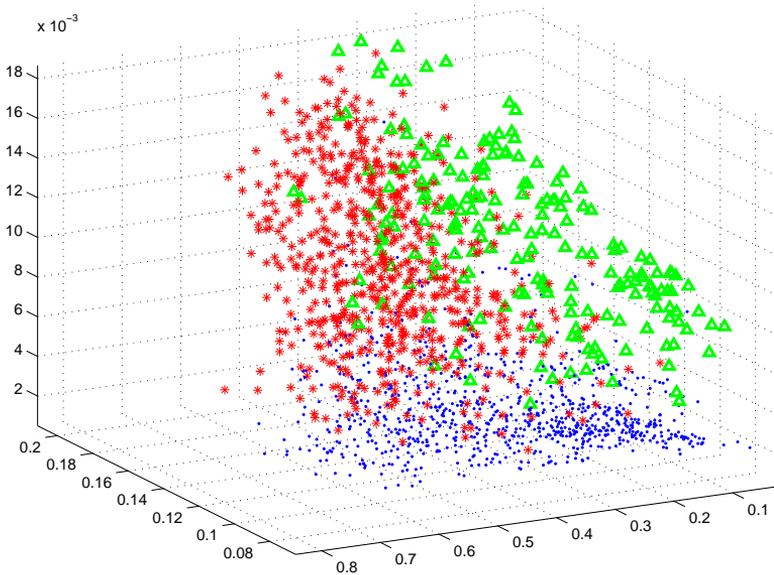


Figure 3.18: Three clusters in a three-dimensional feature space. The features represent energy ratios and thus are unitless.

Feature sub-space modeling and separation is performed by the fuzzy logic based classifier. An example of modeling the clusters presented in Figure 3.18 by triangular Membership Functions (MF) is presented in Figure 3.19. If the clusters do not separate naturally in the feature space (which is often the case), the extracted rules are expected to have a high degree of overlap (Figure 3.19 top). By excluding a portion of overlapping feature samples

(marked by circles in Figure 3.19) from the feature sub-spaces the overlap can be significantly reduced (Figure 3.19 bottom). Moreover, usually this comes at no or only minor loss of classification accuracy [124].

In the proposed hierarchical procedure of Figure 3.16 the fuzzy logic based classifier may be applied in two different manners. First, a general cluster corresponding to all vehicle types in question may be estimated and the resulting classifier used for pure detection purposes just to distinguish all vehicle produced noise from ambient noise. Final classification is then performed by correlation analysis. Alternatively, a separate cluster for each vehicle class is built and the classifier is applied for specific vehicle type estimation. In this case both the fuzzy classifier and the correlation analysis produce separate class estimates, which may reinforce each other. In Section 3.2.3 both methods are used during the multistage procedure testing. For classifier training the entire procedure discussed in Section 2.5.3 is applied, which results in a minimal rule, minimal feature classifier with eliminated outliers.

Correlation Analysis

The final stage of vehicle identification is the correlation analysis (2.21) between the incoming magnitude spectrum vector $|X_t|$ of length K and the reference spectrum vectors $|X_r|_i$, each corresponding to a single vehicle type class $i = 1, \dots, C$, where C is the number of classes:

$$\rho_i = \frac{\left[K \sum_{k=1}^K |X_t[k]| \cdot |X_r[k]|_i - \sum_{k=1}^K |X_t[k]| \sum_{k=1}^K |X_r[k]|_i \right]}{\left[\sqrt{K \sum_{k=1}^K |X_t[k]|^2 - \left(\sum_{k=1}^K |X_t[k]| \right)^2} \cdot \sqrt{K \sum_{k=1}^K (|X_r[k]|_i)^2 - \left(\sum_{k=1}^K |X_r[k]|_i \right)^2} \right]}. \quad (3.2)$$

For more rigorous classification several reference vectors per class may also be used. Correlation coefficients are simple and effective metrics for similarity estimation, however, this method is very susceptible to noise. A spectrum of heavy background noise may correlate to any of the reference spectra enough to produce incorrect classification results. Application of the fuzzy classifier in the previous stage of the algorithm reduces the influence of background noise.

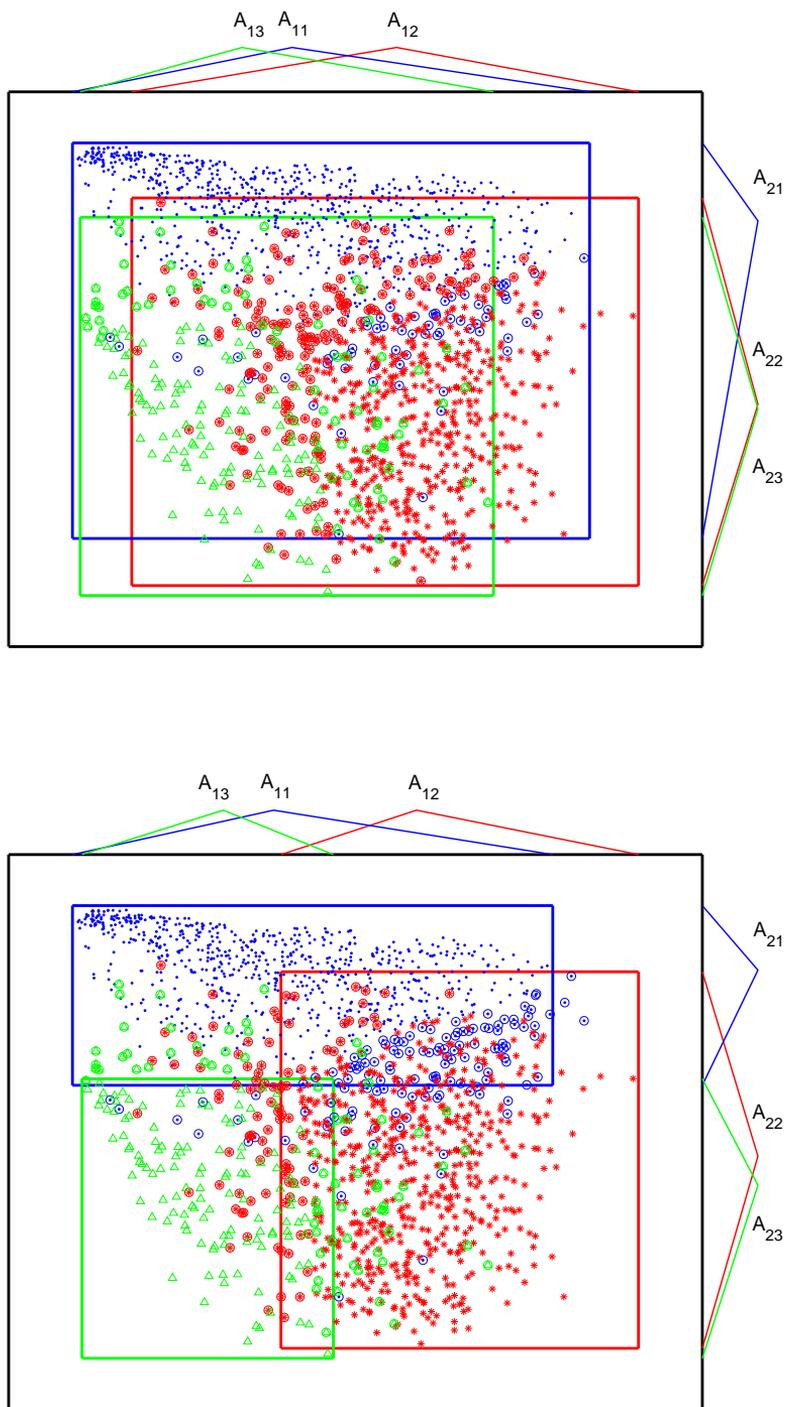


Figure 3.19: Initial (top) and refined (bottom) representation of three signal feature clusters by triangular Membership Functions.

Attack Sustain Release Envelope

The process of a vehicle passing the measurement point consists of three stages: approach (signal intensity increases), passing (signal intensity remains stable), and retreat (signal intensity decreases). This dynamic pattern is detected by estimating the Attack Sustain Release (ASR) envelope. It is conducted by analyzing the RMS energy (2.6) of successive frames.

The amount of deviation of RMS energy of the present frame $\text{RMS}_t(x_t)$ at discrete time instance t is estimated by the difference between it and the mean value of N_E previous RMS energy readings. RMS energy deviation is coded into three states by the following principle:

$$s_t = \begin{cases} 1, & \text{RMS}_t(x_t) > (1 + \delta) \cdot \overline{\text{RMS}}_t; \\ -1, & \text{RMS}_t(x_t) < (1 - \delta) \cdot \overline{\text{RMS}}_t; \\ 0, & \text{otherwise,} \end{cases} \quad (3.3)$$

where 1 denotes energy increase, 0 denotes stable energy levels and -1 denotes energy decrease. The parameter $\delta \in [0, 1]$ is the lower threshold of energy deviation. The mean $\overline{\text{RMS}}_t$ of N_E previous energy levels is calculated as

$$\overline{\text{RMS}}_t = \frac{1}{N_E} \sum_{t-N_E-1 \leq i \leq t-1} \text{RMS}_i(x_i). \quad (3.4)$$

Therefore the transitions $1 \rightarrow 0 \rightarrow -1$ and $1 \rightarrow -1$ are suspected for a car passing event and the quantities of -1 , 0 , and 1 coded frames denote the lengths of attack, sustain, and release components, respectively.

The detection of the signal energy ASR dynamic complements the past identification results. If the ASR pattern is detected, a notification is generated and presented along with the final class estimate. The class labels generated during the detected ASR period are inspected for the most frequent one (mode in the statistical sense), which is presented in the notification. This reduces inconsistencies in the series of class estimates, e.g., when the vehicle type cannot be strictly classified. Furthermore, if a vehicle is better identifiable during the moment of passing the measurement point (when some specific acoustic component is prevalent in the pattern and the Doppler Effect is minimal), class labels of the sustain signal portion may be used for final decision generation. Additional restrictions may also be applied to the ASR envelope detection. If the potential velocity of the moving object is *a priori* known, the lower and upper bounds for the attack, sustain or release components may be specified, so the detection is invalid if these restrictions are not met. For example, if the vehicles are known to stop at the measurement point, the expected value of the sustain component has to be large in order to not confuse this stop with multiple vehicles. On the other hand, for

fast passing vehicles (e.g., on a highway) the ASR components are expected to be short.

3.2.3 Multistage Procedure Test Results

The performance of the algorithm is tested on signals acquired in Experiments 1 and 2, discussed in Section 3.2.1. The signals are manually analyzed prior to algorithm accuracy evaluation. This is done in order to estimate the number of classes that will be used in the algorithm and to assign reference class labels to every test frame. Each signal is divided into a training and test portion. In case of Experiment 1 the training is performed on a signal containing six car pass instances, and testing is performed on another signal containing six pass instances of the same cars. The signal acquired during Experiment 2 is split in half, where the first half is used for testing and the second half is used for training. The training portion is used for fuzzy classifier training and for choosing reference vectors for the correlation analysis procedure. For fuzzy classifier training the features are extracted from every frame of the training signal and concatenated into the training dataset of the form (2.23). For correlation analysis several magnitude spectrum vectors corresponding to different classes are manually chosen. The test portion of the signal is used for the estimation of detection and classification accuracy.

Experiment 1 Results

The signal feature vector (3.1) comprises eight features: four band energy features (four bands of 1–824, 824–2616, 2616–6514, 6514–15000 Hz), spectral centroid, spectral roll-off and spectral slope. For classification a total of 2 classes is used: 1 for Mercedes and 2 for Mazda. The reference spectral vectors used in correlation analysis are estimated by averaging several spectra of sounds produced by vehicles corresponding to the same class; one reference vector per class is applied.

The results of algorithm testing are presented in Figure 3.20. The figure consists of four subplots; from top to bottom these are:

1. Test signal with 6 instances of passing vehicles (blue); final estimated labels with values 0.05 corresponding to class 1 and 0.1 to class 2 (black);
2. RMS energy readings per frame (green); signal energy threshold exceeded or not (black line); energy peaks approximated by ASR envelope (violet stems);
3. Coded RMS energy dynamic of the ASR envelope;

4. Intervals of nonzero fuzzy membership to vehicle feature subspace (black line); coefficients of correlation with the reference spectral vectors (blue — class 1, green — class 2).

For Experiment 1 every vehicle is detected and successfully classified. The second and third subplots of Figure 3.20 show that each vehicle passing instance ASR envelope is correctly detected. Though it can be noticed that approximately on the 107th frame the ASR dynamic is falsely detected, the energy of the signal is below the threshold and the fuzzy classifier gives no classification decision, consequently the detection does not occur. Also approximately at the 145th frame the ASR dynamic is present but not detected, as for the known vehicle speed corridor the attack and release components of the envelope are set to be no less than 2 frames in duration for the dynamic to be detected. For this signal the fuzzy classifier is trained for detection purposes, i.e., distinguishing between all car types combined and ambient noise. The detection decisions of the fuzzy classifier concur with the ASR envelope.

The fuzzy logic based algorithm, trained to identify the general vehicle feature space, succeeds in doing so for the majority of signal frames thus allowing the correlation coefficient calculation procedure to analyze only the frames corresponding to vehicle pass time intervals. The fourth subplot of Figure 3.20 shows that the correlation coefficient values are unreliable during the periods between vehicle pass instances. During vehicle pass instances, however, they become more separate indicating a distinct leader.

Experiment 2 Results

Feature vectors comprise eight features, which are the same as for Experiment 1, except the bands for the band energy features are less spread: 220–818, 818–2592, 2592–6438, 6438–14780 Hz. For obtaining the reference spectral vectors the same technique as for Experiment 1 is used. Two broad vehicle classes are chosen: 1 for light vehicles (passenger cars) and 2 for heavy vehicles (trucks and buses).

The fuzzy classifier in this case is trained for both detection and identification purposes. It is also trained to model the ambient noise feature sub-space. The class labels for the fuzzy classifier are set as: 1 for ambient noise, 2 for light vehicles, and 3 for heavy vehicles. The final structure of the classifier rule base is presented in Figure 3.21. This rule base consists of 7 rules, which only use features 1, 2, and 8 (the first two energy bands and the second parameter of spectral slope). These features are automatically chosen during classifier training. Correctly classified feature samples are denoted by blue dots and the misclassified samples are denoted by red dots. The training dataset produced by the Experiment 2 signal has highly overlapping data, as

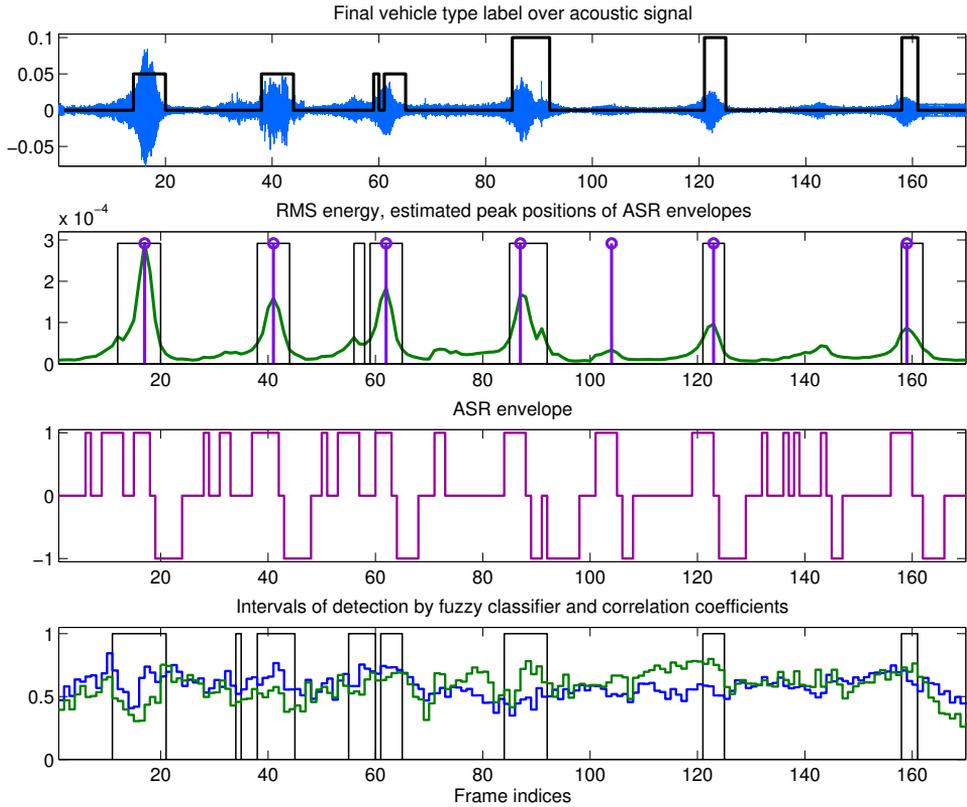


Figure 3.20: Multistage procedure test results for the signal of Experiment 1.

was shown in Figure 3.18 and Figure 3.19. Nevertheless, the fuzzy classifier succeeds in feature sub-space separation with 84.9% accuracy.

The results of signal analysis are presented in Figure 3.22, which consists of two subplots:

1. Test signal with instances of passing vehicles (grey); reference labels with values 1 corresponding to class 1 and 2 to class 2 (black);
2. Final estimated labels with values 1 corresponding to class 1 and 2 to class 2.

As the time intervals between car passes are very short and often nonexistent altogether, the reference class labels are plotted over the acoustic signal in first subplot for presentation clarity. The intermediate results are not presented due to low presentation clarity. For Experiment 2 the following results were obtained: out of 46 instances of class 1 vehicles 37 were successfully detected and classified, 5 were undetected and 4 were confused with class 2; for 11 instances of class 2 vehicles 9 were correctly classified, 1 was not detected

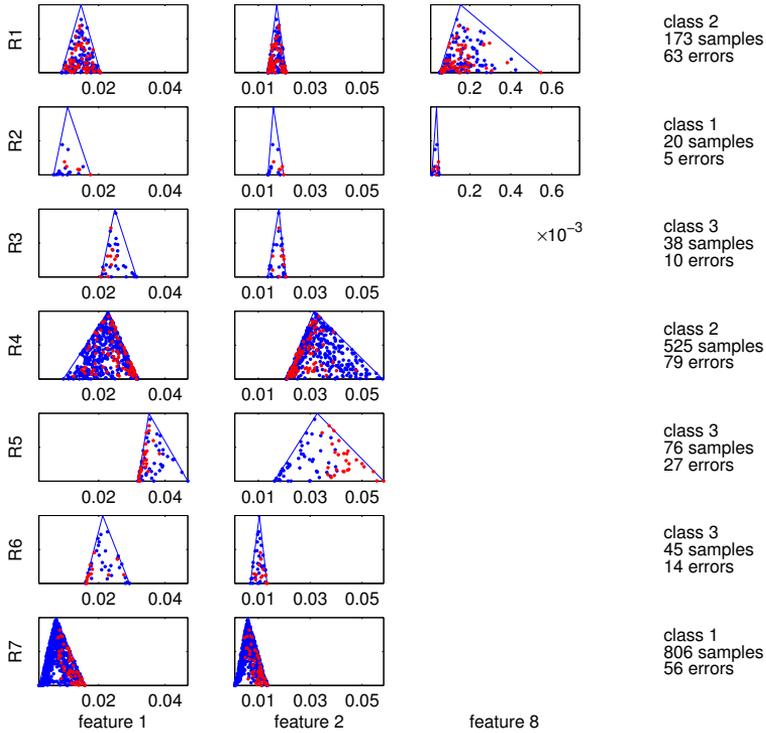


Figure 3.21: Fuzzy rule base for Experiment 2 data.

and 1 confused with class 1. Considering the harsh environmental conditions, the overall detection and classification accuracy is acceptable.

The main problems causing reduced classification accuracy are situated with the following circumstances.

- Severe corruption of the whole signal frequency band with high levels of ambient noise.
- Due to dense traffic the time interval between vehicle passes is very short and often does not allow for distinguishing between successive vehicle passes. Furthermore, sounds of vehicles driving on lanes of opposite direction may overlap and distort one another.
- Sound masking. A heavy truck can emit a noise loud enough to mask the sound of a nearer but lighter car, thus making this car undetectable.
- Intermediate vehicle types (e.g., minibus or pickup truck) make the boundary between light and heavy vehicles more ambiguous. As a result for some specific types of vehicles precise classification is not possible.

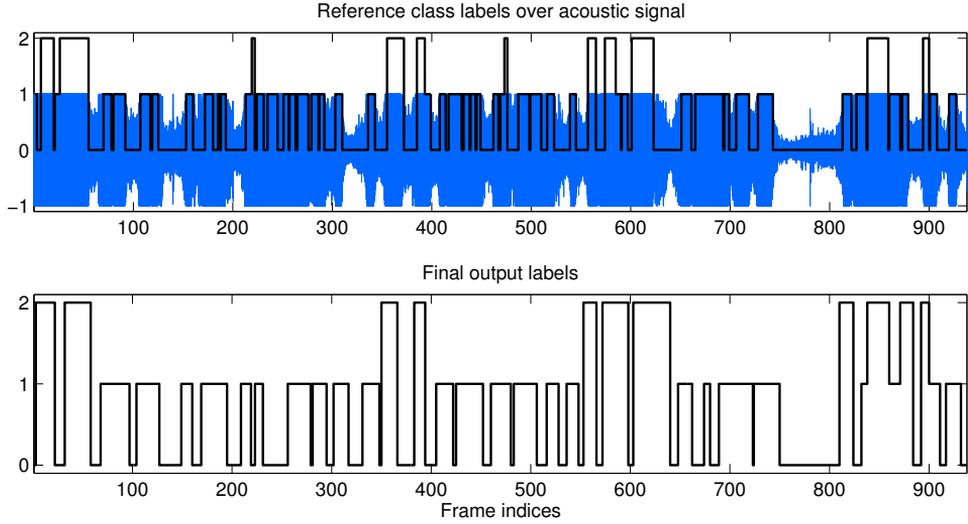


Figure 3.22: Multistage procedure test results for the signal of Experiment 2.

General Testing Results

The multistage procedure operates sufficiently well in both cases of low and heavy traffic. However if the flow of vehicles is consistent and very dense, a decrease of identification quality is witnessed. The influence of background noise, though reduced due to the multistage decision-making logic, cannot be eliminated completely. The algorithm is applicable under different weather conditions, which is demonstrated on the examples of both high and low SNR signals. The sensitivity of the procedure can be adjusted to the needed extent by tuning component parameters. This provides the opportunity to apply the algorithm for classification of various types of moving objects not limited to motorized vehicles.

3.2.4 Real-Time Operation on Embedded Device

Algorithm Complexity Minimization

The most time consuming procedures of the multistage algorithm are feature extraction and correlation coefficient calculation due to a large number of lengthy vector operations. To reduce the number of summations several feature extraction techniques were specifically chosen with similar summands. Analyzing equations (2.16), (2.10), (2.14), (2.17), and (2.18), the recurring elements are $\sum_{k=1}^K |X_t[k]|$, $\sum_{k=1}^K |X_t[k]|^2$, and $\sum_{k=1}^K k \cdot |X_t[k]|$, the first two of which are also present in the correlation calculating equation (3.2). Computing these sums only once and minimizing the number of cycles during feature extraction greatly reduces the number of overall operations.

Equations (2.17) and (2.18) may be further simplified if k is taken as an integer vector index of the corresponding frequency component. The closed form for the sum of K first successive integers is equal to

$$\sum_{k=1}^K k = \frac{1}{2}K(K+1), \quad (3.5)$$

and the sum of squares of K first successive integers is

$$\sum_{k=1}^K k^2 = \frac{1}{6}K(K+1)(2K+1). \quad (3.6)$$

Even if k is chosen non-integer, the sums of the resulting recurrences may still be evaluated [61], however, these closed forms will definitely require more computations than it is needed for (3.5) and (3.6). Calculating the sums of reference vectors and the sums of squared reference vectors only once during the offline stage of algorithm parameter specification turns (3.2) to a more lightweight equation with only one specific summation, which must be performed for each correlation coefficient calculation $\sum_{k=1}^K |X_t[k]| \cdot |X_r[k]|_i$.

Using a power of two as the signal frame length also reduces computation complexity. FFT operation is optimized for frame lengths multiple to powers of two [57]; also in this case many multiplications and divisions are replaced by simpler and faster bit-wise arithmetic shifts.

Implementation and Processing Time Evaluation

For the implementation the embedded device Gumstix Overo Water with System-on-Chip OMAP3530 (600 MHz ARM-Cortex-A8), 256MB RAM and a 4GB microSD card was chosen. For the real-time operation experiment the test signal of Experiment 2 is used. After training the fuzzy rule base and tuning all the parameters of the algorithm, the signal file is streamed to the device input buffer bypassing the ADC at the rate of the sampling frequency in order to simulate real-time data acquisition and operation [16].

For the experiment the frame length was set to $2^{14} = 16384$, $2^{13} = 8192$, and $2^{12} = 4096$ samples, which corresponds to 371.5 ms, 185.8 ms, 92.9 ms at $f_s = 44.1$ kS/s sampling rate, respectively. To be able to operate in real-time each iteration of the identification procedure, therefore, must take less time than the duration of one frame. The length of 16384 samples was chosen initially during the development of the algorithm; two others are added in order to test the processing time and are not well suitable for this specific application due to the decrease in identification quality. Nevertheless, these lengths may be suitable in other applications, and also show the dependence of the required processing time on frame length. The test signal itself is

625.27 seconds long, which corresponds to 1683 frames of length 16384, 3366 frames of length 8192 and 6732 frames of length 4096.

The processing time is measured for the following procedures:

1. FFT execution,
2. RMS energy calculation,
3. Feature extraction,
4. Fuzzy classification,
5. Calculation of correlation coefficients,
6. ASR envelope estimation.

During the experiment the algorithm is made to run to full extent, not terminating during negative detection (i.e., after 2. or 4.) in order to achieve consistent results.

Table 3.6: Processing time mean values for different frame lengths in ms

Frame length	Algorithm sub-procedures				
	1.	2.	3.	5.	Total
4096	24.9	1.4	9.2	3.0	38.6
8192	28.3	1.7	9.9	3.5	43.4
16384	35.0	2.2	11.0	4.4	52.6

The mean values of processing times are presented in Table 3.6. Operations 4. and 6. are excluded from the table, as for all frame lengths the times of 4. are either 30–31 or rarely 61 μ s, and for 6. the processing times are 30–31 μ s. As expected, the most time-consuming operations are FFT execution (consuming more than half of the total processing time) and feature extraction. Process 2. (RMS energy calculation) takes very little time, so during non-detection the system resources are greatly spared. Correlation computation also consumes much time increasing along with frame length and the number of reference vectors. Thus, applying a faster alternative to this method will increase system performance. Altogether, the mean total processing time is significantly shorter than frame duration for all tested frame lengths. Therefore, the algorithm can easily operate in real-time on the given platform.

The distributions of processing times of the steps with the most variance are presented in the form of histograms in Figures 3.23–3.25. The variance of processing time is small and thus the predictability of computation time is high, given that other parallel processes do not interrupt operation flow. A small amount of values noticeably larger than the mean exist for every

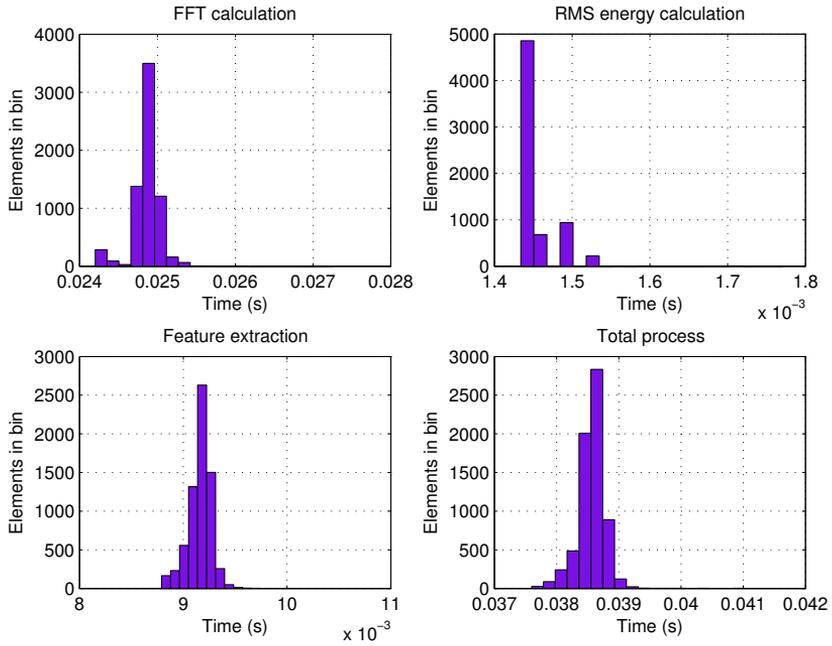


Figure 3.23: Processing time histograms for the frame length of 4096 samples.

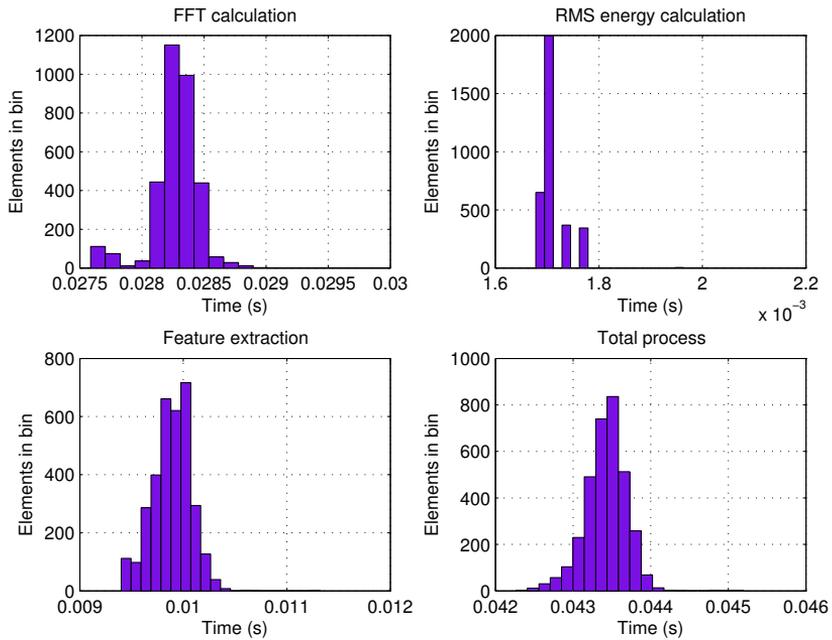


Figure 3.24: Processing time histograms for the frame length of 8192 samples.

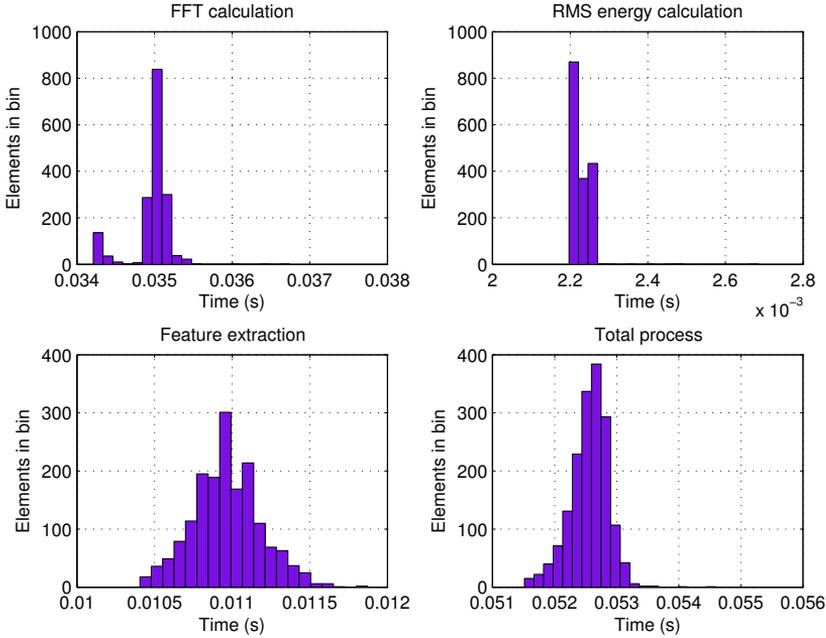


Figure 3.25: Processing time histograms for the frame length of 16384 samples.

sub-procedure and the total process. These abnormalities of long processing duration are most certainly influenced by side processes and problems with memory allocation. Taking this into account and calculating the worst-case total processing times, consisting of maximal values of every sub-procedure time, gives 56.4 ms for frame length 16384, 47.5 ms for 8192 and 44.2 ms for 4096 sample frame lengths, respectively. These estimates are still much shorter than each corresponding signal frame duration.

3.2.5 Multistage Procedure Discussion

The main purpose of the proposed multistage procedure lies in the separation of event detection and identification. Performing computationally expensive operations, such as FFT and feature extraction, only when the initial signal analysis indicates possible event detection, greatly reduces the work load, conserving computational resources and extending battery life. A similar approach is also beneficial for the typical pattern recognition approaches, such as the one presented in Section 3.1. A time domain feature, e.g., the signal energy envelope or the zero crossing interval, can specify signal intensity or half period, respectively. As such properties can be clearly defined for stationary processes, event detection can be quite straightforward.

For non-stationary processes, as the vehicle pass dynamic, neither time nor frequency domain signal properties can be strictly defined. Model based classification aids in confining the general feature sub-spaces, however, this

requires discarding overlapping feature samples belonging to different classes. This may result in information loss and specific parts of the process not being correctly classified. For the vehicle pass process these overlapping feature samples correspond to, e.g., early stages of vehicle approach. Discarding these samples and classifying the vehicles only at the moment of the pass itself is quite permissible, however, for other processes this may be impossible to do without disrupting the identification procedure. Generally, inseparable feature sub-spaces indicate poor choice of the feature set.

The most difficult problem for operation in open environments is situated with high levels of noise with unpredictable properties. Classifier training quality first and foremost depends on the quality of the training data set. The reference features must be chosen with moderate amounts of background noise. Very noisy reference features will most likely produce large, sparse and heavily overlapping feature clusters dependent on the stationary properties of this particular noise. On the other hand, increasing the number of used features will provide more information on the distribution of points in the feature space, thus allowing for more efficient parameter tuning. In simple classification techniques, like correlation analysis, heavy noise may produce misclassifications by correlating with the references. Therefore, if heavy noise is an expected property of the environment, model based classifiers are preferable [106, 134].

The influence of noise coupled with increased uncertainty of an open environment make the discussed procedure applicable only in cases, where system operation bounds are strictly defined. For passing vehicle identification, for example, these bounds may be defined by the sensor's specific placement scheme and FOV (e.g., roadside, one or two lanes). For a more general application field multi-modal signal analysis is preferable with signals of different modalities reassuring each others identification results. Applying multi-channel signal analysis also provides additional information, e.g., object bearing or position estimates. For passing vehicles the constant change in bearing, for example, indicates the pass event and also object direction of movement.

3.3 Conclusions

The discussed procedures of single-sensor signal analysis produce localized situation assessments. Though suffering from problems situated with noise and unknown process states, individual smart sensors are shown to produce accurate process state identification results for bounded application specific problems. The operation scope of each individual smart sensor surely cannot be wide enough to solve each of the tasks assigned to the entire CPS. The simple situation assessments are aggregated through the system WSN, and

presumably false assessments are eliminated by the generation of the global situation assessment. Therefore, process state estimation accuracy of each individual sensor does not have to be perfect in order for the system as a whole to make the correct assessment. Real world process and environment characteristics, such as undefined process states and transitions, undefined number of noise sources with unknown properties, all produce unavoidable classification errors during pattern recognition.

Stationary process state identification naturally presents less problems for pattern recognition and can be performed with basic feature extraction and classification schemes. Non-stationary process identification requires more rigorous signal analysis and is apt to produce less accurate results due to increased complexity in process dynamics. To ensure reliable identification, thorough initial analysis of the process and the signals produced by it is required. During initial analysis the chosen set of signal modalities, signal processing parameters, set of extracted features, classification procedure type and properties are iteratively refined to determine the well separable process defining characteristics. Multi-modal smart sensors have an advantage in this regard, as they are able to analyze a greater set of process characteristics produced by different types of signals.

Chapter 4

Multi-Sensor Solutions

The chapter discusses multi-sensor methods of Direction of Arrival (DOA) and acoustic localization in a sensor network of individual sensor nodes equipped with microphone arrays. The proposed DOA estimation and localization methods are developed under the constraints imposed by the limitations of ad-hoc WSN operation. The chapter first reviews the planar DOA estimation approach for a linear microphone array configuration [10,23]. After that volumetric DOA estimation is discussed for a conical array consisting of a Uniform Circular Array (UCA) with additional vertical microphones [21]. The chapter then continues with the proposed distributed localization approach [10,20]. Experimental verification results [12,13] are presented during the course of every topic discussion. Physical limitations of the array configurations [12,21] and their implementations on embedded hardware [12] are also presented and verified.

4.1 Direction of Arrival Estimation with Linear Arrays

A linear sensor array is the most simple structure for planar Direction of Arrival (DOA) estimation and acoustic localization. In a linear array several sensors are placed facing the same direction along a straight line with specific distances between consecutive sensors. As all the sensors lie in the same plane, volumetric DOA estimation is not feasible for this kind of array structure. This section discusses methods of azimuth ϕ , or the horizontal Angle of Arrival (AOA), estimation for linear arrays with the frontal Field of View (FOV) of $[-90^\circ, 90^\circ]$. The frontal FOV is used because unless using specific directional sensors the linear array structure does not allow for distinguishing between the DOA of the sources in front or at the back of the array. For 360° FOV more reliable array structures are applied. The azimuth

is chosen for the AOA component of DOA, as it is assumed that the array is situated in the x - y plane along the x -axis with zero rotation.

4.1.1 Reduced Functional for SRP-PHAT

As discussed in Section 2.6.1, computing a Steered Response Power (SRP) value for every discrete point if the FOV requires an unjustified large number of operations (2.31). Therefore, it is proposed to reduce the SRP-PHAT functional \mathbf{a} to a set of points along a circumference of a half circle, which covers the front FOV of the linear array. The first two quadrants of the horizontal plane in Cartesian coordinates is divided into n_h possible azimuth angles. A single angle increment is calculated as $\phi_h = \frac{\pi}{n_h}$. The evaluation points are chosen in the planar FOV along a circle with a radius r_{FOV} . The SRP-PHAT evaluation is performed over the half circumference $[0, \pi]$ for the points $\mathbf{a}_h(i) = (x_{h,i}, y_{h,i})$:

$$\begin{aligned} x_{h,i} &= r_{FOV} \cos(i\phi_h), & (0 \leq i \leq n_h), \\ y_{h,i} &= r_{FOV} \sin(i\phi_h), & (0 \leq i \leq n_h). \end{aligned} \quad (4.1)$$

The azimuth is estimated in the direction of elevated SRP values $P(\mathbf{a}_h)$. For a single source case the final azimuth is equal to

$$\phi = \arg \max (P(\mathbf{a}_h)) \cdot \phi_h. \quad (4.2)$$

An example of applying SRP-PHAT with a reduced functional to the signal of man and woman simultaneous speech is presented in Figure 4.1. The figure constitutes a simplified representation of the SRP image presented in Figure 2.14. For the generation of both these figures the same signal frame and the β -PHAT coefficient of $\beta = 0.8$ is used. The circle radius in Figure 4.1 is equal to $r_{FOV} = 1.5$ m (green half circle), and the number of discrete angles is $n_h = 500$, which results in the angular resolution of 0.36° . The SRP values (blue lines) for every point in the functional are scaled to the circle radius with the largest ones pointing in the directions of the two acoustic sources. By comparing Figure 4.1 to Figure 2.14 it is clear that both contain equivalent information in terms of DOA estimation, however, the reduced functional requires 500 SRP evaluations (2.31), as opposed to 3375 for the full SRP image.

4.1.2 Proposed Approach to DOA Estimation

SRP-PHAT is functional-driven in the sense that it calculates its SRP values only for the discrete directions specified by the given functional. In this section a simple technique of DOA estimation for all possible directions (bounded only by the discrete signal sampling interval) is proposed. For

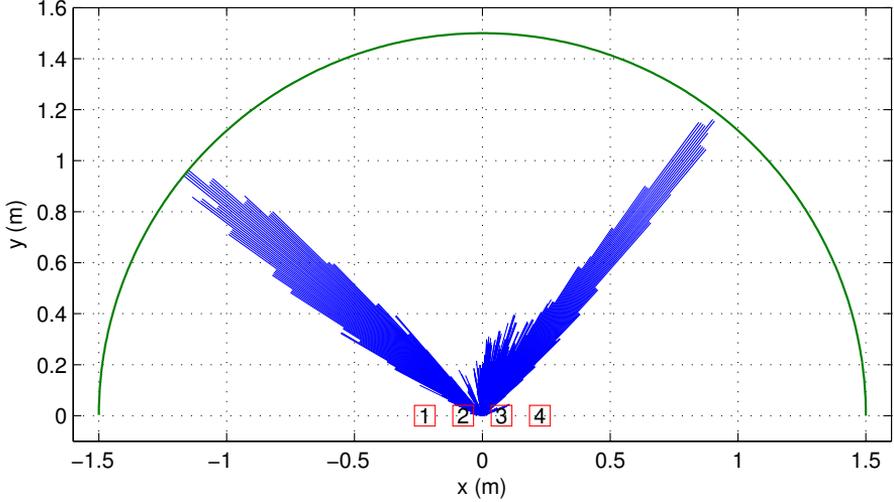


Figure 4.1: Result of applying SRP-PHAT with a reduced functional to the man and woman speech signal frame of Figure 2.14.

Time Difference of Arrival (TDOA) estimation it applies cross-correlation in the time domain, which eliminates the need of applying the computationally heavy FFT. Generally, cross-correlation can be estimated in both the time and frequency domains (which is applicable to both SRP-PHAT and the proposed method), however, calculating the FFT for every frame of each channel of the multi-channel signal requires tremendous amounts of resources and is superfluous if other signal processing procedures do not also use the same data.

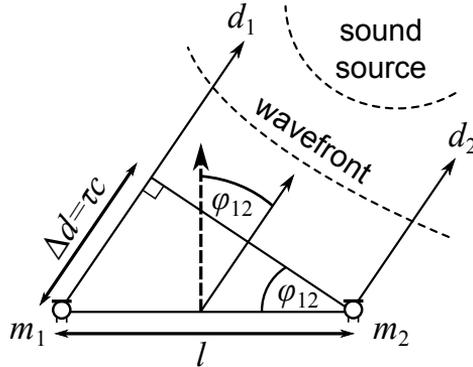


Figure 4.2: DOA estimation for a pair of microphones.

The considered narrow-aperture linear array consists of M microphones. The DOA is estimated for the array front view, i.e., from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$. Estimation is performed for all $\binom{M}{2}$ combinations of microphone pairs. Considering Figure 4.2, the sound wave originating from a source in the far field is acquired

by the microphones m_1 and m_2 with a time delay $\tau = \Delta d/c$, where c is the speed of sound in m/s, calculated using equation (2.30). The TDOA is bounded by the interval $\tau \in [-\tau_{\max}, \tau_{\max}]$, where $\tau_{\max} = l/c$ is the delay of sound traveling directly from one microphone to the other (i.e., at $\pm \frac{\pi}{2}$).

DOA estimation is performed over a $N \times M$ multi-channel signal frame of form (2.1). A separate azimuth estimate $\hat{\varphi}_{ij}$ is made under the far field assumption for every pair of microphones (m_i, m_j) , $i = 1, \dots, M$, $j = 1, \dots, M$. For any pair (m_i, m_j) of consecutive microphones the azimuth estimate is obtained by

$$\hat{\varphi}_{ij} = \sin^{-1} \left(\frac{\tau_{ij} \cdot c}{l} \right) = \sin^{-1} \left(\frac{\Delta k_{ij} / f_s \cdot c}{l} \right), \quad (4.3)$$

where l is the distance, and τ_{ij} is the TDOA of the wavefront between microphones m_i and m_j , respectively. Depending on the chosen pair of microphones, l will vary from l to $l \cdot (M - 1)$. Time delay τ_{ij} is also represented in (4.3) in terms of delay in samples Δk_{ij} and the sampling frequency f_s . To estimate Δk_{ij} cross-correlation is applied to the pair of signals as

$$R_{ij}(\Delta k) = \sum_{k=0}^{N-1} x_{m_i}[k] \cdot x_{m_j}[k - \Delta k], \quad (i < j), \quad (4.4)$$

where N is the length of the signals in samples. The maximum of the cross-correlation then defines the TDOA:

$$\Delta k_{ij} = \arg \max (R_{ij}(\Delta k)). \quad (4.5)$$

The quality of the estimate $\hat{\varphi}_{ij}$ is measured as cross-correlation peak distinctness from its mean level:

$$q_{ij} = \max (R_{ij}(\Delta k)) - \text{mean} (R_{ij}(\Delta k)). \quad (4.6)$$

At this point data validation may be performed. If the maximum of correlation is less than some threshold, the intermediate azimuth estimate $\hat{\varphi}_{ij}$ may be discarded. This way in absence of a sound source or in case of heavy noise invalid estimates are avoided during this initial stage of the procedure. Estimate $\hat{\varphi}_{ij}$ passes validation if it satisfies the control condition

$$\max (R_{ij}(\Delta k)) > (1 + \epsilon) \cdot \text{mean} (R_{ij}(\Delta k)), \quad (4.7)$$

where ϵ is the threshold of correlation peak distinctness, which depends on SNR. During algorithm testing ϵ is set in the interval of 0.2–0.3.

Having calculated $C \leq \binom{M}{2}$ intermediate azimuth estimates (varying slightly due to different inter-microphone distances and naturally occurring errors), the final DOA estimate is derived taking into consideration the following special cases:

1. If the estimates are uniformly spread, no common DOA can be derived.
2. If the estimates are consensual with slight variance, the common DOA can be calculated as some measure of the pair-wise estimates.
3. If the estimates are consensual with slight variance, but some estimates are outside the consensual group, these should be excluded from common DOA calculation.
4. If several distinct groups of consensual estimates exist, the one with more members and better quality should be chosen for a single source case, and several may be considered for multi-source applications.

These conditions are satisfied by applying a partitioning procedure, which performs the task of clustering the $\hat{\varphi}_{ij}^*$ estimates. The coherent estimates of the resulting clusters must lie within sectors with a central angle of no more than φ_{\max} . For example, if $\varphi_{\max} = \frac{\pi}{6}$, then each cluster's coherent estimates must lie no more than $[-\frac{\pi}{12}, \frac{\pi}{12}]$ from the cluster's centroid. The resulting clusters Φ_p , $p = 1, \dots, P$, where P is the number of clusters, each contain n_p estimates $\hat{\varphi}_k$, $k = [1, n_p]$, and the associated quality q_k . The clusters are evaluated in order to find the largest cluster, containing estimates of best quality [21]. Algorithm 4.1 handles the final azimuth calculation for the single source case. The real-valued parameter $\sigma = (0, 1)$ is the threshold of tolerance and the integer parameter n_{\min} is the lower bound for the largest cluster size. The final azimuth estimate ϕ cannot be made if there are insufficient coherent estimates, or if they are of low quality.

Algorithm 4.1 Final azimuth ϕ estimation for a single source

Require: Φ_p , q_k of every $\hat{\varphi}_k \in \Phi_p$, $p = 1, \dots, P$
 get largest cluster size $|\Phi|_{\max}$, maximum quality q_{\max}
if $|\Phi|_{\max} = n_{\min}$ **or** $q_{\max} < \text{allowed}$ **then**
 return $\phi \leftarrow \emptyset$ ▷ initial criteria not met
else if Φ_p of size $|\Phi|_{\max}$ contains $\hat{\varphi}_k$ with q_{\max} **then**
 return $\phi \leftarrow \sum_{k=1}^{n_p} q_k \hat{\varphi}_k / \sum_{k=1}^{n_p} q_k$ ▷ weighted mean
else
 for $i = |\Phi|_{\max} - 1$ **to** $i > n_{\min}$ **do** ▷ search in smaller Φ_p , $n_p > n_{\min}$
 if $\exists q_k \geq \sigma \cdot q_{\max}$ for any $\hat{\varphi}_k \in \Phi_p$, $|\Phi_p| = i$ **then**
 return $\phi \leftarrow \sum_{k=1}^i q_k \hat{\varphi}_k / \sum_{k=1}^i q_k$
 end if
 end for
 return $\phi \leftarrow \emptyset$ ▷ estimates of sufficient quality not found
end if

The partitioning procedure performs final estimate validation by σ and n_{\min} parameters. Tuning the validation parameters allows to adjust the sensitivity and noise tolerance of the procedure. If low variance of estimate quality q is expected, the weighted mean may be replaced by the common mean operation to spare resources. Algorithm 4.1 is easily adjustable to the multi-source case by picking not one but several biggest clusters with best quality estimates.

4.1.3 Performance at Low Sampling Rates

Implementation of the DOA estimation procedure on embedded hardware is limited by the available sampling rate provided by specific embedded ADC. As a result, sampling the multi-channel acoustic signal at the commonly applied rate of around 40 kS/s is not always possible. Furthermore, DOA estimation may be applied to signals of other modalities sampled at significantly lower sampling rates. Therefore, analysis of low sampling rate influence on DOA estimate resolution and accuracy is provided.

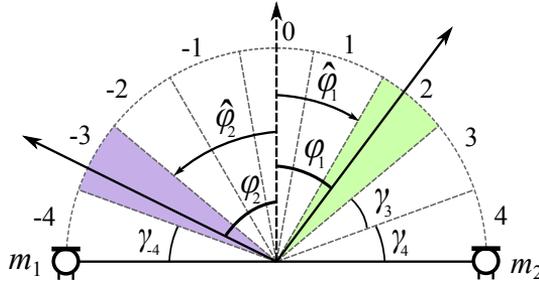


Figure 4.3: Discretization of the AOA resolution defined by Δk_{\max} .

The essential operation for AOA estimation is the signal cross-correlation (4.4). As the time delay $\tau \in \mathbb{R}$ is bounded by τ_{\max} and τ is expressed in delay in samples Δk , then $\Delta k \in \mathbb{Z}$ is also bounded by a maximal sample shift $\Delta k \in [-\Delta k_{\max}, \Delta k_{\max}]$, where Δk_{\max} is calculated as

$$\Delta k_{\max} = \left\lfloor l \cdot \frac{f_s}{c} \right\rfloor, \quad (4.8)$$

where $\lfloor \cdot \rfloor$ denotes rounding to the largest previous integer (floor function). Consequently, the AOA resolution of the sensor pair is reduced to the number of possible discrete AOA values

$$n_{AOA} = 2 \cdot \Delta k_{\max} + 1. \quad (4.9)$$

Figure 4.3 presents a AOA divided into 9 discrete sectors. For any actual AOA φ , only its discrete approximation $\hat{\varphi} \in [\gamma_{-\Delta k_{\max}}, \gamma_{\Delta k_{\max}}]$, corresponding

to the argument of correlation maximum (4.5), can be estimated. For devices capable of operating only at low sampling rates this poses a problem in terms of compromise between the values of l and Δk_{\max} . Consider Table 4.1, which illustrates the dependence of the AOA resolution on l and f_s at the speed of sound $c = 343$ m/s, calculated using (2.30) for the ambient temperature of $t^\circ = 20$ °C. The standard CD sampling rate of $f_s = 44.1$ kS/s is used for reference, and l is calculated by inverting (4.8) as $l = \Delta k_{\max} \cdot c / f_s$. The table shows that to provide even the smallest n_{AOA} the inter-sensor distances must be quite considerable at low f_s . It is clear that if smart sensor dimensions do not exceed, e.g., 15–20 cm, it would not be reasonable to make $l = 1.7$ m to provide the resolution of only 21 possible AOA.

Table 4.1: Required inter-sensor distance for different sampling rates

Δk_{\max}	n_{AOA}	l (cm) for f_s equal to				
		44.1 kS/s	20 kS/s	8 kS/s	2 kS/s	500 S/s
1	3	0.8	1.7	4.3	17.2	68.6
2	5	1.6	3.4	8.6	34.3	137.2
3	7	2.4	5.2	12.9	51.5	205.8
10	21	7.8	17.2	42.9	171.5	686
20	41	15.6	34.3	85.6	343.0	1372
50	101	38.9	85.8	214.4	857.5	3430

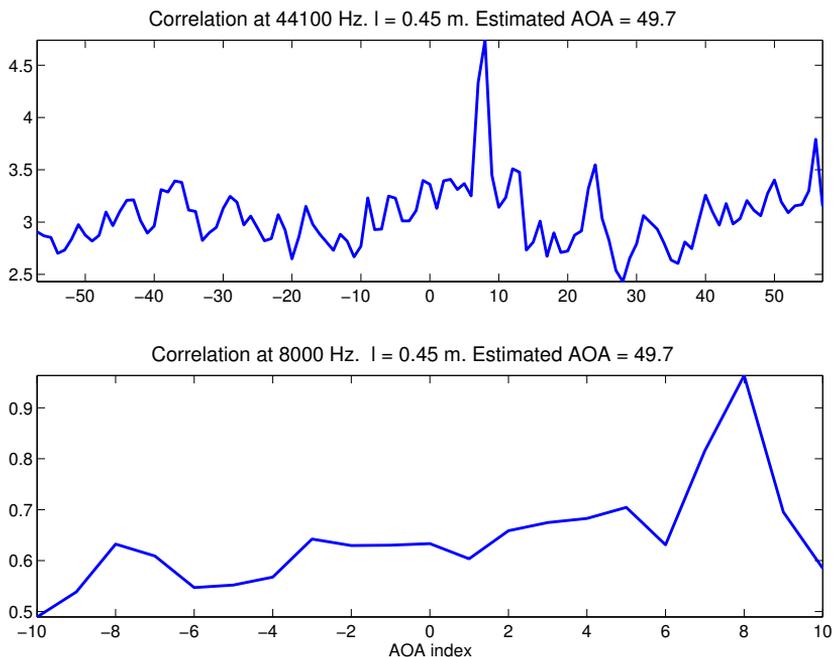


Figure 4.4: Results of signal cross-correlation at different sampling rates.

Low sampling rates also influence cross-correlation coefficients in two ways. Firstly, if the signal contains many strong components in the higher frequencies and they are not acquired at low sampling rates, aliasing may occur, which, in turn, reduces the correlation reliability. A precise peak corresponding to a single Δk loses its steepness and spreads to several values. This makes AOA estimation and the control metric (4.7) less reliable. Secondly, the cross-correlation yields exactly n_{AOA} coefficients, and if this number is low, the correlation peak cannot stand out from the average correlation level as much as in the case of high sampling rates. At low SNR the peak becomes almost uniform with the average level and control metric (4.7) declares the result invalid for the majority of signal frames. Both effects are evident from Figure 4.4. The upper plot presents the result of cross-correlation of two signals sampled at 44.1 kS/s and the lower plot — at 8 kS/s. For both cases the inter-microphone distance is equal to $l = 0.45$ m. The AOA to the acoustic source, as well as signal power, are the same. As n_{AOA} is more than five times larger in the case of $f_s = 44.1$ kS/s, and more signal energy information is contained in a single frame, the correlation peak is much more steep and evident than in the $f_s = 8$ kS/s case. Generally, at a fixed f_s correlation results are improved by increasing the signal frame length, thus providing more signal energy information. Here a compromise between correlation result reliability and the device refresh rate, as well as the amounts of required memory must be reached.

4.1.4 Discussion

SRP-PHAT and the proposed approach to DOA estimation are similar in nature in the sense that both apply the cross-correlation metric for signal phase shifting and estimating TDOA. It may seem that SRP-PHAT is not bound by the discretization of the FOV and AOA by the signal sampling rate. If SRP-PHAT applies an integer lag step correlation function (or beamformer), it is bound by the same limitations, as shown in Figure 4.22 of Section 4.1.3. If the resolution of the SRP spatial vector is higher than the actual resolution bounded by signal properties, the latter can be artificially increased by signal modeling and interpolation. However, the same can be applied to the proposed method.

The performance of the proposed DOA estimation method is evaluated in Section 4.4, and its computational efficiency is evaluated in Section 4.4.3. The computational efficiency of the proposed method is not evident in case of a single narrow-aperture sensor array, however, for a wide-aperture array or a set of distributed arrays the reduction of computational resources is quite considerable if compared to SRP-PHAT.

4.2 Volumetric Direction of Arrival Estimation

In this section the task of 2D DOA estimation is discussed. It consists of estimating the horizontal AOA, or azimuth ϕ , and the polar AOA, or elevation θ , of incoming acoustic waves (see Figure 2.12).

4.2.1 Conical Array Structure

The microphone array structure for this task is designed to meet the desired FOV for localization of ground objects with slight elevation. The targeted application field assumes localization of slowly moving sources (no more than 5 m/s) in both indoor and open outdoor environments. The array consists of M_h microphones, which are equidistantly mounted in a horizontal circular shell, forming a Uniform Circular Array (UCA), and M_v microphones placed vertically upwards from the center of the circular shell. This conical configuration with $M_h = 6$, $M_v = 2$, is depicted in Figure 4.5 along with the desired FOV of the array.

Circular arrays are appealing because they provide 360° FOV with a very simple geometry, however, they can suffer from the influence of acoustic sources not-of-interest arising in this broad FOV. The shell mounting is used to reduce the directivity of the UCA horizontal microphones in order to tackle this problem. In a closed indoor environment this helps to reduce the susceptibility to reverberation. In an open environment, where an unpredictable number noise sources may arise at any given moment, the microphones facing the sound source of interest are less affected by noise sources arising in other directions.

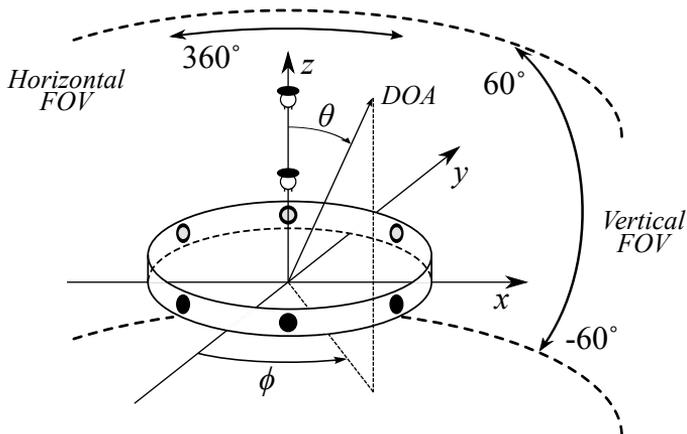


Figure 4.5: Configuration and the desired FOV of the considered conical microphone array.

The application of vertical microphones enables the estimation of elevation θ . As the horizontal microphones are situated in the x - y plane, UCA

alone cannot reliably estimate DOA with non-zero elevation, as depicted in Figure 4.6. The figure portrays a SRP three-dimensional slice of the fourth quadrant of the horizontal plane with array center at $(0, 0, 0)$ (i.e., the right-most corner of the image), with the DOA to the acoustic source being equal to $\phi = 300^\circ$, $\theta = 30^\circ$. As indicated by the confined red region, the application of both vertical and horizontal microphones enables to form a narrow beam of 2D DOA.

The array must be compact (environment integrable and inconspicuous) for the considered applications in CPS. For the first experimental prototype a circular shell with the radius of $r = 7.5$ cm is used. For UCA the angle between two successive microphones relative to the array center O is

$$\alpha = \angle m_i^h O m_{i+1}^h = \frac{2\pi}{M_h}, \quad (1 \leq i < M_h). \quad (4.10)$$

This angle is equal to $\alpha = \frac{\pi}{3}$ for the case of six microphones. The two vertical microphones m_1^v , m_2^v are set at distances of 0.1 and 0.2 m from the horizontal plane to meet the desired vertical FOV of $[-60^\circ, 60^\circ]$.

4.2.2 Volumetric Functional for SRP-PHAT

Computing SRP-PHAT for a volumetric spatial grid requires significantly more resources than is required for the planar case considered in Section 2.6.1. Considering the example presented in Figure 4.6, the volumetric grid of the FOV fourth quadrant, measuring 0.5 m in length, depth and height, is partitioned into discrete points with a distance of 0.01 m along every axis. Thus, the volume vector \mathbf{a} consists of $0.5^3/0.01^3 = 125000$ discrete directions, and so the evaluation (2.31) must be performed the according number of times. For the entire 360° coverage along every axis all eight such FOV segments must be evaluated, which brings the total number of evaluations to one million.

To reduce the SRP-PHAT functional, it is proposed to perform horizontal and vertical DOA estimation separately. In this manner the horizontal plane is divided into n_h and the vertical plane — into n_v possible AOA angles, respectively. A single angle is calculated, similarly to (4.10), as $\phi_h = \frac{2\pi}{n_h}$ and $\theta_v = \frac{\pi}{n_v}$. The points are chosen in the volumetric FOV along a spherical surface with radius r_{FOV} . The horizontal SRP-PHAT evaluation is performed over the entire circumference $[0, 2\pi)$ for the points $\mathbf{a}_{h,i} = (x_{h,i}, y_{h,i}, 0)$:

$$\begin{aligned} x_{h,i} &= r_{FOV} \cos(i\phi_h), & (0 \leq i < n_h), \\ y_{h,i} &= r_{FOV} \sin(i\phi_h), & (0 \leq i < n_h). \end{aligned} \quad (4.11)$$

The azimuth ϕ is estimated in the directions of elevated SRP values. For a single source case it is equal to

$$\phi = \arg \max (P(\mathbf{a}_h)) \cdot \phi_h. \quad (4.12)$$

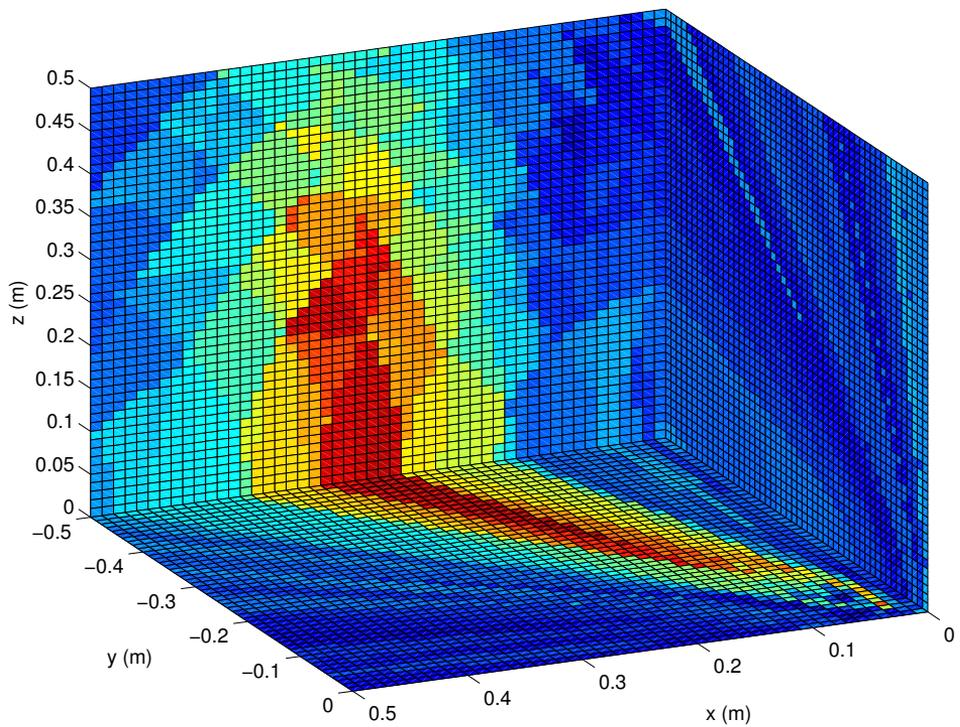
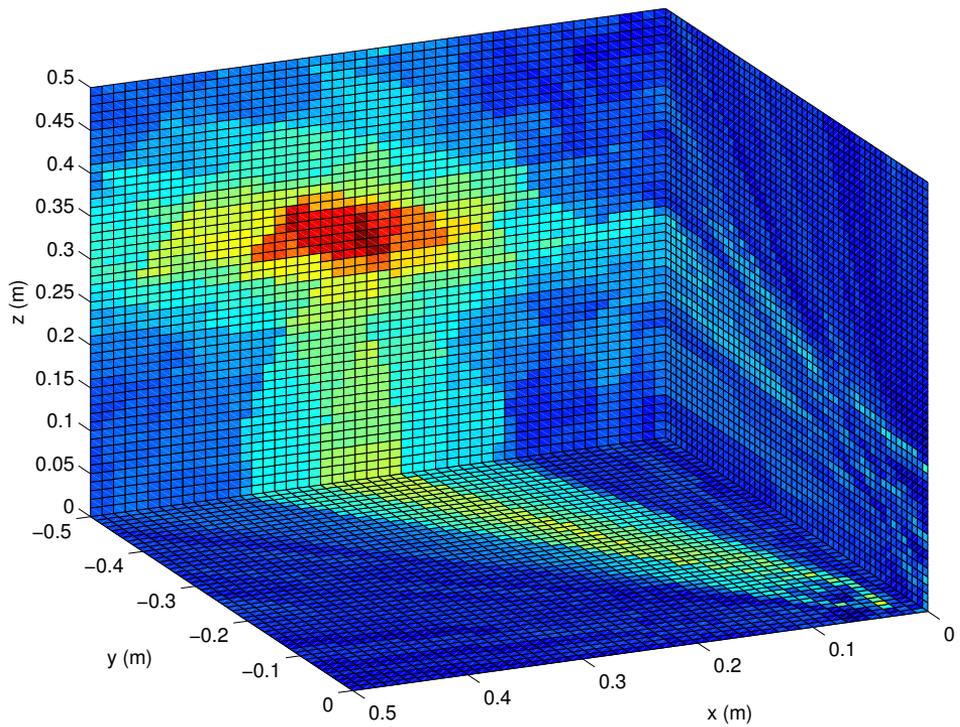


Figure 4.6: Volumetric SRP images computed with (top) and without (bottom) the application of vertical microphones. Acoustic source at $\phi = 300^\circ$, $\theta = 30^\circ$.

After obtaining ϕ , the vertical SRP-PHAT evaluation is performed over the vertical half-circumference from the positive z -axis downward, i.e. $[0, \pi]$, in the direction of established azimuth ϕ for the points $\mathbf{a}_{v,i} = (x_{v,i}, y_{v,i}, z_{v,i})$:

$$\begin{aligned} x_{v,i} &= r_{FOV} \cos(\phi) \sin(i\theta_v), & (0 \leq i \leq n_v), \\ y_{v,i} &= r_{FOV} \sin(\phi) \sin(i\theta_v), & (0 \leq i \leq n_v), \\ z_{v,i} &= r_{FOV} \cos(i\theta_v), & (0 \leq i \leq n_v). \end{aligned} \quad (4.13)$$

The elevation angle is estimated in the direction of elevated SRP, and also brought to a more comprehensive interval $[\frac{\pi}{2}, -\frac{\pi}{2}]$ from the positive z -axis downward:

$$\theta = \frac{\pi}{2} - \arg \max (P(\mathbf{a}_v)) \cdot \theta_v. \quad (4.14)$$

In a multi-source case where, several ϕ of elevated SRP values are chosen, the elevation angle pass must be performed for each ϕ , which increases resource demand, however, is less computationally expensive than, e.g., passing all discrete points along a spherical surface.

4.2.3 Proposed Volumetric DOA Estimation Approach

Even with a reduced functional, SRP-PHAT still requires significant resources and processing time, because it performs cross-correlation between all pairs of microphones for all specified directions. The proposed approach focuses on reducing the number of microphone pairs for cross-correlation and the number of discrete directions per each pair [21].

The proposed two-stage approach is designed for a conical array with M_h horizontal and M_v vertical microphones, discussed in Section 4.2.1. As the horizontal microphones take a directional DOA estimation approach, the pairs of microphones for azimuth estimation are chosen such, that their inter-sensor angle is less than $\frac{\pi}{2}$: $\alpha_{ij} = \angle m_i^h O m_j^h < \frac{\pi}{2}$. Let us denote the set of these pairs as

$$A_h = \left\{ \left(m_i^h, m_j^h \right) \subseteq S_2^{M_h} \mid \alpha_{ij} < \frac{\pi}{2} \right\}, \quad (4.15)$$

where $S_2^{M_h}$ is the set of all combinations of horizontal microphone pairs with cardinality $|S_2^{M_h}| = \binom{M_h}{2}$. A separate azimuth estimate $\hat{\phi}_{ij}$ is computed for every pair $(m_i^h, m_j^h) \subseteq A_h$, and a final estimate ϕ is made.

For the elevation angle estimation a set of horizontal microphones situated within the half-circumference of direction ϕ is chosen:

$$A_{\text{act}} = \left\{ m_i^h \mid -\frac{\pi}{2} \leq \alpha_{\phi, m_i^h} < \frac{\pi}{2} \right\}, \quad (4.16)$$

where $\alpha_{\phi, m_i^h} = \phi - \alpha_{m_i^h}$ is the angle between the azimuth estimate and the microphone position. Here A_{act} denotes the set of so-called ‘‘active’’ horizontal microphones. The pairs participating in elevation estimation consist of:

pairs between every active microphone and every vertical microphone; every pair of vertical microphones. Lets denote this set as

$$A_v = \left\{ \left(m_i^h, m_j^v \right) \mid m_i^h \in A_{act}, j = [1, M_v] \right\} \cup S_2^{M_v}, \quad (4.17)$$

where $S_2^{M_v}$ is the set of all combinations of vertical microphone pairs with cardinality $|S_2^{M_v}| = \binom{M_v}{2}$. A separate elevation estimate $\hat{\theta}_{ij}$ is computed for every pair $(m_i^*, m_j^*) \subseteq A_v$ and the final estimate θ is made.

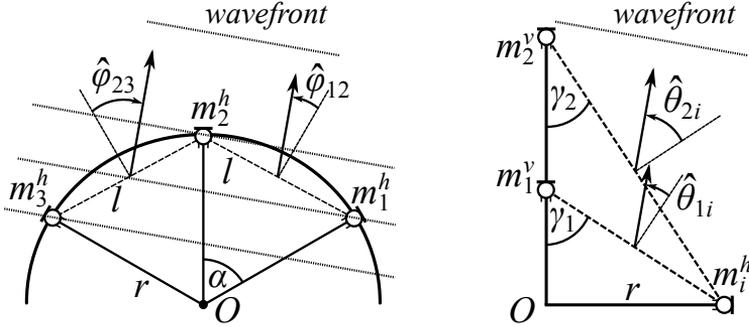


Figure 4.7: Horizontal (left) and vertical (right) AOA estimation for a far field acoustic source using a conical microphone array.

Azimuth Angle of Arrival Estimation

For azimuth estimation the far field disposition of the acoustic source is assumed. The initial azimuth estimates are made for every pair of horizontal microphones $(m_i^h, m_j^h) \subseteq A_h$, as portrayed in Figure 4.7 (left). For any pair (m_i^h, m_j^h) of consecutive microphones the azimuth estimate is obtained, similarly to Section 4.1.2, by applying (4.3), where l is the distance between two consecutive microphones, calculated as

$$l = 2r \sin\left(\frac{\alpha}{2}\right) = 2r \sin\left(\frac{\pi}{M_h}\right), \quad (4.18)$$

where α is calculated for UCA by applying (4.10). For non-consecutive microphones, l is calculated by substituting α in (4.18) with its multiple. For estimation of TDOA τ_{ij} in terms of delay in samples Δk_{ij} , the cross-correlation (4.4) is applied to the pair of signals (m_i^h, m_j^h) as it is done in Section 4.1.2. The quality metric (4.6) is calculated per each estimate $\hat{\varphi}_{ij}$, and verification (4.7) is performed.

Each estimate $\hat{\varphi}_{ij}$ is made for the middle point if the inter-microphone distance (see Figure 4.7 left) and takes the values of $\hat{\varphi}_{ij} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, assuming

negative values if the source is left, positive if the source is right, and zero if the source is in front of the pair. Thus, $\hat{\varphi}_{ij}$ are adjusted to array angle coordinates as

$$\hat{\varphi}_{ij}^* = \hat{\varphi}_{ij} + ((i-1)\alpha + (j-1)\alpha) / 2. \quad (4.19)$$

Consensual directions are found among the estimates by applying the partitioning procedure discussed in Section 4.1.2. The procedure performs the task of clustering the $\hat{\varphi}_{ij}^*$ estimates such, that the coherent estimates must lie within a sector with a central angle no more than φ_{\max} , and computes the final azimuth ϕ by applying Algorithm 4.1.

Elevation Angle of Arrival Estimation

Elevation estimation is performed once the azimuth is set. As the horizontal microphones are not in line with the wavefront, incoming at ϕ AOA, the signals of active microphones in A_{act} must be shifted to meet the TDOA at ϕ . In computing this shift the far field assumption is also made. Its inconsistency produces negligible error, as it is discussed in Section 4.2.4.

After the signals $x_{m_i^h \in A_{\text{act}}}$ have been shifted to meet the ϕ AOA, elevation estimates $\hat{\theta}_{ji}$ for all microphone pairs in A_v are obtained by applying (4.3) and (4.4). The pairs of horizontal and vertical microphones, as portrayed in Figure 4.7 (right), are set at angles γ_j , $j = [1, M_v]$, relative to the vertical plane. Thus, the initial estimates are steered by

$$\hat{\theta}_{ji}^* = \gamma_j - \hat{\theta}_{ji} = \tan^{-1} \left(\frac{r}{z_{m_j^v}} \right) - \hat{\theta}_{ji}, \quad (4.20)$$

where $z_{m_j^v}$ is the distance from O to m_j^v on the z -axis. Final elevation AOA estimation is performed by applying the previously discussed partitioning procedure to clusters Θ_p and computing the final elevation angle estimate θ by applying Algorithm 4.1.

4.2.4 Signal Shifting and Influence of Near Field Error

As for elevation estimation the azimuth estimate is fixed, the signal of the active horizontal microphones in A_{act} must be shifted to meet with the ϕ AOA. Consider Figure 4.8. The far field planar acoustic wave reaches microphone m_i^h in the horizontal plane with a spatial delay h , which is defined as

$$h = a \cos \beta = 2r \sin \left(\frac{\alpha_{\phi, m_i^h}}{2} \right) \cos \left(\frac{\pi - \alpha_{\phi, m_i^h}}{2} \right). \quad (4.21)$$

The signal of microphone m_i^h must thus be left-shifted for $\Delta k_f = \left\lceil h \frac{f_s}{c} \right\rceil$ samples.

Table 4.2: Influence of near field error at different d and f_s

$c = 343$ (m/s), $r = 7.5$ (cm), $\alpha_{\phi, m_j^h} = \frac{\pi}{2}$				
d (m)	f_s (S/s)	$R - d$ (cm)	Δ_{nf} (cm)	Δk_{nf} (S)
0.05	48000	9.51	2.06	3
0.05	32000	9.51	2.06	2
0.05	24000	9.51	2.06	1
0.1	48000	8.97	1.52	2
0.1	32000	8.97	1.52	1
0.1	24000	8.97	1.52	1
0.2	48000	8.44	0.99	1
0.7	48000	7.81	0.36	1

$M_v = 2$, $z_{m_j^v} = \{0.1, 0.2\}$ m. For signal acquisition Vansonic PVM-6052 condenser microphones are used. The microphones are connected through an amplification board to an Agilent U2354A data acquisition device, which samples the signals at $f_s = 48$ kS/s per channel. The data is acquired and processed in the MATLAB environment using the Data Acquisition Toolbox. The processing is performed frame-by-frame with a step of 0.1 seconds for both the proposed method and SRP-PHAT.

The experiments for single source DOA estimation are performed indoors under SNR conditions of approximately 20 dB. A loudspeaker, reproducing human speech, is placed at certain angles within a distance of 1–2 m from the array. A total of four experiments is considered. For the first two the speaker remains at $\phi = 300^\circ$ (i.e., in the direction of the 6-th microphone) and is lifted and lowered to test θ estimation accuracy. The speaker either remains at a certain θ for 3–4 seconds (Experiment I), or moves constantly (II). For the other two experiments the speaker is moved around the array, DOA varying in both ϕ and θ . Here the speaker is also displaced either fixing the position with stops (III), or moving constantly (IV).

The parameters for the SRP-PHAT approach, reviewed in Section 4.2.2, are set as $r_{FOV} = 0.5$ m, $n_h = 500$, $n_v = 250$, which results in unit angle accuracy of $\phi_h = \theta_v = \frac{\pi}{250} \simeq 0.72^\circ$. The approach thus requires 750 evaluations (2.31), as opposed to hundreds of thousands in case of exhaustive search. The result of DOA estimation for a single frame is presented in Figure 4.9. In the figure the azimuth estimates (on the left) are normalized to the FOV radius r_{FOV} , and the elevation estimates (on the right) are normalized to the interval $[0, 1]$. Both global maxima of $\phi = 300^\circ$ and $\theta = 30^\circ$ are distinctive among local surplus peaks of SRP, likely arising due to reverberation. The elevated θ levels at approximately -10° are noteworthy, apparently arising due to sound reflections from the floor surface, as studied in [138].

The proposed approach utilizes 19 pairs of microphones according to the definitions of A_h and A_v in (4.15), (4.17): 6 consecutive pairs $m_i^h m_{i+1}^h$ ((1, 2), (2, 3), etc.); 6 pairs over one microphone $m_i^h m_{i+2}^h$ ((1, 3), (2, 4), etc.); 6 pairs between two vertical m_j^v and three horizontal m_i^h microphones; 1 pair of vertical microphones $m_i^v m_{i+1}^v$. Thus the number of pairs is less than for the SRP-PHAT case, which utilizes $\binom{8}{2} = 28$ pairs. This difference will be more evident, if applied to a larger number of microphones. The clustering parameters are set to $\varphi_{\max} = 30^\circ$, $\theta_{\max} = 10^\circ$, $\sigma = 0.6$. The result of DOA estimation with the proposed method for a single frame is presented in Figure 4.10. In the figure the azimuth estimates (on the left) are colored by quality from worst to best as follows: black, blue, green, red. Elevation estimates (on the right) are presented for 10 successive frames up to the frame at hand with the colored lines denoting the intermediate estimates by microphone pairs in A_v . The thick black line denotes the final estimates for both azimuth and elevation. The initial azimuth estimates of Figure 4.10 strongly resemble the SRP peak distribution pattern of Figure 4.9, indicating a similar reaction to reverberation.

The DOA estimates of both methods for signal segments of Experiments I and IV are presented in Figure 4.11 and Figure 4.12, respectively. Static position of a highly elevated source seems to disrupt ϕ estimation for both methods, which is evident in the interval of $\phi = [50^\circ, 60^\circ]$ in Figure 4.11. During constant movement this static error effect is not observed, however, certain irregularities do arise during rapid movement, e.g., during the interval of 3–6 s in Figure 4.12. Generally, the proposed method is not inferior to SRP-PHAT in the considered FOV, while requiring less processing time. Table 4.3 presents the processing time (average over 10 runs) and DOA estimate Root Mean Square Error (RMSE) for the four experiments. As the processing time is measured in MATLAB running on a PC, it is by no means an adequate measure of computation speed on embedded hardware, however, it shows that the proposed method operates more than 3 times faster than SRP-PHAT over the reduced functional. As the testing was performed manually, the error in speaker placement cannot be strictly accounted for, and thus exact reference AOA are unspecified. Therefore, the RMSE is calculated as a difference between the estimates produced by SRP-PHAT and the ones produced by the proposed method:

$$\text{RMSE}(\phi) = \sqrt{\frac{1}{N_E} \sum_{i=1}^{N_E} \left(\phi_i^{(\text{SRP})} - \phi_i^{(\text{Prop})} \right)^2}, \quad (4.25)$$

where N_E is the total number of estimates. The RMSE remains in reasonable bounds for all experiments, if considering close range localization.

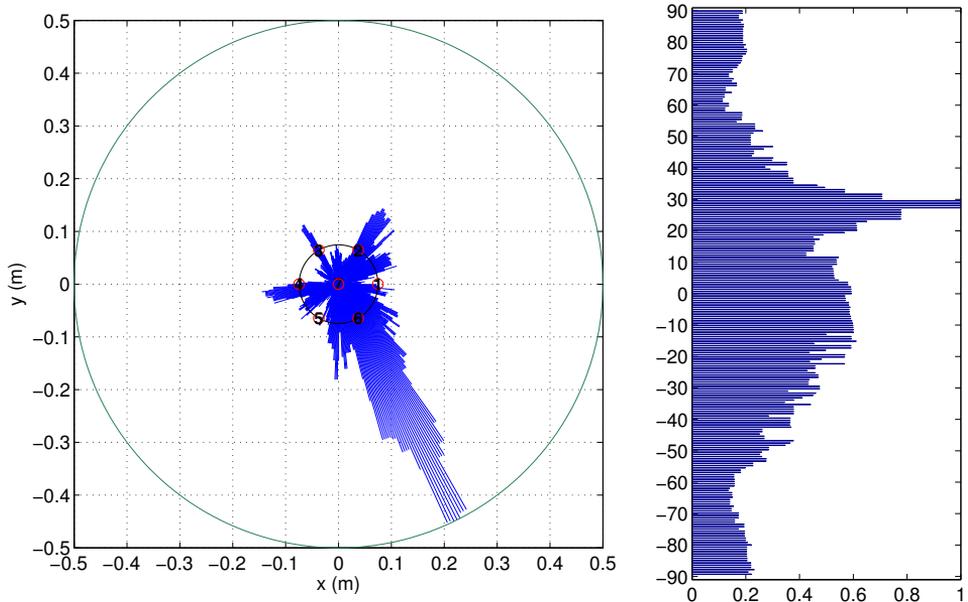


Figure 4.9: Result of SRP-PHAT over reduced functional for a frame of Experiment I.

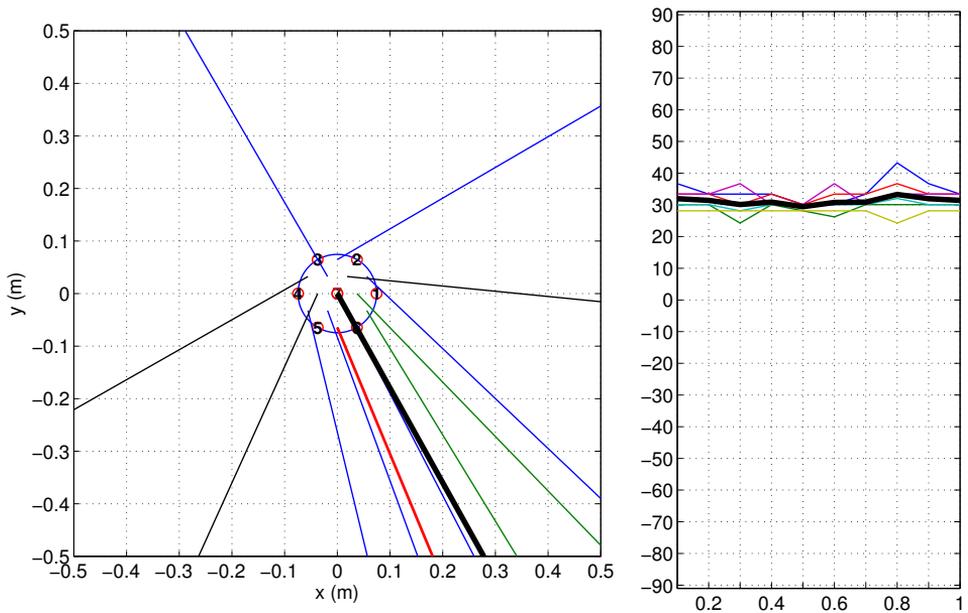


Figure 4.10: Result of the proposed method for a frame of Experiment I.

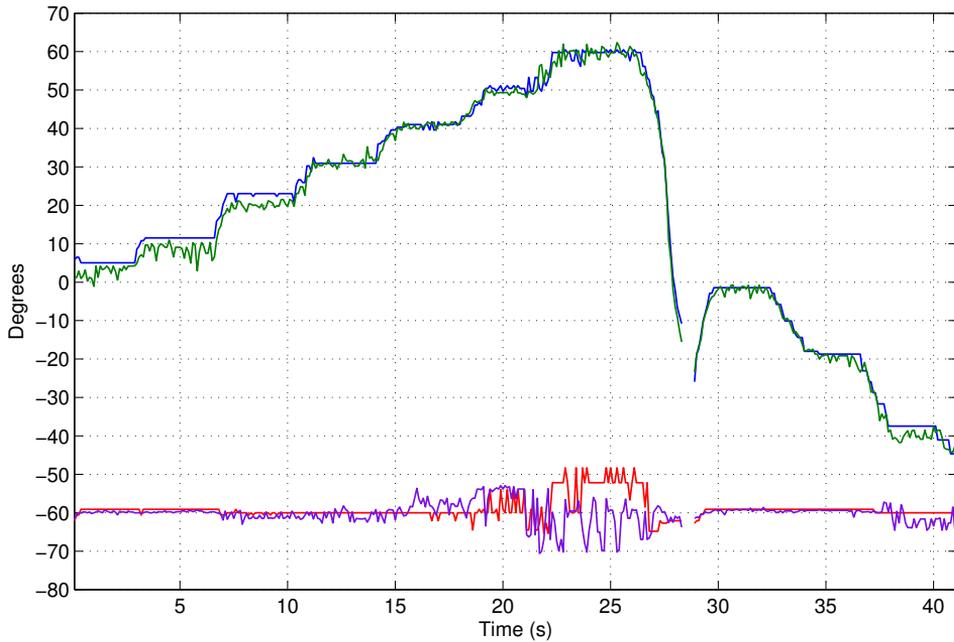


Figure 4.11: DOA estimates for Experiment I. Red — ϕ by SRP-PHAT, purple — ϕ by the proposed method, blue — θ by SRP-PHAT, green — θ by the proposed method.

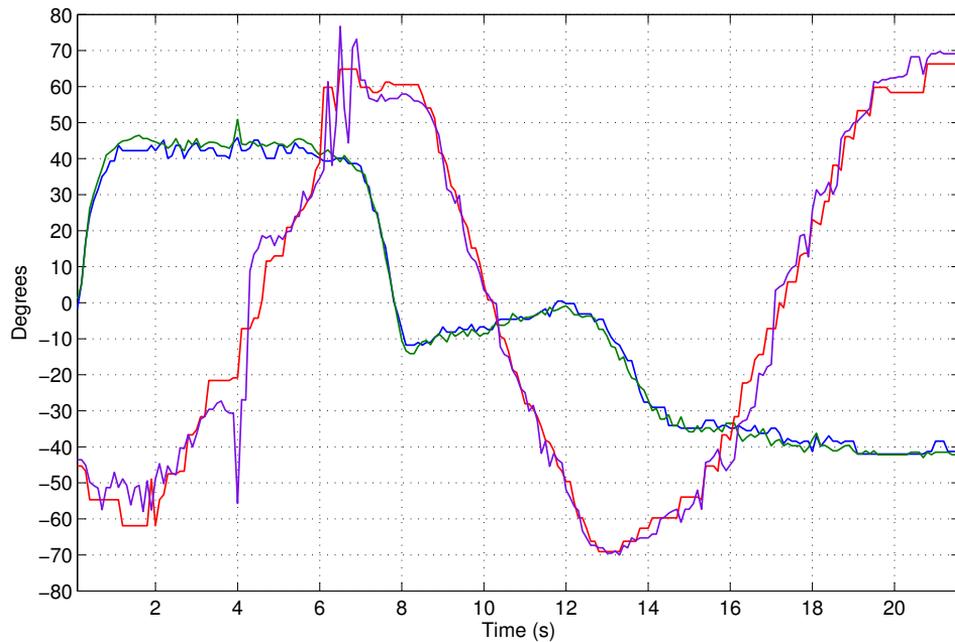


Figure 4.12: DOA estimates for Experiment IV. Red — ϕ by SRP-PHAT, purple — ϕ by the proposed method, blue — θ by SRP-PHAT, green — θ by the proposed method.

Table 4.3: Experimental signal processing times and RMSE of DOA

Experiment	Length (s)	t. proc. (s) SRP-PHAT	t. proc. (s) proposed	RMSE ϕ (deg)	RMSE θ (deg)
I	70.0	283.45	84.25	4.51	2.29
II	62.0	248.77	74.63	5.22	1.86
III	86.0	343.30	103.68	9.43	2.84
IV	52.0	210.97	62.88	13.67	1.99

4.3 Acoustic Localization

Even optimized localization algorithms, e.g, SRP-PHAT with Stochastic Region Contraction (SRC) reviewed in Section 2.6.2, require a significant amount of resources while starting the evaluation on the initial search area or volume. Furthermore, the convergence on a sharp maximum may be guaranteed only if it exists in the FOV. For many applications and monitored object types this is not always true. Large objects, like vehicles or other machinery, do not have a single point of sound emission, rather they appear as spread regions of heightened acoustic energy with several maxima. On the other hand, if no sound source is present in the FOV, the localization algorithms will search for maxima in ambient noise, which produces useless results while consuming resources. The reduction of the initial search area, firstly, allows to estimate the presence of the sound source in the FOV, and secondly, greatly reduces the computational load of localization.

This section presents the proposed method of Initial Search Region Reduction (ISRR) designed for distributed sensor networks. The method operates on a set of distributed sensor arrays and confines the region of acoustic source search. The global maximum of cumulative acoustic energy is estimated by applying SRP-PHAT with SRC to the confined region for wired synchronous system experiments. As neither inter-node wired communication, nor WSN node clock synchronization are assumed in the targeted system design, the regions confined by ISRR are used as estimates of approximate localization in the prototype implementation.

4.3.1 Distributed Localization in WSN

The considered WSN system is designed for localizing grounded acoustic sources. The sensor arrays are placed in the monitored environment in the horizontal plane, and localization is performed by estimating the coordinates (x, y) of sound emitting objects. The proposed WSN architecture is designed for applications in both open outdoor (urban, woodland, etc.) and confined indoor (home, office, industrial facility, etc.) environments. The network consists of two types of nodes: smart sensors and fusion nodes. Multi-channel

smart sensors acquire acoustic information and perform local computations. Fusion nodes gather information from the smart sensors and perform further steps of localization. Such a configuration increases robustness due to high decentralization. A large number of distributed measurement points also simplifies multiple source localization, as the monitored area is divided into smaller local regions.

The multi-channel sensors are dispersed in the monitored environment either in an orderly or random fashion. In confined environments an orderly placement is more likely, because sensors are usually mounted on room walls or ceilings. In open environments, however, it is rarely the case as the sensors may be attached to buildings, light posts, trash bins, etc., in urban environments and to trees, rocks, etc., in natural environments. Thus, a general case is assumed, where the sensor's location is defined by the coordinates of its point of reference (x_r, y_r) and the angle α , by which the sensor is steered from the global angle of reference, as it is shown in Figure 4.13. For example, sensor location may be estimated via the Global Positioning System (GPS), in which case the point of reference is the GPS unit. For environments, where GPS signals are unavailable, other location algorithms based on Radio Frequency (RF) [149] or sound [95] may be adopted. The global angle reference may be defined by Earth's magnetic field, and the angle α estimated using a digital compass. The central point of the microphone array (x_0, y_0) , for which the DOA is actually estimated, is defined by the reference point (x_r, y_r) and may coincide with it. The coordinates of each microphone in the array are then calculated based on the array geometry. In a linear array the coordinates of the i -th microphone (x_i, y_i) are shifted from (x_0, y_0) and steered by α as

$$\begin{bmatrix} x_i^{(\text{rot})} \\ y_i^{(\text{rot})} \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x_i - x_0 \\ y_i - y_0 \end{bmatrix}. \quad (4.26)$$

Sensor nodes are partitioned into groups, where a single node can belong to any number of groups. The whole network may consist of several groups, or each group can constitute a separate sub-network. Group partitioning is in essence a clustering task, for which two aspects are taken into consideration. Firstly, nodes must have a common FOV for all arrays to observe the same area. In this regard, the observed area is not necessarily enclosed by sensor nodes, as shown in Figure 4.13, but may be observed from one or several sides. Secondly, a group must have a certain degree of homogeneity. Nodes located too far from the group's centroid may be useless to the localization effort in low SNR environments, or when the sound emitted by the source of interest is too weak. Furthermore, non-homogeneous groups present additional challenges for wireless communication.

Fusion nodes of the WSN perform sensor node grouping during network initialization and later participate in localization. For an orderly configuration of sensor nodes a single fusion node may be assigned to coordinate the activity of the whole WSN. In a random configuration each sensor node may be a part of several groups and each group may be governed by several fusion nodes. In order to ensure coverage of all groups, fusion nodes reach an agreement concerning which node will govern which sensor node group. In this process communication signal strength is taken into account, meaning that a fusion node will adopt a group, to which it has the strongest connection. However, if there exists an ungoverned group, a redundant (i.e., covering an already covered group) fusion node closest to it will switch to that group.

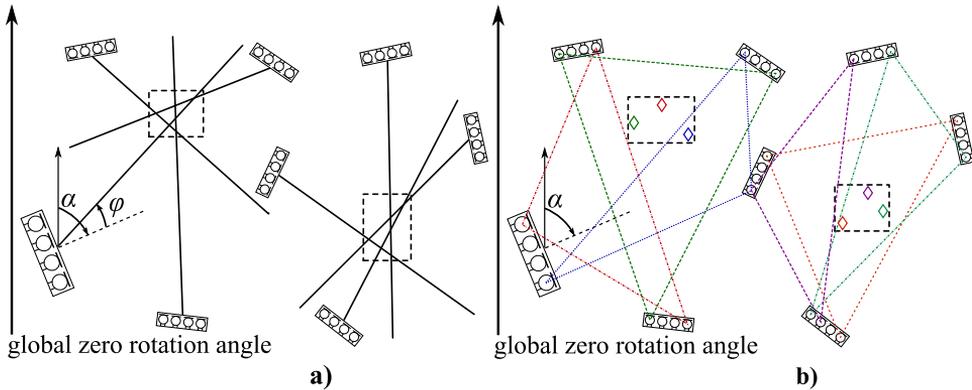


Figure 4.13: Initial search region estimation by (a) DOA estimation and (b) multilateration in a random configuration of sensor array blocks.

4.3.2 Initial Search Region Reduction

In a distributed WSN architecture each array sub-block is implemented on a separate sensor node. From a classical standpoint sub-arrays with common FOV form a single wide-aperture array with a wide FOV. If raw signal communication was possible in the WSN, real-time operation of classical localization techniques utilizing every pair of microphones would still not be feasible as the number of pairs for a combined set of sub-arrays is far greater than the number of pairs in each separate sub-array. ISRR performs search region confinement without utilizing inter-node information and raw signal communication.

ISRR is performed by estimating the DOA for every sub-array and finding the region of common DOA (i.e., the intersection of DOA vectors) as is shown in Figure 4.13a. An alternative approach of choosing sensor triplets and performing multilateration to retrieve the source coordinate estimates is also considered, where the aggregate of coordinate estimates denotes the search

region (Figure 4.13b). The multilateration approach cannot be implemented in WSN as it requires node synchronization and raw signal communication for TDOA estimation. It is used in this work merely for comparison with the proposed DOA approach.

Though the ISRR procedure principle is depicted in Figure 4.13 with linear sub-array blocks, it does not depend on any specific array configuration. This portrayal is chosen for presentation conformity because the procedure is further tested on linear array nodes.

The DOA Approach to ISRR

Having K microphone arrays observing a common FOV, the ISRR procedure is performed in the following steps:

1. Estimate the DOA for each of K arrays.
2. Generate vectors spanning from the array centers to the bounds of the FOV in the directions of DOA.
3. Find points of intersections of these vectors.
4. Find groups of points no farther than D_{\max} distance units (meters) from their centroid and enclose the areas in which these groups reside in rectangles (or other appropriate geometrical shapes).
5. Perform control of false detection, discard areas that do not meet specific criteria (optional).

Step 1. is performed on each sensor node, steps 2.–5. are performed on the group's fusion node. The DOA for planar localization consists of the azimuth ϕ , which is estimated for every array by applying the DOA estimation procedure for linear arrays, presented in Section 4.1.2. If other types of microphone arrays are used in the system, the appropriate procedure of DOA estimation is applied to them. For example, azimuth estimation for UCA is performed for the horizontal microphones defined by (4.15), as discussed in Section 4.2.3.

The fusion node receives $K_1 \leq K$ DOA estimates ϕ_i , $i \in (1, \dots, K_1)$ and adds the sensor node rotation angles α_i to them. As only part of the smart sensors may perceive the signal and estimate validation may fail on others, the number of incoming DOA estimates can be less than the number of sensor nodes in the group. For each device, which has estimated DOA, a vector \overrightarrow{AB}_i is computed with the starting point $A_i = (x_{1,i}, y_{1,i})$ being the coordinate of i -th array's center and the ending point $B_{i^*} = (x_{2,i}, y_{2,i})$ being the point at the bound of the FOV steered by ϕ_i from the array's center. Intersection

points of all pairs $\overrightarrow{AB}_h, \overrightarrow{AB}_q$, such that $h \in (1, \dots, K_1)$, $q \in (1, \dots, K_1)$, $h \neq q$, are calculated as

$$I_{hq} = (I_x, I_y)_{hq} = \left(\frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(x_{1,q} - x_{2,q}) - (x_{1,h} - x_{2,h})(x_{1,q}y_{2,q} - y_{1,q}x_{2,q})}{(x_{1,h} - x_{2,h})(y_{1,q} - y_{2,q}) - (y_{1,h} - y_{2,h})(x_{1,q} - x_{2,q})}, \right. \\ \left. \frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(y_{1,q} - y_{2,q}) - (y_{1,h} - y_{2,h})(x_{1,q}y_{2,q} - y_{1,q}x_{2,q})}{(x_{1,h} - x_{2,h})(y_{1,q} - y_{2,q}) - (y_{1,h} - y_{2,h})(x_{1,q} - x_{2,q})} \right). \quad (4.27)$$

As a result we have a set of I_{i^*} intersections, $i^* \in (1, \dots, K_2)$, $K_2 \leq \binom{K_1}{2}$. To get the confined search areas these intersection points are partitioned by their relative distance. For the maximum distance D_{\max} the partitioning is performed according to Algorithm 4.2. In the algorithm the set of all intersection points \mathbf{I} of dimensions $K_2 \times 2$ consists of K_2 coordinate vectors $I_{i^*} = (I_{i^*,x}, I_{i^*,y})$. Therefore, the initial cardinality of \mathbf{I} is $|\mathbf{I}| = K_2$. The procedure separates the points by distance into partitions $\mathbf{P}_j \subset \mathbf{P}$, each consisting of some number of vector coordinates $P_{j,i^{**}} = (P_{i^{**},x}, P_{i^{**},y})_j$. As Algorithm 4.2 operates in planar Cartesian coordinates, each coordinate and centroid value is defined by two parameters. Centroid $\mathcal{C}(\mathbf{I})$ contains $\mathcal{C}(I_s)$, $s = 1, 2$, for x and y parameters, respectively. Centroid $\mathcal{C}(\mathbf{P}_j)$ contains $\mathcal{C}(P_{j,s})$, $s = 1, 2$, for x and y parameters, respectively. For other applications a different dimensionality for the Euclidean distance can be chosen.

Algorithm 4.2 Intersection point partitioning procedure

Require: $\mathbf{I} \neq \emptyset$

if $|\mathbf{I}| = 1$ **then** ▷ only 1 point

return $\mathbf{P} \leftarrow I_1$

end if

$j = 0$ ▷ initial number of partitions

while $|\mathbf{I}| > 0$ **do**

$\mathcal{C}(\mathbf{I}) = 1/|\mathbf{I}| \cdot \sum \mathbf{I}$ ▷ centroid of free points

$D_k = \sqrt{\sum_{s=1,2} (I_{k,s} - \mathcal{C}(I_s))^2}$, $k = 1, \dots, |\mathbf{I}|$ ▷ Euclidean distance

$p = \arg \min(D)$, $j = j + 1$, $\mathbf{P}_j \leftarrow I_p$, $I_p \leftarrow \emptyset$ ▷ new partition

do

$\mathcal{C}(\mathbf{P}_j) = 1/|\mathbf{P}_j| \cdot \sum \mathbf{P}_j$ ▷ partition centroid

$D_k = \sqrt{\sum_{s=1,2} (I_{k,s} - \mathcal{C}(P_{j,s}))^2}$, $k = 1, \dots, |\mathbf{I}|$

if $\min(D) \leq D_{\max}$ **then** ▷ allowed distance

$p = \arg \min(D)$, $\mathbf{P}_j \leftarrow I_p$, $I_p \leftarrow \emptyset$ ▷ add to partition

end if

while $\min(D) \leq D_{\max}$

end while

return \mathbf{P} ▷ return established partitions

After obtaining the partitions \mathbf{P} , their areas are enclosed by rectangles with the edges denoted by the partitions minimal and maximal values of x and y components. A small constant is added to single point partition areas in order to ensure minimal area (0.1 m is chosen for the experiments). As a result several regions may occur in the same FOV. Also while a vector of DOA from one array may cross with several other vectors, redundant “echoing” regions may arise. These may be removed by additional control metrics or by using tracking filters. For a wired implementation the maximal SRP value of each confined region can be compared to the values of other regions in order to locate the strongest acoustic source.

The procedure is also applicable to multiple target localization. If the sensor arrays are able to provide several DOA estimates corresponding to several consensual directions, ISRR treats all these estimates equally and confines the search regions of two or more sources in the same manner as for the single source case. The only difference here lies in the specification of rules for choosing not one but several best fit regions. Furthermore, as sound pressure decreases exponentially with propagation, each sensor node group can identify a source closest to it. If the groups have overlapping FOV, several targets may be identified by different parts of these groups based on this principle.

Multilateration Approach to ISRR

Multilateration can be applied in ISRR as a TDOA based method for search region confinement, as apposed to the proposed DOA based approach. As it is discussed in Section 2.6.3, the distance between the sensor with coordinates (x_i, y_i, z_i) and the acoustic source is defined as a vector length (2.34). For the multi-sensor WSN ground applications the solution is simplified by constant z dimension. Thus, having a TDOA τ_{ij} between two nodes i and j , the distance difference between both sensors and the source (2.35) is also simplified to

$$d_{ij} = c \cdot \tau_{ij} = c(\tau_i - \tau_j) = \sqrt{(x_i - x)^2 + (y_i - y)^2} - \sqrt{(x_j - x)^2 + (y_j - y)^2}, \quad (4.28)$$

where d_{ij} is the distance difference estimate between sensors i and j , and (x_i, y_i) and (x_j, y_j) are the sensor respective coordinates [91]. If τ_{ij} is represented in terms of delay in samples Δk_{ij} with sampling frequency f_s , then the difference, similar to (4.3), is computed as $d_{ij} = \Delta k_{ij} / f_s \cdot c$. The delay τ_{ij} is calculated using cross-correlation (4.4), also applying the control metric (4.7). For any three separate sensors $\{S_1, S_2, S_3\}$ the acoustic source is localizable by the following system of equations:

$$\begin{cases} d_{12} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} - \sqrt{(x_2 - x)^2 + (y_2 - y)^2} \\ d_{13} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} - \sqrt{(x_3 - x)^2 + (y_3 - y)^2} \end{cases} \quad (4.29)$$

To estimate the solution to this system of nonlinear equations a numerical method called Trust-Region Dogleg [108] is applied.

Multiple sensor triplets are used in order to establish several triangles for multilateration, as shown in Figure 4.13. Every triplet gives a separate position estimate, and then all the estimates are partitioned by minimal distance by applying Algorithm 4.2 in order to get the reduced regions. The general direct multilateration solution in real-time WSN applications is solved with larger number of nodes [157], where the incorrectly placed regions or multiple sound sources are eliminated by feedback from the object tracking stage. However, the convergence rate of solving large systems of nonlinear equations tends to be lower than for simpler systems. Also the TDOA estimation error can disrupt the numerical optimization process for larger systems, which results in unreliable solutions. The utilization of only three sensors for the process thus simplifies and accelerates the solution estimation procedure.

4.3.3 Application to SRP-PHAT with SRC

SRP-PHAT with SRC can be applied as the last step of localization in wired synchronous systems. The proposed approach initializes SRP-PHAT on already contracted areas and is often applied more than once for a single signal frame (due to several contracted regions). The typical approach to SRC discussed in Section 2.6.2 suggests choosing fixed values for the number of contracted region defining points $N_i \equiv N = 100$ and stochastically evaluated points $J_0 = 3000$ for a FOV of approximately 20 m^2 , however this is not suitable for constantly varying initial search areas. For the application of SRP-PHAT with SRC to ISRR the parameters are rather estimated by linear functions. Building on the test results in [49] and considering peak estimation quality, the two functions for the task are derived as

$$\begin{aligned} J_0(s) &= \begin{cases} [297.6 \cdot s + 24], & s < 10, \\ 3000, & s \geq 10, \end{cases} \\ N(s) &= \begin{cases} [9.9 \cdot s + 1], & s < 10, \\ 100, & s \geq 10, \end{cases} \end{aligned} \quad (4.30)$$

where s is the area of the FOV in m^2 , and the bracket $[\cdot]$ denotes the operation of rounding to the nearest integer. The application of these functions optimizes the SRC process by greatly reducing the number of SRP evaluations for reduced regions of acoustic source search [23].

4.4 Experimental Evaluation

For the experimental installation Vansonic PVM-6052 condenser microphones are used. Four microphones are mounted into a plastic board casing with a

distance of $l = 15$ cm between each other, the width of a single linear array sub-block thus being equal to 45 cm. A total of 4 sub-arrays are used, which results in a wide-aperture 16-microphone array. For signal acquisition the Agilent U2354A DAQ is used with the sampling rate set to 8 kS/s per channel. The data is acquired to and processed in the MATLAB environment using the Data Acquisition Toolbox. Processing is performed in frames with a step of 0.2 seconds by conventional SRP-PHAT and by ISRR followed by SRP-PHAT with SRC on confined regions.

4.4.1 Array Regular Configuration

During initial testing the array sub-blocks are placed in three regular shape configurations: linear — the sub-arrays are placed in a straight line with $\alpha = \{0, 0, 0, 0\}$; angular — two sub-arrays are placed perpendicular to the remaining two with $\alpha = \{0, 0, 90, 90\}$; square — sub-arrays are placed at the edges of a square area with $\alpha = \{0, 90, 180, 270\}$. The experiments are performed in an office room with no protection against reverberation. The audio speaker reproducing human speech is placed in several spots of the wide-aperture array FOV, and minute long signals are acquired.

The localization results for these three configurations are presented in Figures 4.14–4.16. The top plots of these figures depict the results of search region confinement by ISRR denoted by rectangles and the point of maximal SRP estimated by SRP-PHAT with SRC on the confined region (black circles with SRP values to the right of them). The DOA based approach to ISRR is tested with the DOA estimate vectors denoted by blue lines and their intersection points — by purple stars. The SRP images computed over all 16 microphones by conventional SRP-PHAT with $\beta = 0.8$ are presented in the bottom plots of Figures 4.14–4.16 for reference. The figures show that the beams of each individual sub-array computed by SRP-PHAT coincide with the DOA estimated by the proposed procedure. The superposition of these beams forms the SRP images with the regions of elevated SRP energy resembling the search regions confined by ISRR.

The reduction in the number of performed SRP evaluations (2.31) and the initial search areas are presented in Table 4.4. SRP-PHAT with SRC manages to locate the global optimum in one iteration for all three configurations, however, the evaluation of the whole FOV leads to 3000 evaluations. ISRR manages to reduce the initial search area to a fraction of a square meter, thus significantly reducing the number of evaluations according to (4.30) for all experiments. The distribution of the mean quantity of evaluations is close to normal, which indicates a good level of computational reliability.

Considering the three tested configurations, it can be noted that the angular and square configurations provide more confined localized regions, compared to the linear configuration because the measurement points are

more widely distributed. Even the slightest error in DOA estimation in the linear configuration leads to the risk of sub-arrays not converging on the common search region. The square configuration performs better than the linear one, however, its localization quality suffers if the acoustic source is not omnidirectional. In this case part of the sub-array blocks inevitably receive reflected acoustic waves, which reduces their DOA estimation accuracy. As a result the square configuration is also more prone to reverberation influence. The angular configuration is more robust in terms of tolerance to reflection and reverberation. It is more practical as well because the FOV does not need to be surrounded by sensors.

Table 4.4: Results of region reduction for regular configurations

Configuration	SRP-PHAT with SRC		With ISRR	
	Mean J_0	FOV (m ²)	Mean J_0	Mean area (m ²)
Linear	3000	18.4	155	0.4412
Angular	3000	11.0	102	0.2632
Square	3000	16.0	117	0.3105

4.4.2 Array Irregular Configuration

For the localization experiment with an irregular sub-array configuration the array blocks are placed in a room as it is portrayed in Figure 4.17. The FOV is set to be 1 meter wider in every direction from the corner points of the array (approximately 18 m²). Sub-arrays S_1 , S_2 and S_3 form an angular configuration, while sub-array S_4 is diverted from the common direction of view, as if belonging to a different sensor group. The speaker takes three paths (returning to the starting point) while walking with an average pace of no more than 1 m/s. Each path is taken twice and in two manners: with the speaker moving constantly and the speaker moving and stopping regularly. This constitutes 12 experiments in total. For the DOA approach to ISRR all four sub-arrays with 16 microphones are used. For multilateration the microphone triplets are chosen for the system of equations (4.29) in the following manner to obtain triangles of various size: $\{m_1, m_4, m_{12}\}$, $\{m_1, m_4, m_{16}\}$, $\{m_5, m_8, m_{12}\}$, $\{m_5, m_8, m_{16}\}$, $\{m_1, m_8, m_{12}\}$, $\{m_1, m_8, m_{16}\}$, $\{m_4, m_5, m_9\}$, $\{m_4, m_5, m_{13}\}$.

The results of speaker localization at the beginning of path 1 and at the end of path 3 are presented in the top and bottom plots of Figure 4.18, respectively. In the figure blue lines denote sub-array estimated DOA, purple stars denote the intersections of DOA vectors, black rectangles denote the confined regions and black circles denote the SRP energy maxima with their values situated to the right. For multilateration the coordinate estimates are denoted by green diamonds, the confined regions are denoted by dotted rectangles and the SRP energy maxima — by black circles with values. In

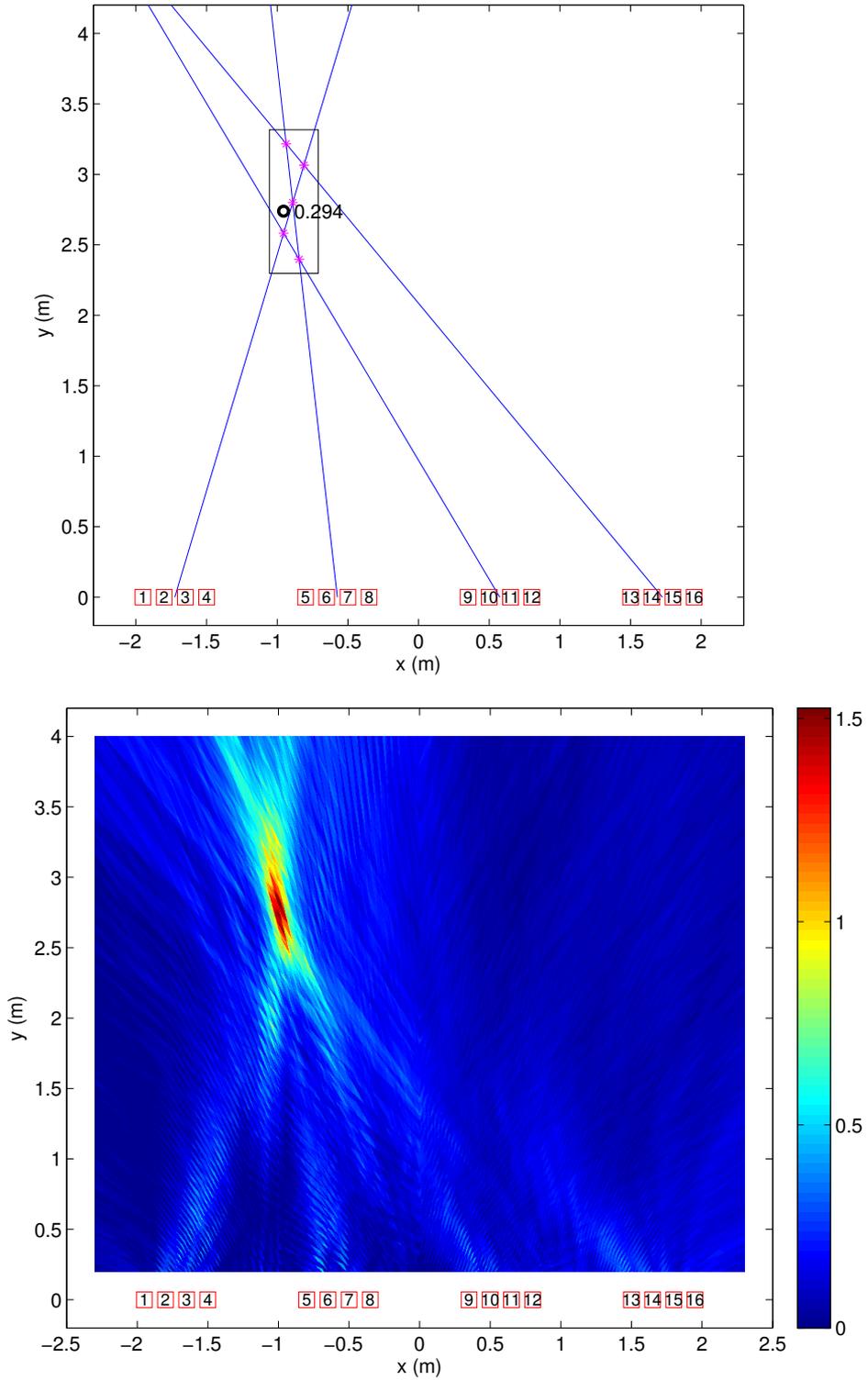


Figure 4.14: Localization by the linear configuration of arrays. Result of ISRR (top) and SRP-PHAT (bottom).

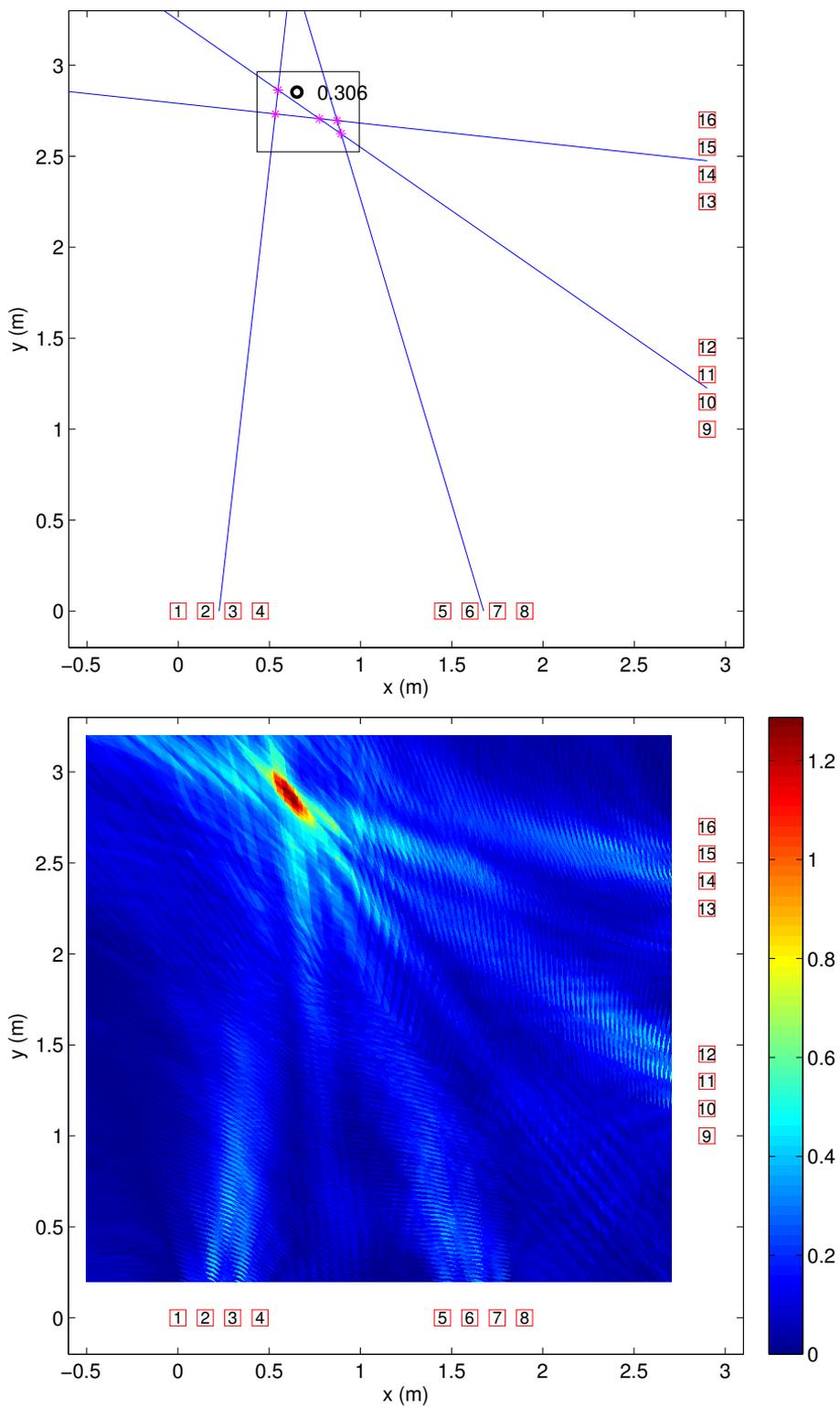


Figure 4.15: Localization by the angular configuration of arrays. Result of ISRR (top) and SRP-PHAT (bottom).

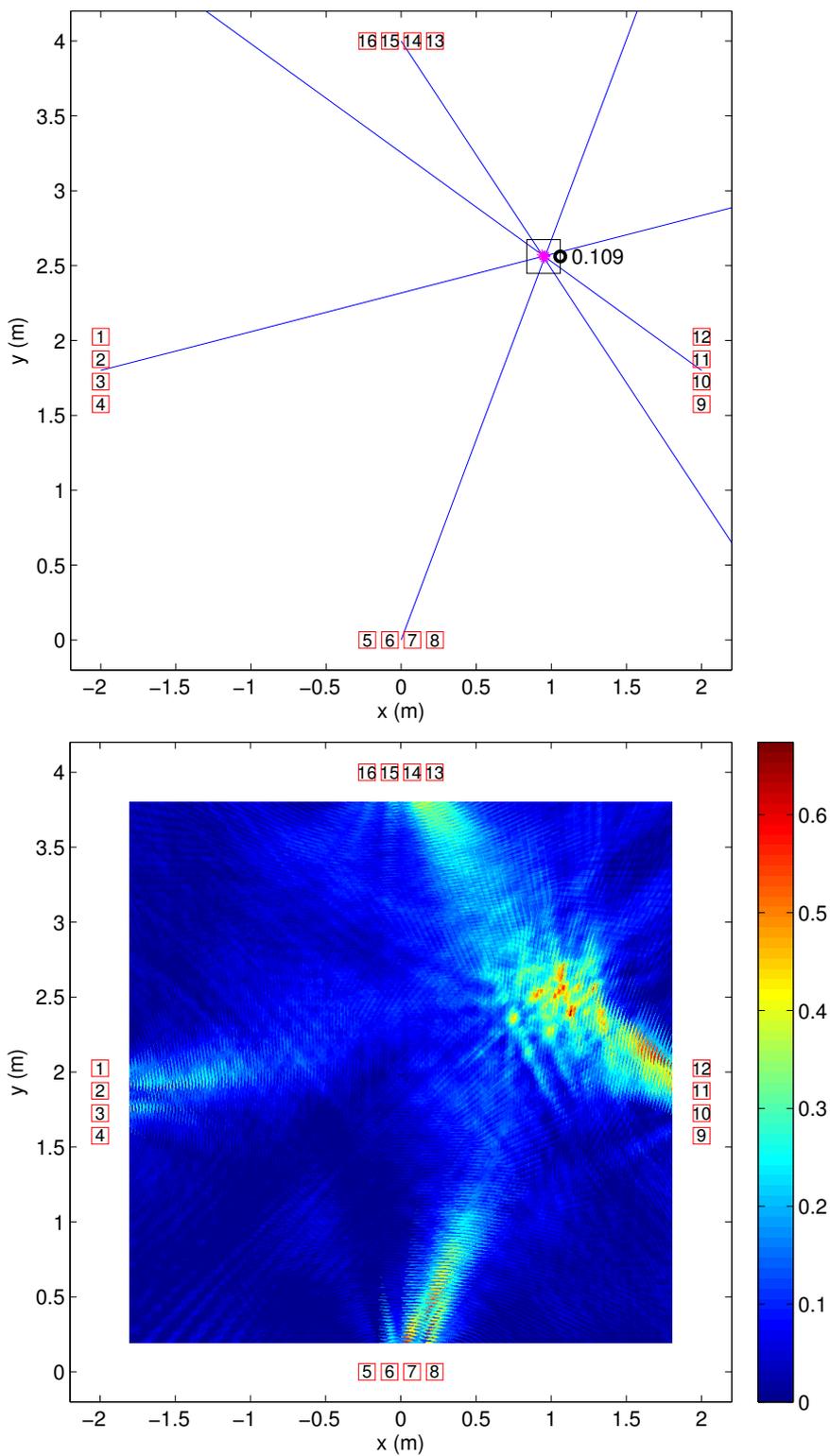


Figure 4.16: Localization by the square configuration of arrays. Result of ISRR (top) and SRP-PHAT (bottom).

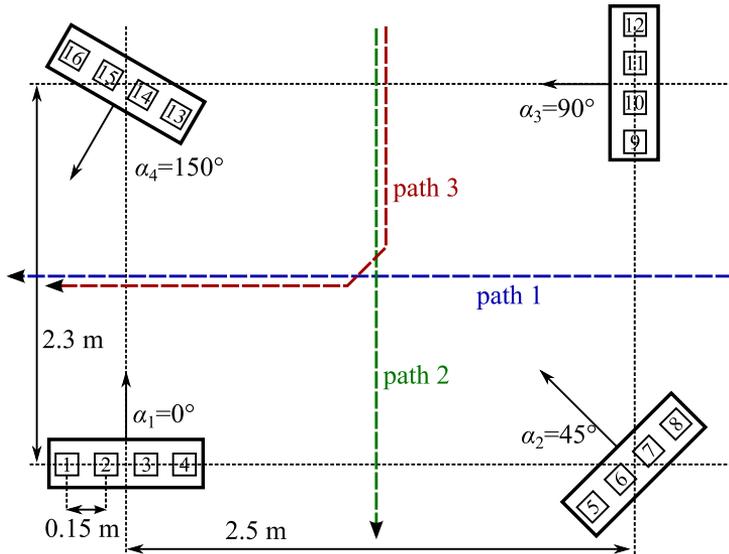


Figure 4.17: Irregular configuration experiment layout with four array blocks and one speaker taking three paths.

this experiment the two ISRR approaches based on DOA estimation and multilateration operate with approximately equal accuracy. Problems arise for both approaches in the region behind and between S_2 and S_3 (path 1), where neither S_2 or S_3 have a sufficient view of the source and S_1 and S_4 are excessively steered away from that position. This problematic region is depicted in the top plot of Figure 4.18. For S_4 the expected DOA totally exceeds the limits of S_4 frontal FOV. A slight advantage of multilateration is, however, evident due to its non-directional approach. The latter part of path 1 and both paths 2 and 3 are well traceable by both approaches. In the leftmost region of the FOV, where S_4 is also active, ISRR achieves the best results, as shown in the bottom plot of Figure 4.18.

Table 4.5: Results of region reduction for the irregular configuration with $\text{FOV} \simeq 18 \text{ m}^2$

Parameter	ISRR based on	
	DOA estimation	Multilateration
Mean area (m^2)	0.1374	0.0621
RMSE x (m)	0.1143	0.1227
RMSE y (m)	0.1107	0.1230

The impact of ISRR is substantial with the mean area being reduced from the 18 m^2 of the entire FOV range to a fraction of a square meter, as shown in Table 4.5. To estimate the divergence from the global SRP maximum estimated over the entire FOV, the difference between the result of SRP-PHAT

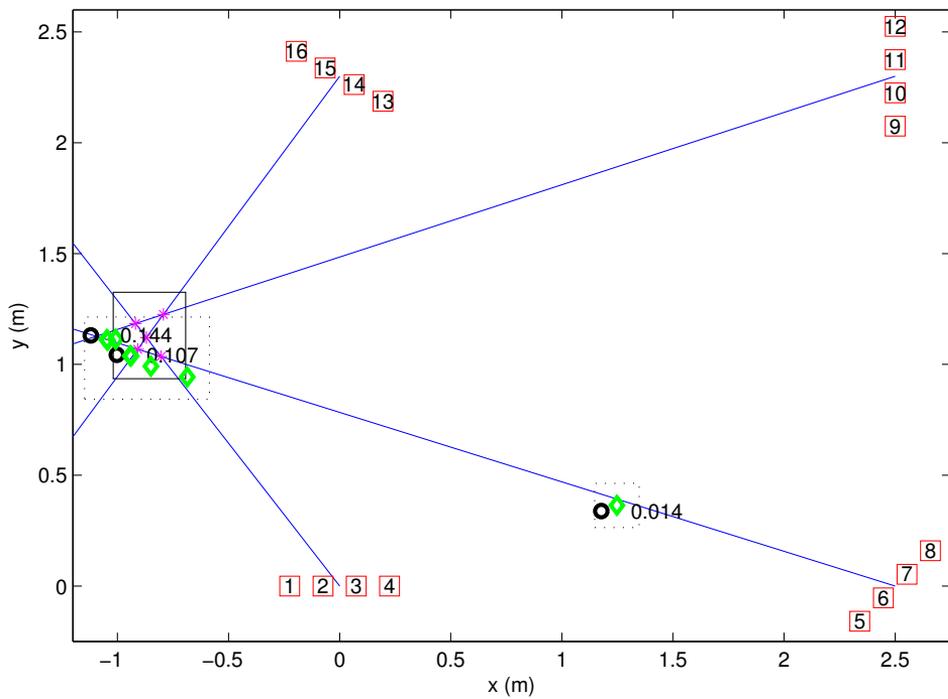
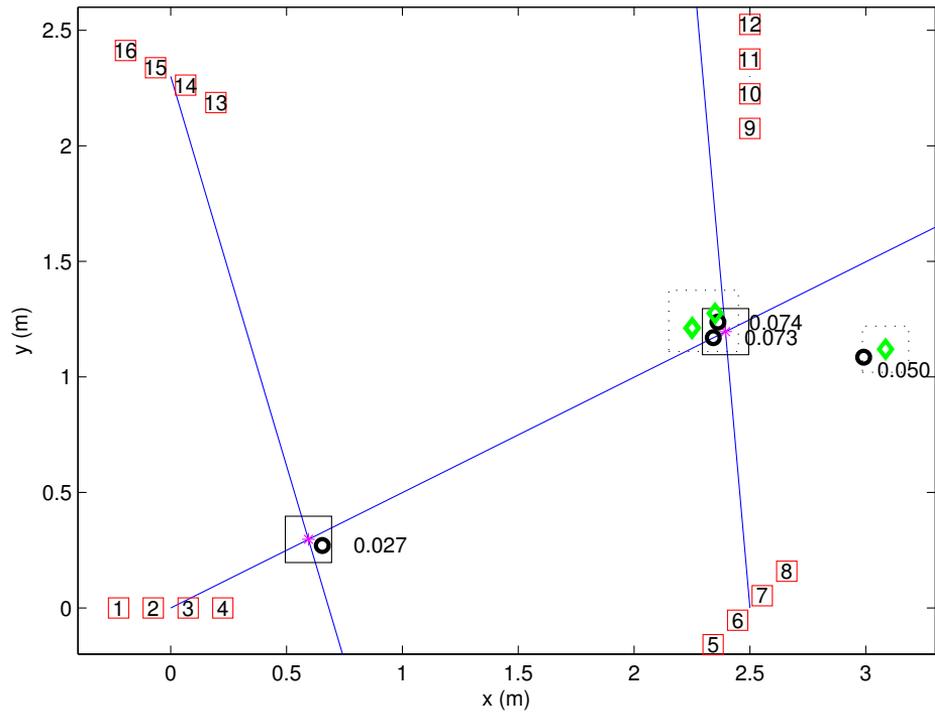


Figure 4.18: Results of speaker localization by ISRR at the beginning of path 1 (top) and at the end of path 3 (bottom).

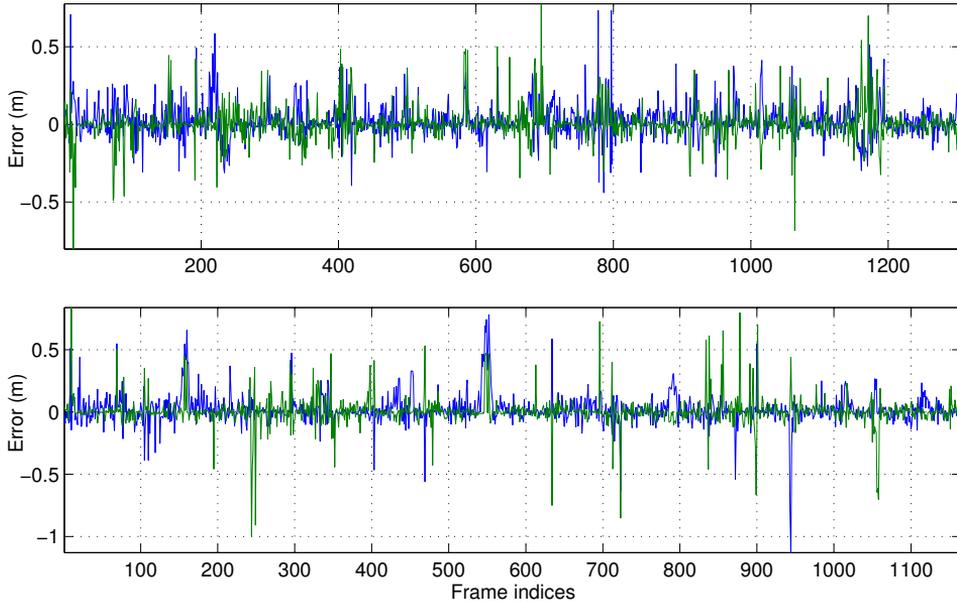


Figure 4.19: Difference in localization between SRP-PHAT over the entire FOV and the DOA based (upper) and multilateration based (lower) approaches to ISRR.

with SRC and the result of ISRR is calculated over all signals acquired for all three paths. Error variation in time is presented in Figure 4.19, where blue lines denote the x and green line — the y coordinate. The Root Mean Square Errors (RMSE) are presented separately for the x and y coordinates in Table 4.5. The x -axis in Figure 4.19 contains more values for the DOA based approach to ISRR if compared to the multilateration based approach. The multilateration based approach suffers from rare divergence at the stage of solving the system of nonlinear equations (4.29) at lower SNR. The DOA based approach also discards less frames due to data validation failure than multilateration. Also the RMSE is slightly lower for the DOA based approach. The overall errors are sufficiently low for speaker localization. Rare bursts of error do occur, however they are instantaneous and appear only during moments of speaker acceleration.

4.4.3 Computational Efficiency Assessment

To determine the increase in computational efficiency of the proposed DOA estimation and distributed localization methods the reduction in the number of cross-correlations required for computing SRP-PHAT and the proposed methods is quantified. The cross-correlation is chosen because it is the most resource demanding operation in both methods (and the majority of conventional phase based DOA estimation approaches, e.g, MUSIC). Other operations of the proposed methods, e.g., final AOA estimation of Algorithm 4.1

or point partitioning of Algorithm 4.2, are insignificant in terms of processing power demand, and are thus omitted from analysis.

Assessment of DOA Estimation for the Linear Array

For DOA estimation using a linear sensor array, SRP-PHAT with the reduced functional, discussed in Section 4.1.1, will calculate $n_h \cdot \binom{M}{2}$ cross-correlations, each with a single lag shift specified by the chosen evaluated direction (see Section 2.6.1). The proposed method, discussed in Section 4.1.2, will calculate δ cross-correlations, where $\delta = \sum \delta_{ij}$ is the total number of shifts required for calculating the cross-correlation (4.4) for all microphone pair (m_i, m_j) combinations. The number of integer lag shifts in the cross-correlation is equal to the number of possible discrete AOA (4.9) for the microphones m_i and m_j :

$$\delta_{ij} = 2\Delta k_{\max}(i, j) + 1, \quad (4.31)$$

where $\Delta k_{\max}(i, j)$ is obtained by applying (4.8).

Let us consider the experimental microphone implementation with $M = 4$ microphones and $l = 15$ cm between consecutive microphones. For $n_h = 500$ SRP-PHAT will calculate $500 \cdot \binom{4}{2} = 500 \cdot 6 = 3000$ cross-correlations. The proposed method will calculate three cross-correlations $\{m_1, m_2\}$, $\{m_2, m_3\}$ and $\{m_3, m_4\}$ for the inter-microphone distance $l = 15$ cm, two cross-correlations $\{m_1, m_3\}$ and $\{m_2, m_4\}$ for the distance $2l = 30$ cm, and one cross-correlation $\{m_1, m_4\}$ for the distance $3l = 45$ cm. By applying (4.8) to the given distances for $c = 343$ m/s and $f_s = 8$ kS/s, the total number of cross-correlations is equal to $\delta = 3 \cdot 7 + 2 \cdot 13 + 1 \cdot 21 = 68$.

Assessment of DOA Estimation for the Conical Array

For the estimation of azimuth and elevation AOA components of DOA using a conical sensor array, SRP-PHAT with the reduced functional, discussed in Section 4.2.2, will calculate $n_h \cdot \binom{M_h}{2} + n_v \cdot \binom{M_v}{2}$ cross-correlations, each with a single lag shift specified by the chosen evaluated direction. The proposed method will calculate $\delta = \delta_h \cdot |A_h| + \delta_v \cdot |A_v|$ cross-correlations, where δ_h is the total number of shifts required for calculating cross-correlations for all microphone pairs $(m_i, m_j) \subseteq A_h$, and δ_v is the total number of shifts for calculating cross-correlations for microphones $(m_i, m_j) \subseteq A_v$.

The experimental array consists of $M_h = 6$ and $M_v = 2$ microphones with UCA radius of $r = 7.5$ cm and vertical microphones elevated above UCA at $z_{m_j^v} = \{0.1, 0.2\}$ m. For the chosen $n_h = 500$, $n_v = 250$, the number of cross-correlations per each SRP-PHAT computation is then equal to $500 \cdot \binom{6}{2} + 250 \cdot \binom{2}{2} = (500 + 250) \cdot 15 = 11250$. The proposed method utilizes 19 pairs of microphones according to the definitions of A_h and A_v in (4.15),

(4.17): 6 consecutive pairs $m_i^h m_{i+1}^h$; 6 pairs over one microphone $m_i^h m_{i+2}^h$; 6 pairs between two vertical m_j^v and three horizontal m_i^h microphones; 1 pair of vertical microphones $m_i^v m_{i+1}^v$. Therefore, by applying (4.8) to the array geometry for $c = 343$ m/s and $f_s = 48$ kS/s, the total number of cross-correlations is equal to $\delta = 6 \cdot 21 + 6 \cdot 37 + 3 \cdot 35 + 3 \cdot 59 + 1 \cdot 29 = 659$.

Assessment of the Localization Procedure

The reduction in the number of computations is even more evident in case of the localization procedure over a wide-aperture array composed of several standalone array sub-blocks. Conventional SRP-PHAT and other similar localization methods, which employ all available microphone signals, will calculate the SRP (2.31) or other metric of signal phase coincidence for all $\binom{M}{2}$ pairs of microphones, where M is the total number of microphones in the wide-aperture array. It has to be also noted that SRP-PHAT will perform this number of computations per every discrete point in its spatial vector. Wideband MUSIC, on the other hand, will perform the computations separately for every analyzed frequency band. The distributed localization procedure will calculate \mathcal{N} signal phase coincidence metrics, where \mathcal{N} depends on the array structure. Assuming that identical sub-array blocks are used in the localization system, for the linear array, discussed in Section 4.1.1, this number will be $\mathcal{N} = K \cdot \binom{M_i}{2}$, where K is the number of individual array sub-blocks and M_i is the number of microphones in each i -th linear array sub-block. For the conical array, discussed in Section 4.2.2, the number of signal phase coincidence metrics will be $\mathcal{N} = K (|A_h| + |A_v|)$, where $|A_h|$ is the number of horizontal microphone pairs and $|A_v|$ is the number of vertical microphone pairs in each of K individual arrays.

For the experimental example of four linear sub-arrays comprising four microphones each, conventional localization methods, like SRP-PHAT, perform $\binom{16}{2} = 120$ evaluations (per every point in its spatial vector). Distributed localization performs $\mathcal{N} = 4 \cdot \binom{4}{2} = 24$ evaluations. If the same four sub-arrays consist of conical arrays with eight microphones, SRP-PHAT performs $\binom{32}{2} = 496$ evaluations. The proposed distributed localization, on the other hand, performs $\mathcal{N} = 4(12 + 7) = 76$ evaluations. It is easy to notice that the number of evaluations for conventional methods grows exponentially to the number of microphones used in the entire localization system. Therefore, expansion of the localization system yields an even greater reduction in the number of signal phase coincidence metric computations. Furthermore, each evaluation consists of a previously discussed number of cross-correlations, which further increases the overall number of computations. The evaluation of the total number of cross-correlation computations in the entire localization system is left for the reader.

4.5 Implementation on Embedded Hardware

A total of three array prototypes were implemented on embedded hardware for DOA estimation and distributed acoustic localization procedure field testing. For the linear array configuration two prototypes were developed. The first prototype was implemented on a small low-power platform, equipped with a Atmel ATmega128RFA1 microprocessor. Testing on this platform gives the impression of the proposed method productivity on CPS smart dust motes. The second linear array prototype was implemented on the BeagleBone Black (BBB), which provides significantly more computational resources, as well as allowing for sampling at higher rates than the Atmel based solution. One implementation of the UCA part of the conical array was developed for testing the planar localization procedure. It is also based on the BBB platform. Several units per each implementation were built and tested during the application-specific experiments, discussed in Chapter 6.

4.5.1 Mote Implementation of the Linear Array

For the implementation of a two-microphone linear array on smart dust motes a platform developed by Defendec Inc. is used. The motes are equipped with Atmel ATmega128RFA1 microprocessors, which provide an on-chip ADC for signal acquisition and a radio transceiver for WSN communication [12]. The microprocessor has a clock speed of 16 MHz and provides 16 kB of SRAM memory for operation with an additional 128 kB of flash memory for program code. The on-chip ADC has a resolution of 10 bits and is able to sample with rates up to 330 kS/s. However, actual experiments were initially carried out with a sampling rate of 2 kS/s (and later increased to 4 kS/s) for each microphone channel, since higher sampling rates provided inconsistent and erroneous results during data acquisition. Through testing it was established that erroneous results were caused by signal leakage from the previous ADC channel to the succeeding channel when switching between channels, but the cause of the leakage could not be determined. Mote-to-mote communication was implemented through the IEEE standard 802.15.4 compliant radio transceiver with an effective indoor communication range of approximately 30 meters. The IEEE 802.15.4 standard supports transfer rates up to 250 kbit/s with packet sizes not larger than 127 bytes. Two Vansonic PVM-6052 electret condenser microphones were used for acoustic signal acquisition with additional circuitry performing signal amplification and the normalization needed for the microprocessor ADC input. For every mote a pair of microphones was mounted facing the same direction in an aluminum housing, which was then attached to the mote's plastic chassis (see Figure 4.23).

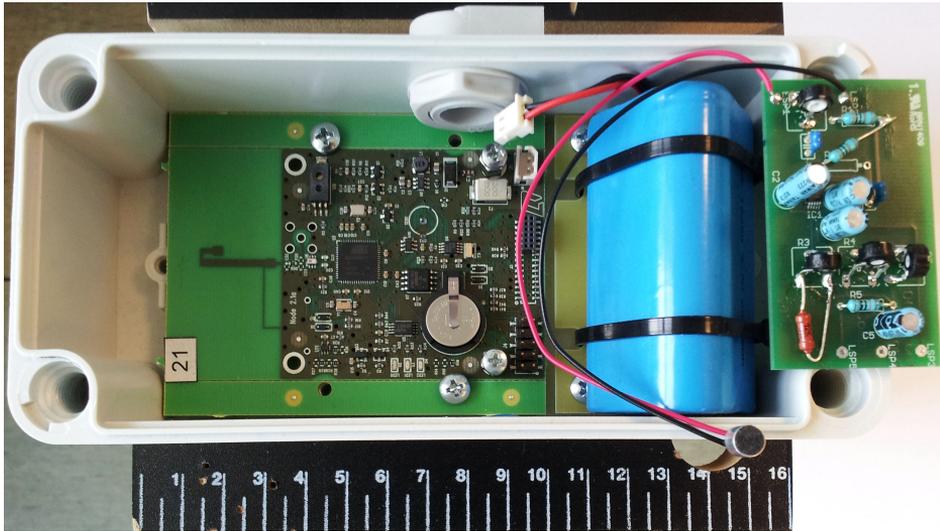


Figure 4.20: Packaged WSN mote with a sensor amplification circuit (scale in cm).

The developed smart sensor mote is presented in Figure 4.20. Microphone amplification circuitry is situated on the right and the microphone itself is shown in the bottom right corner of the figure. The mote is powered by a 3.7 V, 6600 mAh battery block (left from the sensor circuit). The motes are packaged in protective frames 16 cm in length. The poor computational characteristics listed above are typical for smart sensor motes. The reason for this is that these motes must work ubiquitously and autonomously with the battery they are provided for as long as possible. For example, the battery used in configuration at hand can sustain the motes for 1–1.5 years in a low duty cycle mode and approximately a month in constant operation mode. The goal here is to show that if localization and ISRR can be carried out on a smart sensor mote network, it is reasonable to assume it can also be implemented on larger networks with computationally more powerful motes.

Aspects of Operation at Low Sampling Rates

As it was discussed in Section 4.1.3, applying low sampling rates for signal acquisition introduces a number of challenges for smart sensor design. In this case a compromise between a permissible the inter-microphone distance l and the number of discrete AOA n_{AOA} has to be established. A distance of $l = 0.7$ m is chosen for the motes, which gives $n_{AOA} = 9$ possible AOA values with an average step of 19.7° at the applied rate of $f_s = 2$ kS/s. These are calculated using (4.3) and presented in Figure 4.21. The substantial difference with high sampling rates is evident from the figure. For the same sensor distance at a rate of $f_s = 44.1$ kS/s the AOA number is equal to 179 with an average step of 0.92° . A small n_{AOA} introduces additional error

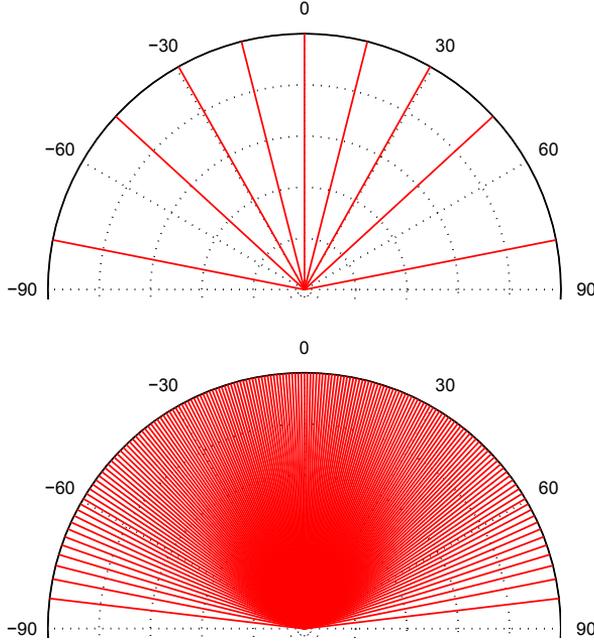


Figure 4.21: Possible AOA values for a sensor pair with $l = 0.7$ m and sampling rate set at $f_s = 2$ kS/s (top), and $f_s = 44.1$ kS/s (bottom).

into the localization process as the ISRR estimated regions may become larger and get shifted from the true area occupied by the sound source. For example an angle step of 19.7° can give an error of 1.8 m if the sound source is situated only 5 m away from the sensors. To manage the situation a large number of motes must be used, preferably steered by different angles α (i.e., not facing in exactly the same direction). Random mote placements allow AOA uncertainty regions to superimpose on one another thus reducing the discrete gaps. At the sampling rate of $f_s = 4$ kS/s the number of possible angles increases to $n_{AOA} = 17$, which significantly improves DOA estimation quality. Generally, the sensor array motes are applicable for estimation of object movement direction, e.g., from left to right or vice versa, and not for precise localization.

To demonstrate the effect of applying low sampling rates on localization quality, an experiment with a wired system of four microphone blocks with two microphones spaced at $l = 0.7$ m is performed similarly to the ones, discussed in Section 4.4. Data acquisition is performed at two sampling rates: $f_s = 8$ kS/s and $f_s = 2$ kS/s per sensor. The array sub-blocks are arranged in an angular configuration with two microphone pairs placed perpendicular to the other two. The speaker is placed at position $(0.7, 2)$ meters and a short speech recording is made. For conventional SRP-PHAT the area discretization value is set to 1 cm^2 .

The results of speaker localization for a signal frame of 200 ms are presented in Figure 4.22. At $f_s = 8$ kS/s both SRP-PHAT and ISRR localize the sound source efficiently. The SRP-PHAT region of particularly high cumulative energy (i.e., orange to red on the scale) is reduced to approximately 0.01 m^2 . The region estimated by ISRR is significantly larger, but proportionate to the SRP-PHAT region of medium cumulative energy (i.e., green on the scale). The sound source is fully confined by the ISRR region, as confirmed by the SRP-PHAT with SRC estimate on this region.

On the other hand, both methods suffer from the problems situated with the low sampling rate of $f_s = 2$ kHz, as it is shown in the lower plots of Figure 4.22. SRP-PHAT high cumulative energy region enlarges to approximately 0.25×0.3 meters with an incidental region situated in the top left corner of the FOV. The decrease of n_{AOA} and the number of signal samples per frame, described in Section 4.1.3, affects SRP-PHAT producing a more rough and less comprehensive image. Nevertheless, SRP-PHAT localizes the source properly. ISRR performs worse, missing the source slightly up the y -axis because the fourth microphone pair fails to estimate the DOA correctly. Although the confined region is close to the source position, it does not confine it fully. This example clearly shows the need for a larger number of microphone pairs (notes) to be used for successful localization.

Real-Time Operation Testing

Measuring the operation time on the embedded platform at hand is not a straightforward task. Due to the hardware limitations the measurement process itself consumes resources, which can affect the measurement accuracy. Furthermore, the microprocessor clock is slightly slower than a reference clock of a PC. Therefore, a direct and indirect approach is used for measurement.

The average processing time measurement results are presented for over 9000 measurements over a period of approximately 40 minutes of embedded system operation. The sampling rate is set to 4000 S/s with the frame length being equal to 400 samples per channel (i.e., frame duration of 100 ms). One full operation cycle (FOC) of the system includes: sampling of values (SV); processing of values (PV), i.e., pre-processing and DOA estimation; other functionality (message transmission, etc.) and various inbuilt system tasks.

For the indirect approach measuring the processing time with computer clock results in an average full operation cycle of $FOC = 282.1$ ms. PV measurement with an oscilloscope results in an average time of data processing of $PV = 169.1$ ms. As the theoretical sampling interval is equal to $SV = 400/4000 = 0.1$ s, or 100 ms, other functionality and system tasks require approximately 13 ms.

The direct approach of measuring the operation time of PV within the embedded system itself using the system clock gives an average of $PV = 165$ ms.

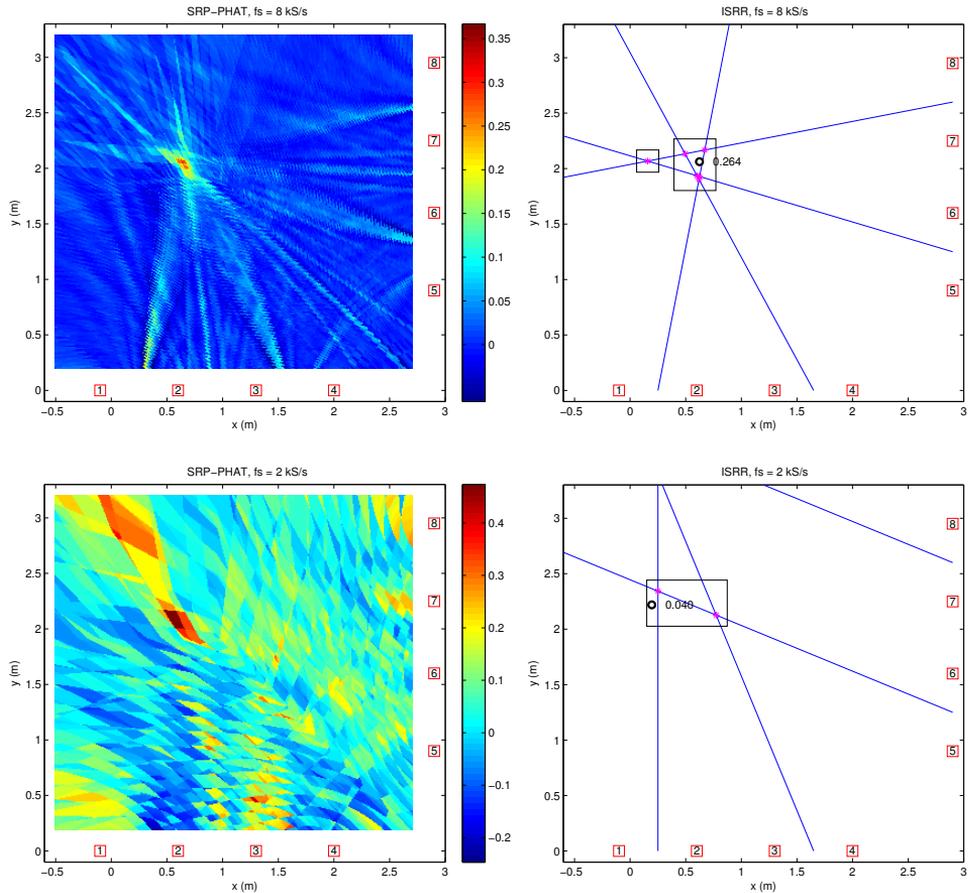


Figure 4.22: Acoustic localization results for four pairs of microphones applying conventional SRP-PHAT and ISRR.

It is normal to have a smaller value here, because the embedded platform clock is slightly slower than the PC clock. Based on the microprocessor's slower clock the total time for sampling of values is slower when calculated using the time concept of the embedded platform: $SV = 400 \cdot 256/1000 = 102.4$ ms. Here the coefficient $256/1000$ is the closest approximation to the desired sampling rate of 4000 S/s. It is, therefore, not possible to use the exact sampling rate of 4000 S/s on the embedded system due to the limitations of microprocessor clock ADC operation. Ignoring various inbuilt system tasks the direct approach gives an average full operation cycle estimate of $FOC = 267.4$ ms, which is almost three times slower than the desired operation period of 100 ms.

On low power embedded hardware, where the tasks of signal acquisition and data processing cannot be performed in parallel, it is impossible to strictly operate in real-time. Even if signal acquisition is performed in

a timely manner, the microprocessor needs to be given sufficient time to process the data. During this time the ADC remains idle. However, for operation in soft real-time, where the processing delay of several frames is permissible, and the consistency of signal frames is not required, this type of solution is perfectly applicable. On the other hand, if the signal needs to be assembled from several successive frames for further processing, and thus sample loss is not allowed, a more powerful embedded platform needs to be used for implementation.

4.5.2 Implementation on the BeagleBone Platform

A more powerful implementation of the microphone arrays is built on the BeagleBone Black platform. The system comprises a BeagleBone Black (BBB) development board, a power bank, and a proprietary standalone communication module, called MURP module. BBB features two programmable real-time units (PRU) with 32 bit RISC processors and an 8-channel 12 bit ADC. This enables the BBB to be used as both a DAQ and processing unit, sampling the data from 6 channels at $f_s = 20$ kS/s separately from the BBB non-real time operating system. The PRUIO library used for controlling the ADC has been developed under LGPLv2 license by Thomas Freiherr. The samples produced by PRU are written into a circular memory buffer implemented by the PRUIO library. A circular buffer is used in order to guarantee continuous online signal processing. The binary raw data is also stored on an external SD memory card for later analysis. The sampled data is then fed frame by frame to other software modules, which perform DOA estimation and other system tasks. A stable configuration with a frame length of 2730 samples per channel is experimentally established, which constitutes a refresh rate of 136.5 ms at the sampling frequency of $f_s = 20$ kS/s. The MURP module has its own Atmel Atmega256RFR2 chip and IEEE 802.15.4 compliant radio transceiver. This module is used for triggering the concurrent start of signal sampling within the sensor node group, as well as for soft synchronization of the WSN node clocks.

Linear and Circular Array Implementations

A linear array implementation on the BBB is presented in Figure 4.23. The array consists of six ADMP401 MEMS microphones (Pololu Corp., USA) enclosed in an aluminum housing with a distance of $l = 10$ cm between consecutive microphones. The total array length is thus equal to 50 cm. The maximum number of possible discrete angles for the sampling rate of $f_s = 20$ kS/s is equal to $n_{AOA} = 59$, which allows for far superior DOA estimation quality compared to the mote implementation.



Figure 4.23: Linear array implementations on the BeagleBone Black (left) and on the mote system with Amel microcontroller (right).

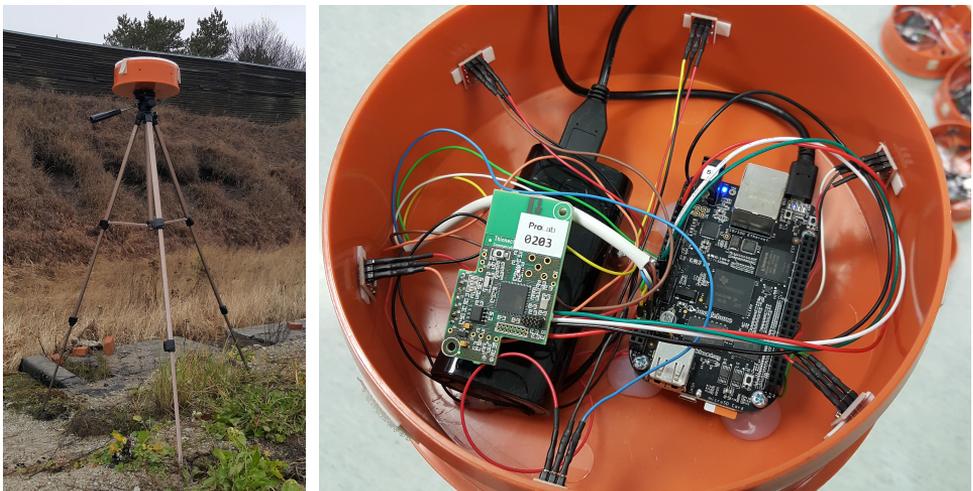


Figure 4.24: UCA implementation on the BeagleBone Black (left) and its inner components inside its plastic housing (right).

The UCA implementation is composed of an enclosed plastic circular shell with the radius of $r = 10$ cm and six ADMP401 MEMS microphones positioned at an angle between two successive microphones relative to the array center of $\alpha = \frac{\pi}{3}$. This implementation is presented in Figure 4.24. The inner components comprising the BBB, the power bank, the MURP module and the MEMS microphones are also presented in the figure.

Real-Time Operation Testing

The processing time of DOA estimation is measured directly on the BBB. The experiment was performed for the linear array configuration invoking the proposed DOA estimation procedure discussed in Section 4.1.2. The linear array is chosen for testing on the platform at hand because it has a larger maximum inter-microphone distance than UCA. The measurement was conducted for the duration of 77500 frames. The system was operating in an ordinary mode with various functionality and inbuilt system tasks enabled. The results of processing time measurement are presented in Figure 4.25. The average measured processing time of the DOA estimation procedure is equal to 80.72 ms, which is significantly less than the frame duration of 136.5 ms. Figure 4.25 shows that the distribution of the measured processing time is close to normal around the mean value, however, a small portion of measurements is detached from the main group at approximately 120 ms. This small outlying group likely results from problems situated with system specific process management. Although these measured times are close to the frame duration, they are sufficiently shorter than the frame duration and their quantity is insignificant compared to the total number of measurements. Furthermore, during the experiment no sample loss was registered. Thus the implementation at hand is shown to operate in real-time.

4.6 Conclusions

The presented approaches to planar and volumetric DOA estimation are reviewed for a single source case, where the maximum of the signal phase coincidence metric (i.e., signal cross-correlation) determines the AOA. For a multiple source case several maxima of this metric need to be chosen to estimate multiple AOA. The task then lies in the estimation of the number of active acoustic sources at any given moment. Blind signal separation is used as a phase-invariant approach for separating uncorrelated or slightly correlated signal mixtures. In this case PCA or ICA is applied to the signal mixture to estimate the number of minimally correlated components that indicates the number of sources [28, 33]. TDOA based approaches applying, e.g., matching pursuit [114] or random finite sets [161], utilize the phase

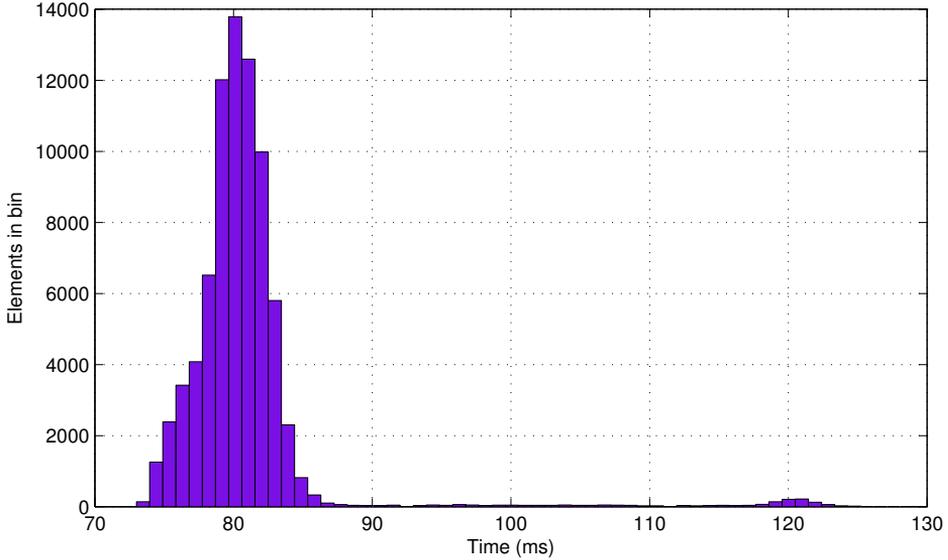


Figure 4.25: Processing time histogram of DOA estimation on the BeagleBone Black for the frame length of 2730 samples and frame duration of 136.5 ms.

information from spatially distributed sensors. These approaches mostly target human speaker separation, where the signal dynamical properties of the speech finite state non-stationary process significantly confine the problem scope. In a general case acoustic sources of various nature cannot be guaranteed to be separable. Furthermore, from the standpoint of acoustic localization if the sources are closely positioned, successful signal separation does not guarantee that the sources will be localized to separate regions due to the limitations of DOA estimation. Persistent acoustic sources are also separable through object tracking and trajectory estimation. For a WSN of multi-channel acoustic sensors, however, the most prominent solution lies in the partitioning of the overall FOV into smaller local regions, where the individual sources are localizable by sensor node sub-groups.

The proposed planar localization procedure is intended for ground object localization, however, it can also be applied to cases where certain amount of elevation also comes into play but does not need to be estimated. As it is shown in Figure 4.6, the direction to an elevated source can be estimated by only the planar AOA without using a volumetric array configuration. Furthermore, if the scale of the overall localization system is significantly larger than the localized object height of elevation, the vertical component is negligible. The proposed method, however, can be expanded to operate in the volumetric space by adding elevation AOA estimation and non-zero z coordinate component to the ISRR computations.

Generally the proposed localization procedure is not intended to operate on acoustic sensor arrays of identical configurations and DOA estimation procedures. On the contrary, the main advantage of a distributed multi-node CPS comes from the diversity of its sensor components. Different array configurations are applied to satisfy specific FOV requirements, and different DOA estimation methods used in conjunction can complement and reaffirm individual results. Though, even the computationally simplified proposed DOA estimation methods consume the majority of computational resources of even a powerful embedded platform like BBB, the low cost of individual CPS components may be sacrificed in order to ensure improved accuracy and robustness for high precision tasks.

Chapter 5

Data Fusion and WSN Communication

The chapter discusses the process of data fusion and reviews application-specific examples of fusion for the previously discussed solutions of process monitoring and object localization. A proposed approach to localized object trajectory estimation is also presented [11]. The second part of the chapter is devoted to the problems situated with WSN communication between individual nodes of CPS. The topics of node synchronization, communication delays and data validation are discussed, and a specific solution to network management is reviewed.

5.1 Data Fusion Aspects

Data fusion in its broad definition is the combination of separate knowledge gained by different system components for situation assessment, estimation of impact that the situation causes, determination and prediction of the required system state and reaction performed in response to the changing situation [6, 88]. Data fusion is performed from the bottom (sensor) components of the system upwards to information aggregation and decision-making components. Human interaction with the system (if any) is performed at the top levels of data fusion [64]. In the past decades several models of data fusion were proposed. One of the most popular is the Joint Directors of Laboratories (JDL) model [88], established in the early 1990s for military and defense purposes [75]. It proposes five levels of data fusion. These grow in complexity from level 1, which performs object detection, identification and tracking, to level 5, which handles system process refinement and resource management for system response capability estimation, governs the lower fusion processes, and involves human operators if needed. A zeroth level is also considered in the model, which performs sensor level data fusion, i.e., inference and ba-

sic decision making over processed data during signal processing. Decisions made on the top levels are addressed back to the bottom levels for system component reconfiguration and designated reaction execution. Alternatives to the JDL model include Dasarathy's functional model [45], the Omnibus process model [27], the Boyd decision (OODA) loop [144], and the Transformation of Requirements for the Information Process (TRIP) model [76]. All these models were developed mainly for military and defense purposes due to the fact that correct time-critical data management and decision making is most crucial in this area, but are applicable in other areas as well. The models abide by the principal of hierarchical data fusion process for decision making, and differ only in the aspect of information and decision flows between the specific hierarchical levels.

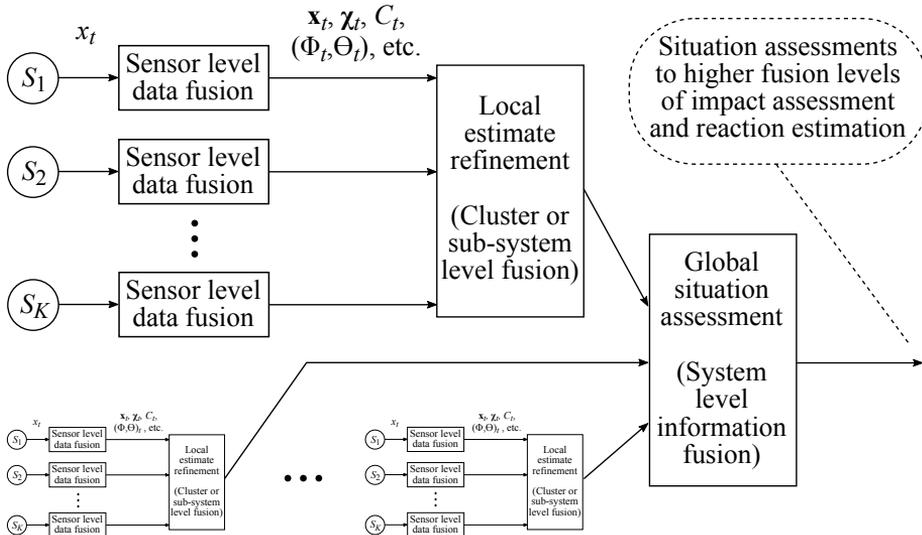


Figure 5.1: Considered hierarchy of data fusion.

5.1.1 Considered Tasks of Data Fusion

This work considers the initial data fusion tasks of inference on individual smart sensors (level 0) and initial situation assessment in the CPS sub-systems (level 1). Data fusion between sensors of different modalities and sensor sub-systems for global situation assessment (level 2) is not handled. The so-called autonomous data fusion architecture [64], presented in Figure 5.1, is adopted. This architecture assumes higher level inference being performed on individual sensor components with local estimates (results of on-sensor detection, identification, bearing to object, etc.) being used in further fusion on higher level components. This architecture opposes the centralized approach, where raw signals are used by the higher level fusion components,

and no inference is performed on the sensors. For WSN applications the centralized architecture is not applicable due to complications imposed by raw signal transmission [119]. As portrayed in Figure 5.1, at any time instance t each sensor node performs the appointed tasks of pattern recognition (consisting of signal pre-processing, feature extraction and classification), estimation of bearing to object, etc., and transmits the intermediate situation assessment data, e.g., its spatial coordinates \mathbf{x}_t , representative features of the observed process χ_t , estimates of process state or object class C_t , the set of bearing angles (Φ_t, Θ_t) and any other relevant data to the data aggregation and fusion WSN nodes, which govern the local cluster or sub-system of sensor nodes. There level 1 fusion ensues and results in local situation estimates, e.g., object localized region, its class, movement trajectory, etc. These local assessments are forwarded to the higher fusion components, which handle global situation assessment, impact estimation and other fusion tasks.

Level 1 fusion performed on data aggregation nodes handles initial estimate data provided by the sensor as a time series, which is evaluated in order to identify short-term faults in the estimates and to produce a most likely situation estimate. The process is very similar to pattern recognition, employing maximum likelihood criteria [134], least square optimization [94], Bayesian estimation [88], Markov chains [78], etc. The errors situated with initial estimate generation, time of estimate generation and measurement point location are modeled as process noise in linear (2.36) or non-linear system state models and estimated by, e.g., recursive Bayesian estimators, like the Kalman filter or Particle filter [88,94]. Data fusion of different entities is performed by various means, including general mappers, like Artificial Neural Networks [113] and Fuzzy Inference Systems [72].

In the autonomous data fusion architecture of the distributed WSN components the fusion performed on sensor nodes can also include analysis of the consecutive initial estimate series. The estimate at time t may be appended depending on the properties of several previous estimates. Furthermore, if the refresh rate of the data fusion components is lower than the refresh rate of the sensor components, a series of estimates can be produced and rectified on the sensor node in a sliding window manner, and later sent to the fusion components in batches. This further reduces network bandwidth consumption, but increases system latency. The choice of process flow and system state refresh rate is purely application specific. Another important task of data fusion in a WSN of distributed components lies in data temporal alignment for concurrency in observed event separation and identification. Data communication through the multi-hop sensor network induces transport delays, which have to be estimated [121]. Inter-node synchronization errors and the time delay of event detection between the spatially distributed measurement points also disrupt data concurrency. Therefore, both spatial and temporal data evaluation needs to be performed during fusion [116,118].

5.1.2 Discussed Data Fusion Methods

The methods discussed in this work can be considered level 0 and level 1 data fusion according to the JDL model. Level 0 fusion includes the inference procedures of pattern recognition (feature extraction and classification), discussed in Section 3.1, the multistage decision making logic of the multistage procedure, presented in Section 3.2.2, and the Direction of Arrival (DOA) estimation methods, discussed in Chapter 4. A level 0 fusion method of moving object detection and speed estimation for the Passive Infrared sensor is presented in Section 6.3. Level 1 fusion methods include the distributed localization procedure, presented in Section 4.3, the trajectory estimation procedure, discussed in Section 2.7, and an application specific procedure of shooter acoustic localization, presented in Section 6.4.

5.2 Trajectory Estimation and Prediction

Moving object localization is complicated by the unknown dynamics of object movement. In case of static object localization the consistency of consecutive localization estimates indicates their correctness. A moving object appears at different positions for consecutive frames, and thus it cannot be assumed that these consecutive estimates are correct or belong to the same object. Trajectory estimation helps to establish the movement pattern and thus indicates the probable region of next localization estimates for future signal frames. Trajectory estimation also resolves several issues of the proposed localization approach, discussed in Section 4.3. For both the DOA and multilateration based approaches to Initial Search Region Reduction (ISRR) several confined search regions may be established at any time instance. In case of a wired system the extent of signal phase coincidence can be measured by SRP-PHAT with SRC per every confined region with the maximal values indicating the true position of the acoustic source. For operation in WSN this approach cannot be adopted, therefore, trajectory estimation is the most viable option. The acoustic source produces consistent regions, while the incidental regions chaotically appear and disappear in the FOV. This consistency of consecutive regions is determined by the object tracking procedure, and future region estimates are made by trajectory prediction.

5.2.1 Proposed Trajectory Estimation Approach

The Kalman filter (KF) constitutes the foundation of the proposed approach to trajectory estimation in WSN. This filter is optimal for a linear system with Gaussian measurement and process noise, as discussed in Section 2.7. As the adopted process model for describing object movement is linear (see Section 2.7.3), the application of KF should be sufficient for the object tracking

problem. However, the Rao-Blackwellized Particle Filter (RBPF) is applied for reference in order to establish the potential improvement in estimation quality. KF and PF are quite common filters applied for trajectory estimation in WSN in, e.g., [39, 148, 153, 155]. The main problem situated with recursive Bayesian estimators applied in a frame by frame manner is situated with the error occurring during rapid change in process dynamics, e.g., during direction change or rapid acceleration. In this case the inertia built up in the previous estimates impedes smooth trajectory estimation for several frames. The proposed method builds upon the filter estimates and forms search regions based on coordinate predictions to improve both localization and trajectory estimation.

For search region prediction at the discrete time instance k , the state estimate $\hat{\mathbf{x}}_k = (\hat{x}_k, \hat{y}_k, \hat{\dot{x}}_k, \hat{\dot{y}}_k)$ is used to estimate the search region for the next time step denoted by the lower $\mathbf{x}_k^l = (x_k^l, y_k^l)$ and upper $\mathbf{x}_k^u = (x_k^u, y_k^u)$ points of the region rectangle diagonal as

$$\begin{aligned} \begin{bmatrix} x_k^l \\ y_k^l \end{bmatrix} &= \begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} - (1 - \kappa) \cdot \Delta t \begin{bmatrix} \hat{\dot{x}}_k \\ \hat{\dot{y}}_k \end{bmatrix}, \\ \begin{bmatrix} x_k^u \\ y_k^u \end{bmatrix} &= \begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} + \kappa \cdot \Delta t \begin{bmatrix} \hat{\dot{x}}_k \\ \hat{\dot{y}}_k \end{bmatrix}, \end{aligned} \quad (5.1)$$

where $\Delta t = \frac{N}{f_s}$ is the refresh rate (i.e., the time interval between consecutive estimates), and $\kappa = [0, 1]$ is the measure of velocity influence on region shift. Its value depends on the localization system FOV and the expected velocity of the object. During the experiments κ is set at $\kappa = 1/3$ for both filtering approaches.

Before trajectory estimation starts, the acoustic source is initially detected by ISRR. In case of a wired system, source detection is performed by applying SRP-PHAT with SRC. If the SRP value of the confined region is sufficiently high, tracking and region prediction are initialized, and region prediction assists ISRR in search region estimation, reassuring ISRR estimates and identifying redundant regions. Here the localization estimate $\tilde{\mathbf{x}}_k^{\text{ISRR}}$ is defined by the point of the SRP maximum. In case of a WSN implementation, where SRP-PHAT estimation is not possible, the primary region can be established by the number of DOA vector intersection points, and the localization estimate $\tilde{\mathbf{x}}_k^{\text{ISRR}}$ can be defined as the centroid of this region, calculated by Algorithm 4.2.

Trajectory estimation begins once the acoustic source is detected and proceeds while the acoustic source region is consistent, i.e., remains in the FOV, sustains high SRP level or other fitness metric and shifts with reasonable velocity. When the source leaves the FOV, disappears, or instantly shifts to another location (which may be an indication of a separate source), tracking and region prediction stop and reinitialize once a source is located

again. During trajectory estimation, KF and RBPF are applied at every time step k to produce the estimates $\hat{\mathbf{x}}_k$. Region prediction proceeds with KF and RBPF estimates $\hat{\mathbf{x}}_k$ to calculate the search regions for the next iteration by using (5.1). The estimates $\tilde{\mathbf{x}}_k^{\text{ISRR}}$ are evaluated by regions $(\mathbf{x}_k^l, \mathbf{x}_k^u)$ predicted at the previous step $k - 1$, and $\hat{\mathbf{x}}_k$ are rectified by the ISRR localization estimates according to Algorithm 5.1. The final position estimate $\mathbf{y}_k = (x_k, y_k)$ for the process model (2.36) is calculated if Algorithm 5.1 passes $\hat{\mathbf{x}}_k$ evaluation. Otherwise the ISRR localization estimate is kept as the final trajectory point estimate. For final trajectory presentation the Rauch-Tung-Striebel (RTS) smoother is applied [26] in order to eliminate trajectory irregularities situated with tracking errors and object movement peculiarities, e.g., sudden stops, abrupt change of direction, pitching, etc.

Algorithm 5.1 Trajectory estimate evaluation

Require: $\tilde{\mathbf{x}}_k^{\text{ISRR}}, \hat{\mathbf{x}}_k, \mathbf{x}_{k-1}^l, \mathbf{x}_{k-1}^u$ for time step k
if $\tilde{\mathbf{x}}_k^{\text{ISRR}}$ belongs to predicted region $(\mathbf{x}_{k-1}^l, \mathbf{x}_{k-1}^u)$ **then**
 return $\hat{\mathbf{x}}_k$
end if
 calculate distance $d = \|\tilde{\mathbf{x}}_k^{\text{ISRR}}, \hat{\mathbf{x}}_k\|$
 if $d \leq \kappa \cdot \Delta t \|\hat{x}_k, \hat{y}_k\|$ **then**
 return $\hat{\mathbf{x}}_k$
 else
 return $f(\tilde{\mathbf{x}}_k^{\text{ISRR}}, \hat{\mathbf{x}}_k)$ ▷ e.g., $f = \tilde{\mathbf{x}}_k^{\text{ISRR}}, f = \frac{1}{2}(\tilde{\mathbf{x}}_k^{\text{ISRR}} + \hat{\mathbf{x}}_k \{1, 2\})$
 end if

The estimates supplied by KF and the localization procedure estimates reinforce each other according to Algorithm 5.1. If the object rapidly changes direction or acceleration, KF estimates will be skewed towards the direction of previous frames. Returning the final position estimate to the localized estimate rectifies the trajectory during such movement irregularities. If, on the other hand, the localization procedure fails to accurately estimate the search region for a number of successive frames, the previous dynamic is maintained by the KF estimates. The coefficient κ can be set equal to the one in (5.1) or independently.

5.2.2 Experimental Evaluation

Experimental evaluation of the proposed trajectory estimation approach is performed on the experimental installation presented in Section 4.4. Four sub-arrays, each consisting of four microphones, are placed in the configuration presented in Figure 4.17. The signals are sampled at 8 kS/s per channel and processed in frames with a step of 0.2 seconds. The speaker takes 3 paths

in two manners: moving constantly and moving and stopping regularly. The microphones chosen for the DOA and multilateration approaches to ISRR are listed in Section 4.4.2.

Table 5.1: Results of region reduction for ISRR, Kalman and Particle filters

ISRR method	Estimates made by	Mean area in (m ²)	RMSE x in (m)	RMSE y in (m)
DOA based	ISRR	0.1374	0.1143	0.1107
	KF	0.0780	0.0994	0.0874
	RBPF	0.0409	0.1014	0.0863
Multilateration	ISRR	0.0621	0.1227	0.1230
	KF	0.0857	0.1084	0.1043
	RBPF	0.0427	0.1083	0.1019

The results of trajectory estimation for the return on path 3 and finishing the direct pass of path 1 are presented in Figure 5.2 and Figure 5.3, respectively. In the figures blue lines denote the vectors of DOA, purple stars denote their intersection points, black rectangles are the search areas estimated by ISRR-DOA. For ISRR using the multilateration approach the green diamonds denote sensor triplet coordinate estimates, and the dotted boxes denote the estimated search regions. The green boxes denote the search regions (5.1) predicted using KF estimates, and the red boxes denote the regions derived from RBPF estimates. Applying Algorithm 5.1 to the ISRR-DOA and KF estimates, as well as the two search regions, results in the final position estimate denoted by the thick violet point. Tracking results of KF are also presented in Figure 5.2 and Figure 5.3, where the black line denotes the trajectory estimated by KF and the red dotted line denotes the final trajectory smoothed by RTS. The path is clearly visible in both figures as the speaker is returning to the starting point of path 3 and reaching the endpoint of path 1.

The effect of final estimate rectification by Algorithm 5.1 is evident from Figure 5.3. Due to the rapid deceleration of the acoustic source (speaker stopping to turn and walk back) the predicted region and KF trajectory keep to the movement inertia and tend forward (left in the figure). As the ISRR estimate no longer resides in the predicted region, it is decided to return the final estimate into the ISRR confined region. On the other hand, in Figure 5.2 the movement direction coincides with the ISRR region, and thus KF estimate is considered to be correct.

During the experiments it was impossible to exclude the naturally occurring trajectory error of manual movement, therefore, the difference between the reference and estimated trajectories cannot be used as an error metric. Rather, the error is calculate as the difference between the estimates of SRP-

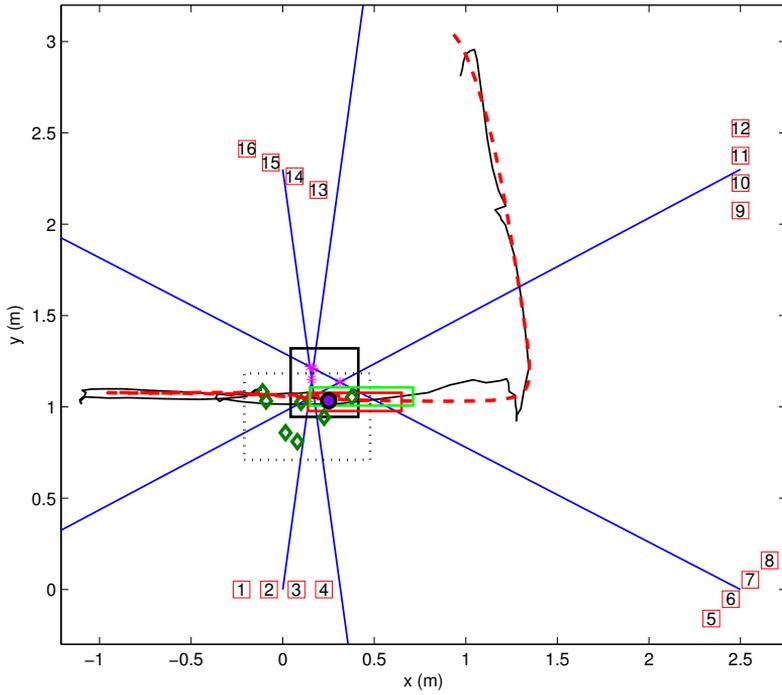


Figure 5.2: Localization and trajectory estimation results for path 3.

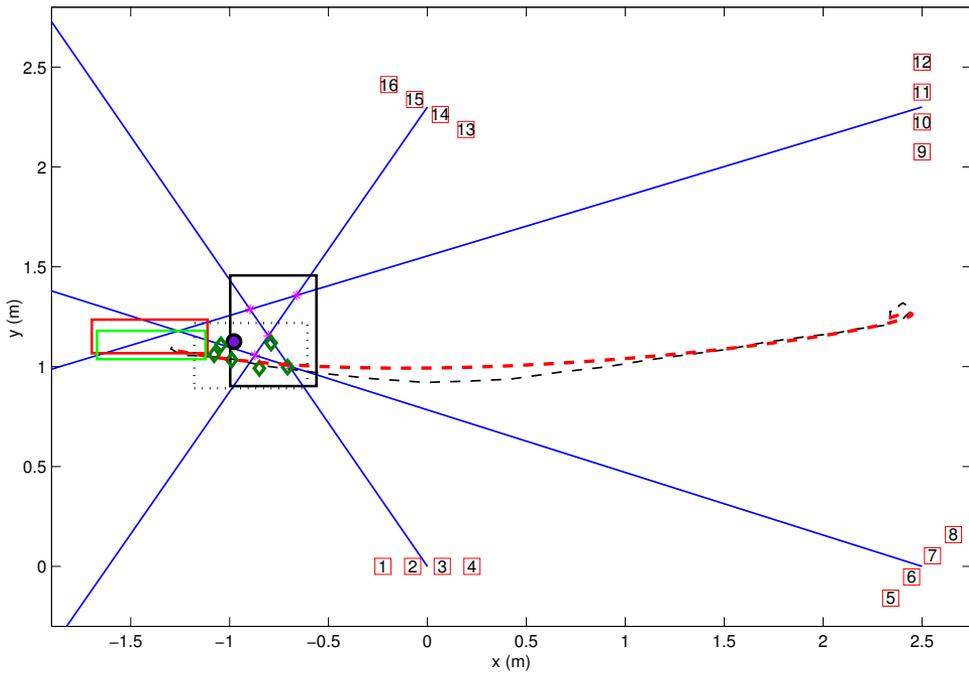


Figure 5.3: Localization and trajectory estimation results for path 1.

PHAT over the whole FOV and the estimates of the proposed trajectory estimation method. The results are presented in Table 5.1. Both mean area and root mean square error (RMSE) tend to decrease for both the DOA and multilateration based approaches to ISRR after applying trajectory estimation with KF and RBPF. As there is no significant difference between KF and RBPF performance, it can be stated that the linear process state system applied to object movement modeling is indeed sufficiently resolved by KF and no significant increase in quality comes from the non-linear properties of RBPF. Considering lower computational cost of the KF, it is therefore preferable in real-time WSN implementations.

5.3 WSN Communication Issues

The presented methods of distributed signal processing do not require sensor nodes to be synchronized. They do, however, require the events detected and identified by different sensors of various modalities and afterwards used in the fusion process to originate from the same physical phenomenon or process instance. For example, in single sensor systems, which produce individual situation assessments, if different nodes assess the observed process state at significantly different time instances, though their individual assessments are correct, the fusion process may fail because there will be no consensual information present in the separate assessments. For the stationary process of industrial machinery operation state this means that if the sensors assess the process during a state transition period and part of the sensors lags behind other sensors, their assessments will be conflicting and the true state will not be resolved during fusion until a steady state is reached and the majority of the sensors will provide the estimates of the same situation. For momentary processes, like vehicle pass, the desynchronized sensors may provide identification estimates for different pass instances and thus the conflicting vehicle type estimates will not result in a mutual decision. The problem is more evident for smart sensors cooperating in the localization process. Though the proposed distributed localization approach does not apply inter-node signal analysis (where synchronization up to the sampling period is required), soft synchronization is required in order for all the separate DOA estimates belonging to the same time instance to be handled as such.

5.3.1 Local Time and Communication Delays

The considered WSN consists of a unspecified number of nodes, which are interconnected in a dynamic ad-hoc manner. The network is self-structuring with the nodes coming in and out of communication at any time instance. Node synchronization in a classical sense, i.e., individual node clock adjustment to the same tick, is both tremendously hard and ultimately useless for

this kind of communication scheme. If the network structure is not static, there is no need to precisely synchronize the nodes that are not guaranteed to communicate for a specified amount of time. Furthermore, different implementation of each node does not guarantee that the clocks can be synchronized at all due to the physical properties of individual hardware [156]. And even precise synchronization is achievable, simultaneous data transmission is not possible through the common RF medium. Therefore, synchronized data will still reach the fusion nodes with significant delays. The introduction of common global time into the system is not necessary for the same reasons — each node is not guaranteed to be able to keep track of the global time flow and the constant synchronization packet broadcasting consumes impermissible amount of traffic [119].

The solution to not having a global time reference consists of not using time instance markers at all. Each node operates in its local time with a common understanding of a single time unit (e.g., second or millisecond). The packets containing application specific data, be it situation assessment or DOA estimates, are marked with data age values. The data age is calculated as the time difference between the moment of signal acquisition and packet composition before transmission. The fusion node accepts the packets and understands the data age at the moment of packet transmission, however, packet communication through the multi-hop network can take significantly more time than signal processing itself. Therefore, individual node operation in real-time does not guarantee real-time operation of the whole WSN without proper network management and packet transport time estimation.

Another problem situated with not having a global time reference lies in the uncoordinated refresh rates of individual nodes. As every node operates in its local time, it cannot know when exactly to start sampling the signal frame in order for it to contain the same signal portion acquired by other nodes of the WSN. As a result the process can be perceived a little differently by different nodes. For moving object localization, for example, this will introduce additional localization error, as the object's instant position depends on the exact time of signal frame acquisition. In an asynchronous network exact time moment specification is unachievable, instead a simultaneity interval of time inaccuracy tolerance is specified. The situation assessments made during this simultaneity interval are considered to be related to the same time instance.

5.3.2 Data Validation

Due to the complications induced by the distributed asynchronous operation of the CPS components the data incoming from the WSN nodes needs to be validated both spatially and temporally during data aggregation and fusion [121]. Spatial validation ensures that the data is received from the correct

cluster of sensor nodes governed by the specific fusion node, which performs data fusion on this cluster. Temporal validation is required to establish the correct sequence, age and simultaneity of incoming data. Performing data validation prior to the fusion process reduces the computational cost of fusion, because unwanted (incoming from a different cluster), outdated or non-sequential (belonging to a time instance already processed) data is discarded early on.

A generalized example of distributed sensor communication is presented in Figure 5.4. Black rectangles on the sensor timelines ($t_{S_1}, t_{S_2}, t_{S_3}$) represent the duration of signal processing on each sensor node, and arrows indicate the transport times. The packets reach the fusion node at times $T_{S_1}, T_{S_2}, T_{S_3}$ (on timeline t_{FN}). The fusion node estimates the total transport and processing times of each packet incoming from S_1-S_3 , denoted by rectangles T_{proc} and $T_{transport}$. The event detection times are then aligned to the estimated time moments $\hat{T}_{S_1}, \hat{T}_{S_2}, \hat{T}_{S_3}$, and compared against the simultaneity interval $I_{simultaneity}$ (on timeline \hat{t}_{FN}). Figure 5.4 illustrates that the processing time and packet transport time for each sensor node may be different, which results in the packets arriving out of order and too far apart to be included in the same simultaneity interval. Therefore, without proper temporal validation event concurrency is not guaranteed to be established. Processing delays originating from sensor platform specifics, clock jitters and drifts, ad-hoc WSN transmission scheduling, limited bandwidth and packet collisions within the network — all can unexpectedly disrupt smooth WSN communication. Proper network management is required to ensure real-time operation of the system as a whole.

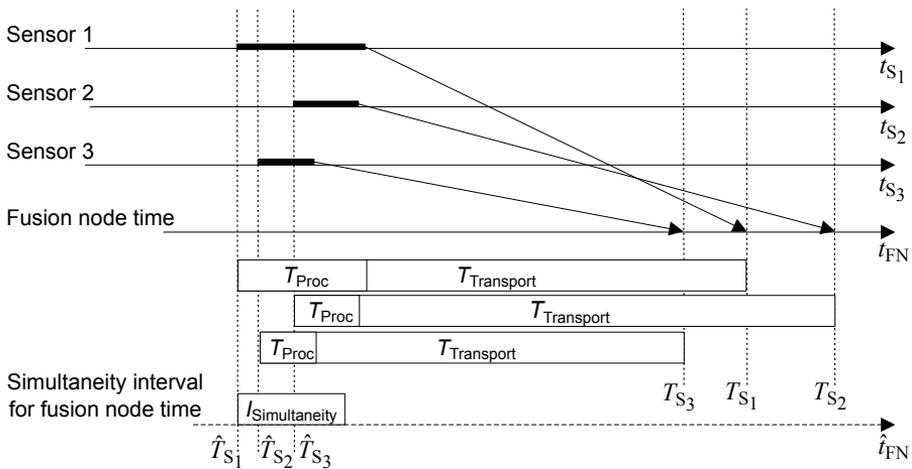


Figure 5.4: Temporal alignment and the simultaneity interval.

5.3.3 Communication Solution

In this work a communication solution proposed in [116, 121] is utilized. This solution is developed by the Research Laboratory for Proactive Technologies of the Tallinn University of Technology and is not included as an approach proposed by the author of the thesis, but rather as an example of WSN network management. The solution consists of a proactive middleware component called ProWare, which is specifically designed to meet the challenges of the dynamic ad-hoc topology and to provide network nodes with the capability of checking the validity of communicated data, while dynamically organizing the network structure at system run-time. Communication between nodes is established by a standalone communication module (proprietary transceiver), which comprises software components referred to as ProWare, and a hardware platform referred to as MURP. The general principle of inter-nodes connection is presented in Figure 5.5. Each node is connected to a MURP module over a serial UART port. The module is programmed with ProWare software and is equipped with a IEEE 802.15.4 compliant radio transceiver to establish physical links between nodes. All necessary network services, such as run-time sensor discovery, end-to-end data transfer timing and validity checking of received data, are provided by ProWare. One of the main advantages of ProWare is that it separates data acquisition and fusion from communication services, thus greatly simplifying WSN design.

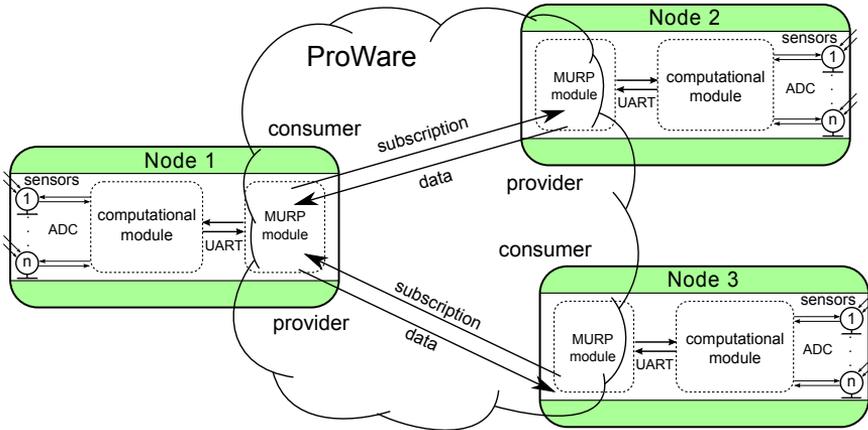


Figure 5.5: WSN node link establishment and data communication using ProWare.

Run-time sensor discovery is achieved through publish-subscribe principles. Fusion nodes express their need for specific data (e.g., situation assessments, localization intermittent data, etc.), and ProWare queries the network to find this data. Sensor nodes that receive the query reply by publishing their ability to provide this data (i.e., to successfully obtain the estimates).

The fusion node considers the replies and subscribes to sensor nodes of its choosing to start receiving data from them. Once these agreements are established, fusion nodes become data consumers and the selected sensor nodes become data providers. It is possible for a node to be both a consumer and a provider at the same time. In Figure 5.5, for example, node 1 subscribes for data from node 2 and is a data consumer in this relationship. Node 1 is also equipped with sensors and is able to provide its sensor data to node 3. It is a data provider to node 3, while at the same time being a data consumer of node 2.

Any sensor node may assume the role of a fusion node (additionally to being a sensor node), and several fusion nodes may exist at the same time for the same cluster of sensor nodes. This increases the reliability and robustness of the WSN compared to solutions utilizing a single predesignated fusion node, since there is no single point of failure. Fusion nodes are elected based on the availability of computational resources, required to perform fusion tasks, and communication signal strength.

ProWare also provides data validity checking for all communicated data [118], ensuring that only relevant, spatially and temporally correct data is received by the fusion node. The constraints for data correctness are determined by the consumer (fusion node) and provided in the subscription's description. Data producers (sensor nodes) decide whether they can service these constraints and, if so, start publishing the requested data. Published data is tagged with metadata (temporal and spatial information), and message transport time is tracked. This enables to check whether the data received from different providers is still valid and mutually compatible at the consumer's side.

In the spatial domain the sensor nodes complement the data packets with respective node positions. This allows the fusion node to subscribe to data from where it is needed (i.e., to redefine the node cluster used for fusion) and to validate whether the packets are received from the correct area. Regarding the temporal domain, the sensor nodes complement their data packets with a validity interval and packet age. The packet's age is specified in each node's local time as a time difference between the moment when the packet is sent and the first moment of event detection. The validity interval, on the other hand, describes how long the packet is usable. During packet delivery ProWare also provides estimations of communication channel delays [121] and computes the estimated duration of total transport time. Before the packet is handed over to the fusion node, its age value is incremented by the time spent on data transport. This allows the fusion node to validate the age of arrived packets and to align packets from different sensors to a unified approximated timeline, in order to compare them against a pre-specified simultaneity interval. Both the simultaneity interval and the validity interval depend on the initial assumptions about the observed process.

5.4 Conclusions

The timeliness of data fusion for global situation assessment is as important for CPS real-time operation as timely operation of each of its individual components. Generally data fusion does not consume significant amounts of computational resources and is thus able to be performed on any number of sufficiently powerful WSN nodes. However, the delays situated with data communication can disrupt smooth operation, and information concurrency and correctness cannot be guaranteed without proper network management and data validation. The temporal and spatial inaccuracy of data caused by WSN asynchronous operation adds to the overall estimate error of the system as a whole. The margin of that error needs to be specified for each specific application in order for the system to present both timely and accurate situation assessments.

Chapter 6

Applications

The chapter discusses several considered applications of the signal processing techniques presented in this work. The chapter begins with the discussion of the industrial machinery monitoring problem [14,19]. Pattern recognition methods presented in Section 3.1 are applied to this problem for machine operation state identification and malfunction detection. The chapter proceeds with the discussion of military Intelligence Surveillance and Reconnaissance (ISR) tasks [15] and presents three topics of acoustic and Passive Infrared (PIR) sensor application in a ISR system of distributed ground sensors. The first topic is situated with military vehicle identification and trajectory estimation. The multistage procedure discussed in Section 3.2 is applied to vehicle identification, and the distributed localization procedure presented in Section 4.3 along with the trajectory estimation approach discussed in Section 5.2 are applied to vehicle trajectory estimation. The second topic considers foot soldier detection and movement speed estimation with PIR sensors. For this application a pattern recognition method based on time domain feature analysis is presented and verified on experimental data. The third ISR system topic considers gunshot detection and shooter acoustic localization [9, 22]. For this application the methods of DOA estimation discussed in Chapter 4 are applied along with a proposed data fusion procedure of gunshot geometry estimation and shooter localization. Experimental results of methods proposed for each of the considered applications are presented along the course of the chapter.

6.1 Industrial Machinery Monitoring

Thorough monitoring and diagnostics of industrial machinery is essential in almost any production process. Receiving on-line feedback on equipment utilization ensures uninterrupted flow of the production process. Analysis of machinery utilization also allows for process optimization and thus increases

efficiency of the production facility [42]. Monitoring machinery performance and condition allows to eliminate faults and malfunctions in equipment prior to breakdown. This decreases the number of accidental failures and reduces maintenance costs.

6.1.1 Problem statement

Most high-end industrial devices are equipped with integrated monitoring solutions that enable device interconnection and process management through a SCADA network. Very often, however, a production process is manually operated or equipped with outdated machinery that lacks integrated sensors and control hardware. This situation typically occurs in the workshop floors designed according to outdated process management standards. Applying a monitoring solution to this kind of machinery presents a set of challenges, the most evident one being situated with hardware integration. Installation of sensors may prove hazardous and also expensive due to interruptions in the production process. Furthermore, if a machine requires a safe zone or is concealed in a safety chamber, sensor installation may be simply prohibited by safety regulations. Applying distributed contactless sensors connected wirelessly in order to reduce installation costs can resolve the problem. Acoustic sensors are chosen as a source of contactless information along with other modalities [51], e.g., visual light, hyperspectral, vibration, etc.

Acoustic information analysis applied to industrial machinery operation state identification [14] was discussed as an example of single-sensor solution in Section 3.1. However, high levels of noise originating from different machines in close proximity in a highly reverberant environment of the shop floor may corrupt the signal to a degree, where the classification of the machine working state becomes unreliable. Using several sensors in an array configuration provides the possibilities of localizing a specific noise source, concentrating on it and attenuating unwanted noise digitally. Microphone arrays in industrial applications are mainly used for acoustic holography, in which machinery or products are examined for areas of high or abnormal noise [82]. These techniques have been applied in a wide range of applications, e.g., diagnostics of vehicle internal and external noise sources [82], estimation of fan and turbine noise [103], detection of compressed air leakage [52], etc. Multi-channel acoustic signal analysis also benefits the optimization of processes, for example laser welding [130]. Research in the field of machinery noise localization at the shop floor has also been discussed in [136] and [24].

6.1.2 State Identification using Multimodal Information

Industrial machinery state identification using acoustic and vibration signal analysis is thoroughly discussed in Section 3.1. Figure 3.1 illustrates the main

principle of multimodal data acquisition at different measurement points. Vibration sensors provide localized information from specific parts of a machine onto which they are mounted. Remote acoustic sensors, on the other hand, are susceptible to a wider variety of noise produced during every registered operation state. During initial signal analysis a set of distinguishing features is specified for every machine under test and the fitness of these sets is evaluated using a simple correlation based classifier and a model based fuzzy logic classifier. Automatic feature selection is additionally handled and the results are verified by the fuzzy logic based classifier. The combined results of process state classification are presented in Table 6.1. The table shows that it is possible to manually specify or automatically select a set of features that yields high classification accuracy for every machine under test and for every considered modality. The majority of the misclassified signal frames originates from the periods of state transition, which is natural because process dynamics are not settled during the transitional period, and the system is trained to identify steady-state parameters.

Table 6.1: Results of process classification in (%) combined from Tables 3.3–3.5

Classifier type	Feature extraction	Saw acoustic	Saw vibration	Router acoustic	Router vibration
Fuzzy	Band energies	94.27	90.62	98.67	99.85
	Instantaneous	98.72	83.75	97.83	99.28
	MFCC	96.49	81.65	98.72	99.49
Correlation	Band energies	81.25	54.06	91.95	98.39
	Instantaneous	74.19	68.91	53.82	96.44
	MFCC	87.90	65.69	95.72	99.85
Fuzzy	DEFS	94.96	70.31	99.08	99.88
	LSFS	95.16	58.12	98.80	99.76
	MCFS	95.56	65.27	98.68	77.73

The overall state identification accuracy can undoubtedly be increased by introducing additional modalities to the identification effort and increasing the number of measurement points. For more sophisticated process identification, where the two discussed modalities do not offer sufficient identification quality, process analysis is performed until the optimal set of modalities and most discriminative measurement points are established.

Production Process Fault Detection

The detection of machinery malfunctions or faults in the production process can be performed by the identification of abnormal patterns in the monitored process. Figure 6.1 illustrates turning tool fault detection in a Computer Numerical Control (CNC) lathe. The lathe is specifically programmed for the

experiment to perform an improper tool movement which can result in the tool breaking. The pattern recognition system is trained prior to the experiment to classify the spindle rotation (class 1) and cutting process (class 2). The fuzzy logic based classifier is trained by modeling the feature sub-spaces by triangular Membership Functions (MF) in order to strictly specify the feature sub-space bounds (see Section 2.5.3). As a result the abnormal behavior, which happens approximately at the 16th second in Figure 6.1, falls out of the confined feature sub-spaces of the specified classes and is left unclassified. Combining this type of detection with abnormal readings from other sensors (e.g., vibration) can indicate a fault, given that the source of signal abnormality does not originate from elsewhere.

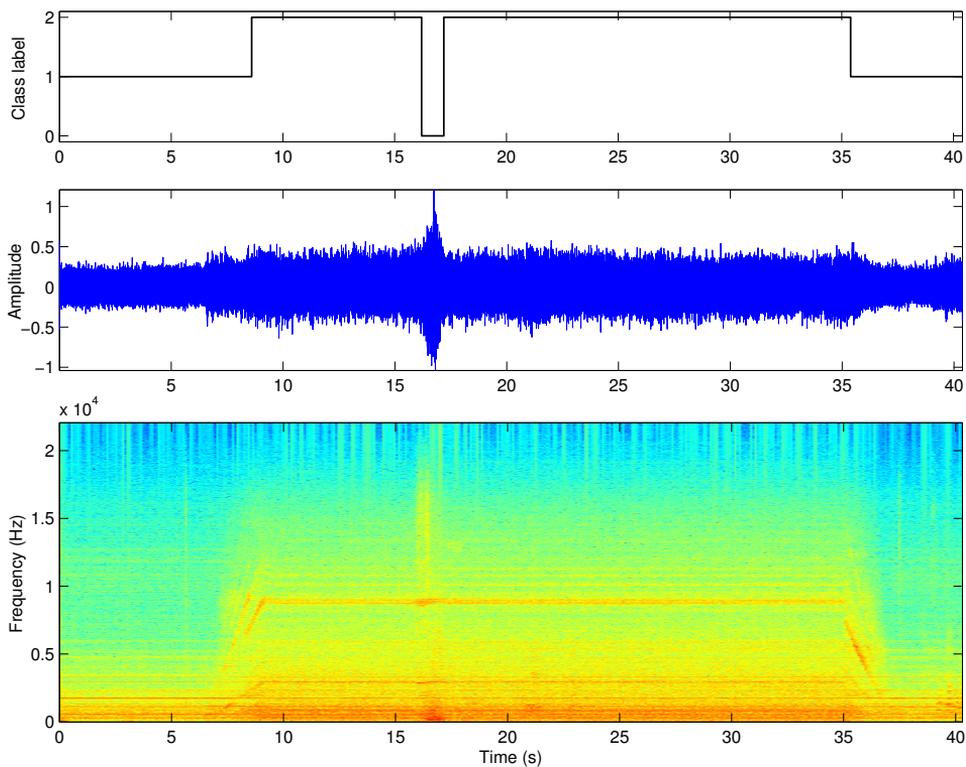


Figure 6.1: Lathe turning tool fault detection by an acoustic sensor.

6.1.3 Localization of Machinery Operating Region

Applying multi-channel sensors to industrial machinery monitoring allows to locate the region of operation of any particular machine in the FOV. Acoustic noise emission from any specific region of a machine provides additional information to the process state identification system and reduces the risk of falsely identifying a similar pattern incoming from a neighboring machine or

mistaking any background noise for a process state. Furthermore, in a well established setting of the shop floor, where the positions of all machines and tools are known, it is possible to concentrate the directivity pattern of the acoustic arrays on specific regions of interest and attenuate unwanted noise via beamforming.

Experimental Setup

The proposed solution consists of several small sub-arrays that form a wide aperture array with a FOV of 15–25 m². With each such array the system monitors a separate machine or part of a larger object, such as a conveyor chain. The applied approach to acoustic localization in the horizontal plane is discussed in Section 4.3. For the experiments the wired array configuration consisting of four sub-array blocks, discussed in Section 4.4, is used. The signals are sampled at a rate of 8 kS/s per channel. Two array placement configurations are used: linear — the sub-arrays are placed in a straight line with the steering angles $\alpha = \{0, 0, 0, 0\}$; angular — two sub-arrays are placed perpendicular to the remaining two with $\alpha = \{90, 90, 0, 0\}$.

The experiments are performed at the shop floor of a small manufacturing facility during common operational conditions. The two machines chosen for testing are a CNC lathe and a manually operated lathe. The array placements for the CNC lathe and manual lathe are presented in the Figure 6.2a and Figure 6.2b, respectively. Such configurations do not occupy the operation space and thus do not impede personnel movement. The CNC lathe is programmed for a short turning cycle with the motor being turned on, the spindle put into rotation (i.e., the transmission put on the spindle), and one lengthwise pass on the manufactured piece with a spinning tool. The same operations are performed manually on the manual lathe with the operator located at his normal working position.

Results of Localization

Both machines under test are successfully localized by the two array placement configurations. However, the drawback of the linear configuration, discussed in Section 4.4.1, is evident in the manual lathe case. Figure 6.3 presents the result of manual lathe localization with both the motor and spindle turned on. For the ISRR approach (top subplot) blue lines denote sub-array estimated DOA, purple stars denote the intersections of DOA vectors, black rectangles denote the confined regions, and black circles denote the SRP energy maxima with their values situated to the right. Multilateration coordinate estimates are denoted by green diamonds, and the confined regions are denoted by dotted rectangles. For the SRP-PHAT approach an SRP image of the FOV spatial grid is presented. Figure 6.3 shows that the

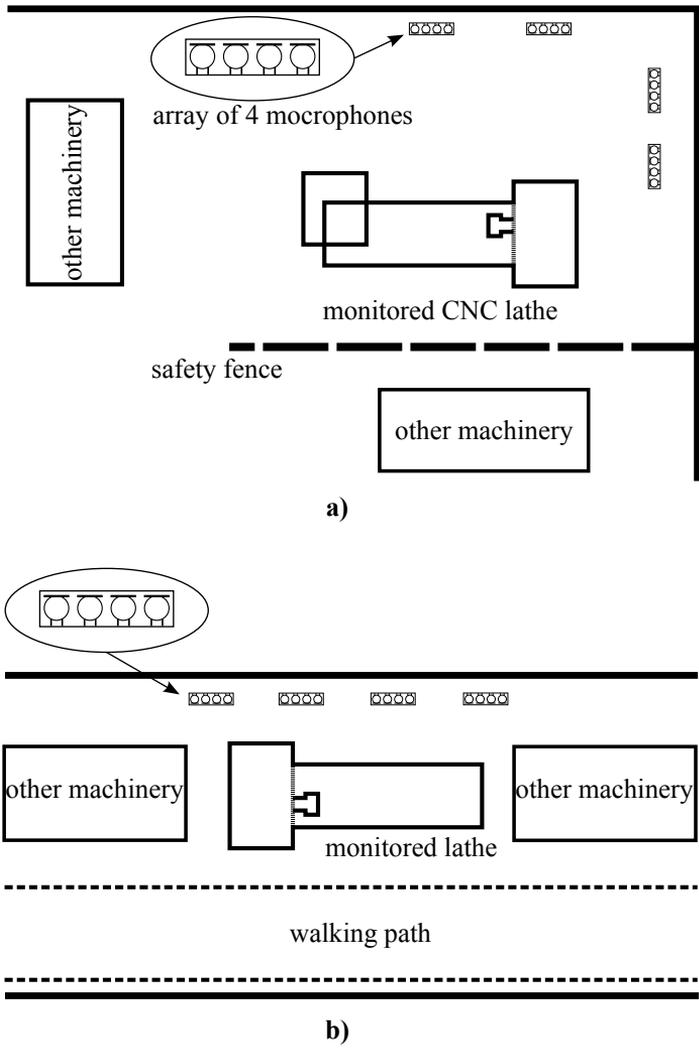


Figure 6.2: Array placement for monitoring the CNC lathe (a) and the manual lathe (b).

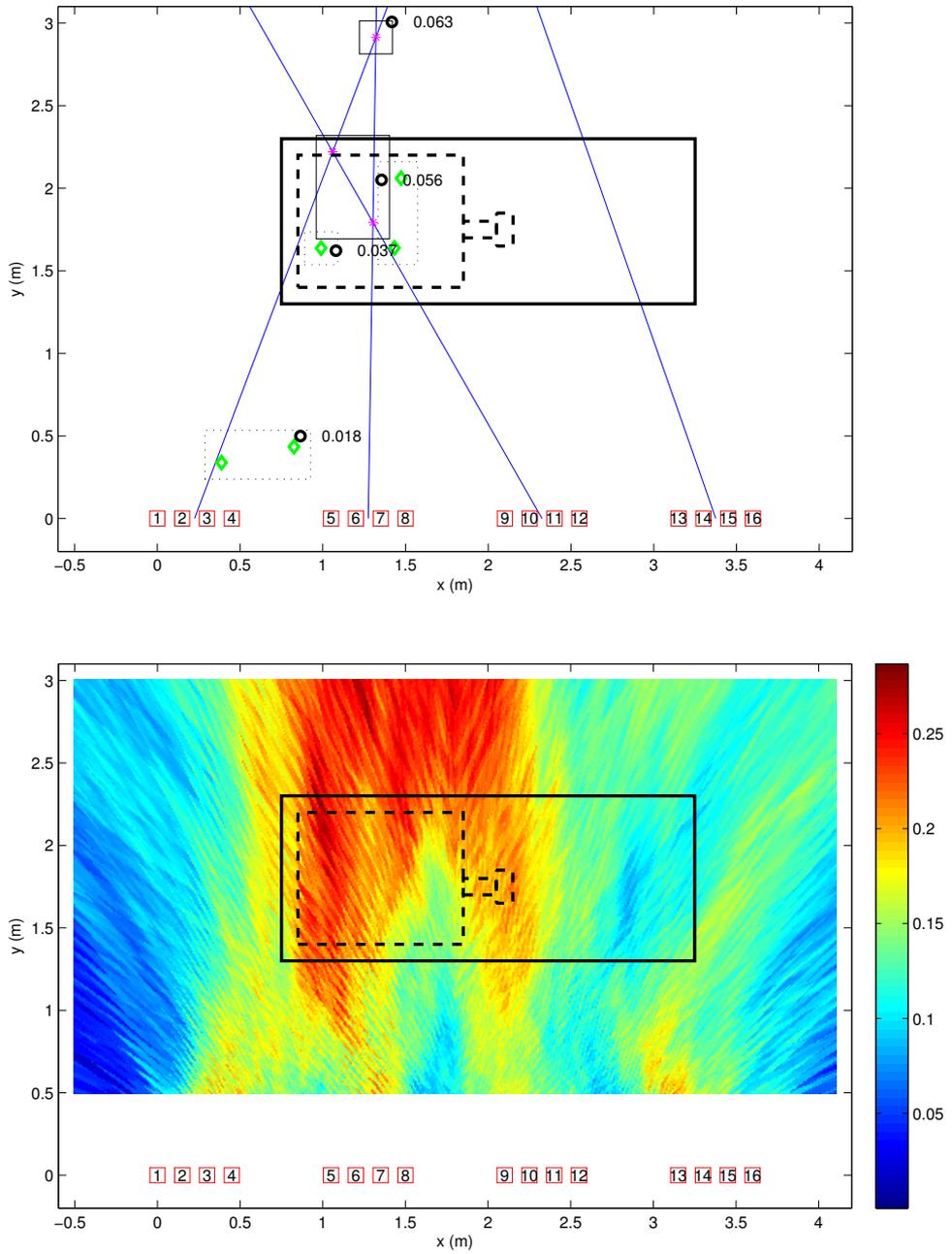


Figure 6.3: Localization result of the manual lathe with motor and spindle turned on produced by ISRR (top) and SRP-PHAT (bottom).

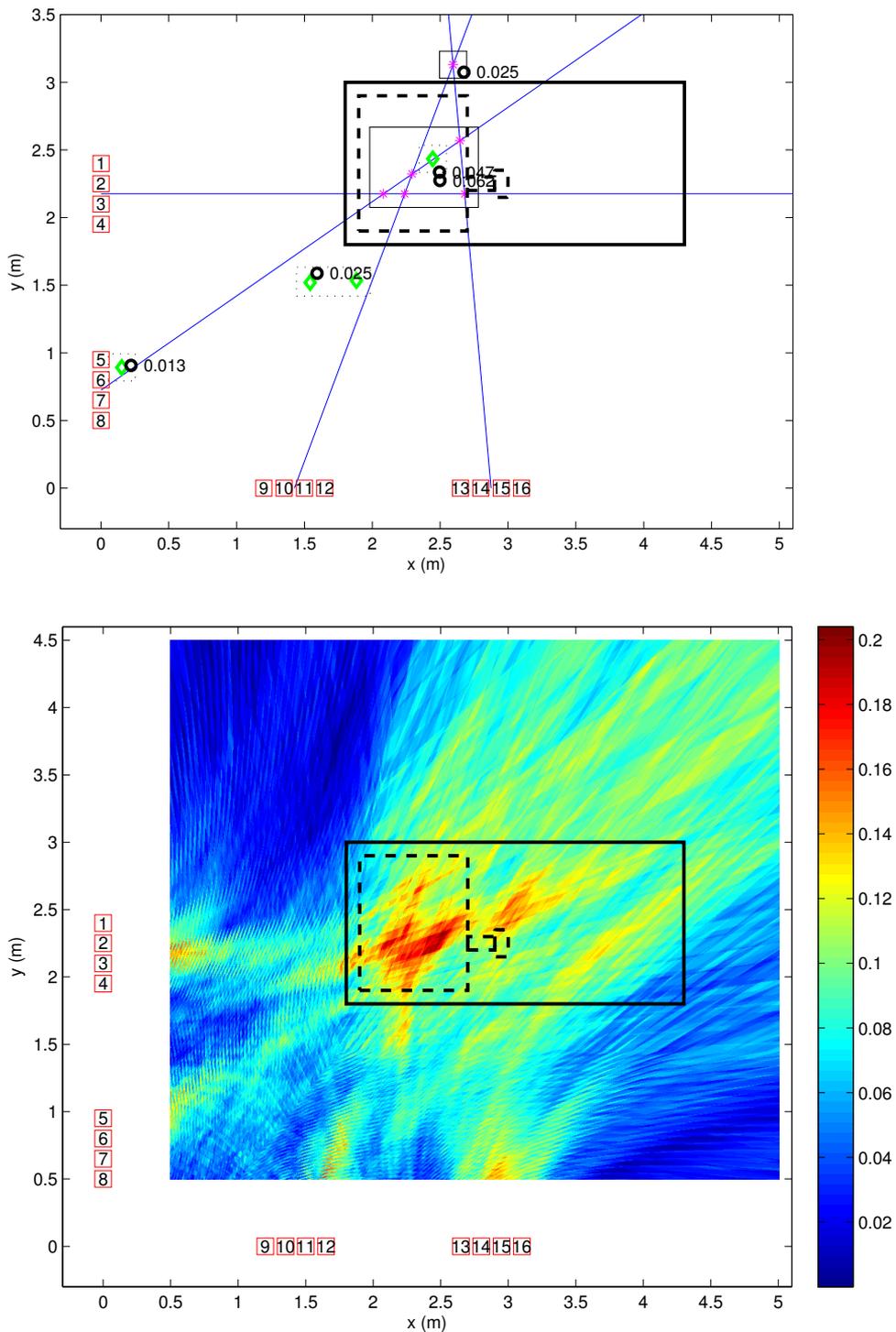


Figure 6.4: Localization result of the CNC lathe with motor turned on produced by ISRR (top) and SRP-PHAT (bottom).

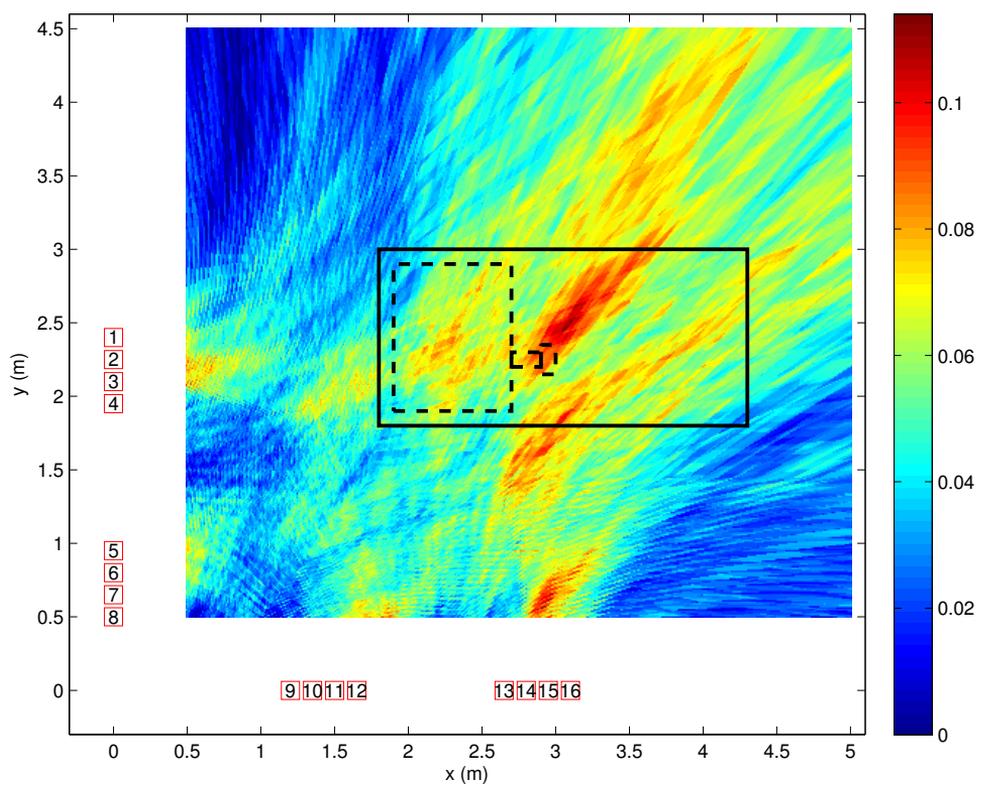
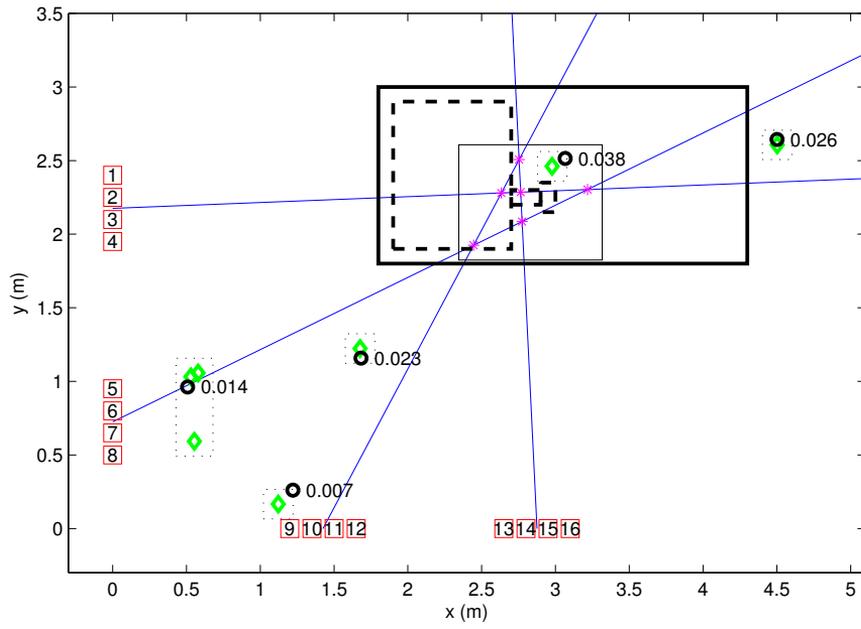


Figure 6.5: Localization result of the CNC lathe performing treatment with a spinning tool produced by ISRR (top) and SRP-PHAT (bottom).

proposed ISRR method of distributed localization succeeds in locating only the motor noise emitting region. SRP-PHAT also produces a rough image with the majority of elevated SRP values located behind the lathe. The figure illustrates that even with a significantly spread wide aperture array close proximity localization of a large object with several noise emitting regions is complicated. Nevertheless, both methods achieve crude localization, which can indicate the operation of this particular machine and reassure state identification estimates.

Localization of the CNC lathe with the array angular configuration proves to be more precise. Lathe components that emit most noise are the DC motor, transmission, spindle and the area of manufactured piece. In the motor no-load state (i.e., the transmission is off the spindle) the motor itself is well distinguishable, as it is shown in Figure 6.4. Spindle rotation and manufactured piece treatment produce a more spread, but, nevertheless, localizable region, as presented in Figure 6.5. The angular configuration performs significantly better than the linear one, localizing separate noise emitting regions. Such level of localization down to the machine specific part in operation can be used in data fusion to reassure the state estimates situated with that particular part.

Regarding the two approaches to ISRR, the multilateration approach generally performs worse for the task of machine noise emitting region localization than the DOA approach. While the DOA approach to ISRR behaves similarly to SRP-PHAT, confining the noise source region when possible, the multilateration approach produces a large number of incidental regions and rarely settles on the true noise source even when it is well established by the two other methods. This indicates the poor ability of multilateration to localize significantly large noise sources at close proximity.

6.2 Identification and Tracking of Military Vehicles

The considered application of military vehicle identification and trajectory estimation is a part of a larger study performed by the author of the thesis and the Research Laboratory for Proactive Technologies of the Tallinn University of Technology on the topic of Military Intelligence Surveillance and Reconnaissance (ISR) system development. Other applications situated with the ISR system include detection and movement speed estimation of ground troops and gunshot acoustic identification and shooter localization, which are handled in Section 6.3 and Section 6.4, respectively.

6.2.1 Intelligence Surveillance and Reconnaissance

The tasks of military intelligence surveillance and reconnaissance consist of constant and systematic monitoring of the area of potential threats or actual assault and immediate notification of every concerned party about the changing situation [111]. The common monitoring tasks include threat detection, identification of friend or foe, identification of various objects (e.g., vehicles [150,160], munitions [132], etc.) and events [88]. In traditional ISR systems human operators handle raw data incoming from human components, such as reconnaissance teams and scout patrols, and computer components, such as surveillance cameras, reconnaissance drones, motion detectors, etc., and perform situation evaluation themselves. In state of the art systems with higher levels of computer system integration raw data is processed by the autonomous system and human operators handle the provided situation assessments, which increases overall system agility and area coverage [66]. In modern ISR systems ground situation assessment is performed by Unmanned Ground Sensors (UGS) with support from aerial patrolling by Unmanned Aerial Vehicles (UAV) [117, 120]. In the environments, where aerial view can often be obstructed (e.g., urban and woodland), the ground ISR system serves as the primary source of situation information.

The individual components of an ISR system are distributed in the wide area of operation and communicate through RF channels forming an interconnected WSN. This WSN must be decentralized because the viability of each component is expected to be low. In a decentralized network the loss of any individual sensor or data aggregation component will not disrupt overall system operation. Thus an ISR system serves as a very good example of a CPS with distributed autonomous components. Furthermore, modern ISR systems consist of standalone sub-systems (e.g., ground, aerial and manned), while forming an even more complex Systems of Systems (SoS). Situation evaluation in such systems is performed *in situ* by every sub-system rather than in the information hub, and system-to-system interaction greatly exceeds machine-to-human interaction [111].

The considered ground ISR comprises multiple distributed smart sensors (UGS), which perform environment monitoring by analyzing signals of different modalities and collaborate on global situation assessment. The larger SoS consists of various ground and aerial ISR systems, information accumulation and fusion databases. The situation evaluation results produced by autonomous components are available to human operators in the operation center, as well as to local scout patrols. In the scope of the entire SoS the ground ISR component is a provider of ground information, and other sub-systems are considered external information consumers. The principle diagram of the considered system is presented in Figure 6.6. The smart sensors employ multiple modalities, including acoustic, magnetic and Passive

Infrared (PIR). Acoustic sensors are used for both acoustic source identification and localization. Magnetic and PIR sensor readings complement acoustic information in detection of vehicles and foot soldiers.

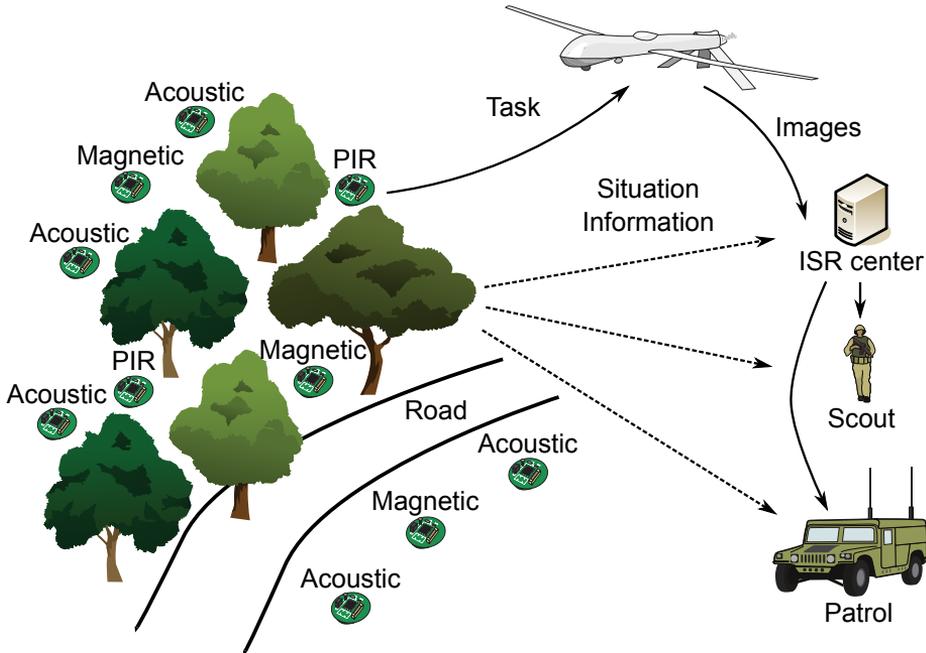


Figure 6.6: Principle diagram of the ground ISR system and other SoS components.

6.2.2 Military Vehicle Experiment Setup

The experiment was performed in a military base in the city of Tartu, Estonia. The military vehicles were provided by the Estonian Defence Forces under the European Defence Agency project IN4STARS. The experiment was performed in early September under varying weather conditions, i.e., clear calm weather with occasional short rains and wind gusts of up to 15 m/s and temperature of $t^\circ \simeq 18^\circ \text{C}$.

A total of twelve UGS equipped with linear microphone arrays were used for the task of acoustic localization and trajectory estimation. The UGS were manually grouped into three clusters, each containing four UGS. Each cluster was operating in the common FOV of its UGS independently from the other two clusters. The clusters were distributed along the vehicle driving path as portrayed in Figure 6.7. In the figure the red squares marked by S_i denote the linear array UGS. These squares are disproportionate to the image scale, being drawn bigger for better presentation quality, however, they retain true array positions and steering angles. UGS latitude/longitude coordinates were measured using a standalone GPS device (Trimble R8 GNSS) and later

converted into a local planar coordinate system for offline analysis with UGS S_1 of cluster 3 being set as the zeroth coordinate. UGS steering angles relative to the Earth magnetic north were measured with a high-precision compass.

Each linear array is implemented on the BeagleBone Black platform, as discussed in Section 4.5.2. The array consists of six microphones enclosed in an aluminum housing with a distance of $l = 10$ cm between consecutive microphones. The acoustic signal is sampled at the sampling rate of $f_s = 20$ kS/s per channel and processed in frames with lengths of 2730 samples per channel. The left image of Figure 6.8 presents one of the UGS used in the experiment.

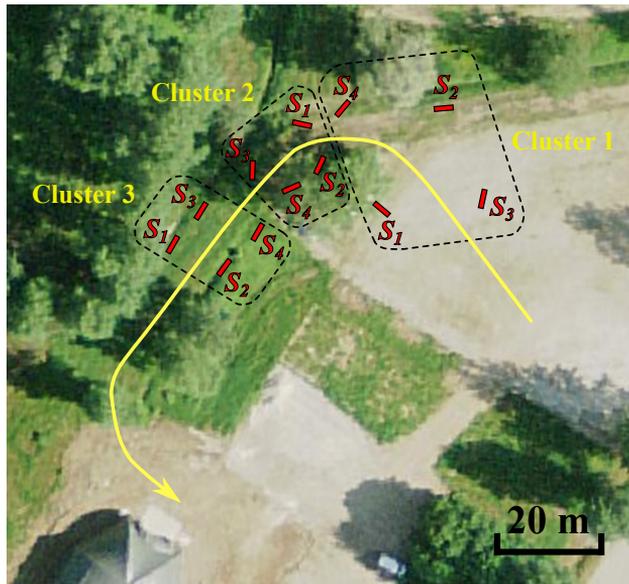


Figure 6.7: Acoustic array UGS placement and vehicle driving path.

The UGS perform the dedicated tasks (i.e., DOA estimation, etc.) autonomously and send the intermediate information through data aggregation nodes to a data fusion and situation assessment station implemented on a dedicated server. The subscription to data provided by the UGS is performed by means of the middleware component, reviewed in Section 5.3.3. A synchronized start time of the WSN is achieved by broadcasting a sequence of specially timed messages from the control node (six messages counting down from 100 ms with 20 ms intervals), which is used to trigger the concurrent start of signal sampling within all three UGS clusters. Two data aggregation nodes are presented in the right image of Figure 6.8. Additionally to forwarding the intermediate data each UGS also saves it along with the raw multi-channel signal on a memory card for later offline analysis.

The four vehicles used in the experiment were: a light patrol jeep, pre-



Figure 6.8: Linear array UGS prototype (left) and two data aggregation nodes (right).

sented in Figure 6.9, a personnel truck (with an empty back), presented in Figure 6.10, a medical utility truck, presented in Figure 6.11, and an armored personnel carrier, presented in Figure 6.12. Each vehicle took multiple drives without stops along the path highlighted in Figure 6.7 with speeds raging from 10 km/h to 35 km/h.

6.2.3 Vehicle Localization and Trajectory Estimation

Vehicle localization was performed by the proposed approach to distributed acoustic localization, discussed in Section 4.3. As the experiment was performed using wirelessly connected asynchronous (softly synchronous) sensors, the multilateration approach to ISRR could not be applied for comparison. Instead, only the DOA approach was used. For trajectory estimation the approach presented in Section 5.2 was applied. The Kalman Filter (KF) was applied for solving the dynamic system (2.36) of the object movement model, and the Rauch-Tung-Striebel (RTS) smoother was used for final trajectory smoothing.

General Results

The system succeeded in localizing each of the vehicles as a single confined region for the majority of vehicle pass instances. As strict temporal data validation was not performed by the middleware component, an inter-node synchronization error not exceeding 5 ms remained in the WSN. This resulted in a slight decrease in localization accuracy as the DOA estimation in each of the UGS clusters was performed with a slight lag between the nodes. The trajectory estimation procedure partly corrected sequential localization estimates. All estimated trajectories resemble the curved path presented in Figure 6.7 despite several cases of the localization procedure failing to follow the course of the acoustic source.



Figure 6.9: Light patrol jeep in the vicinity of UGS cluster 3.



Figure 6.10: Personnel truck in the vicinity of UGS cluster 3.



Figure 6.11: Medical utility truck in the vicinity of UGS cluster 3.



Figure 6.12: Armored personnel carrier in the vicinity of UGS cluster 2.

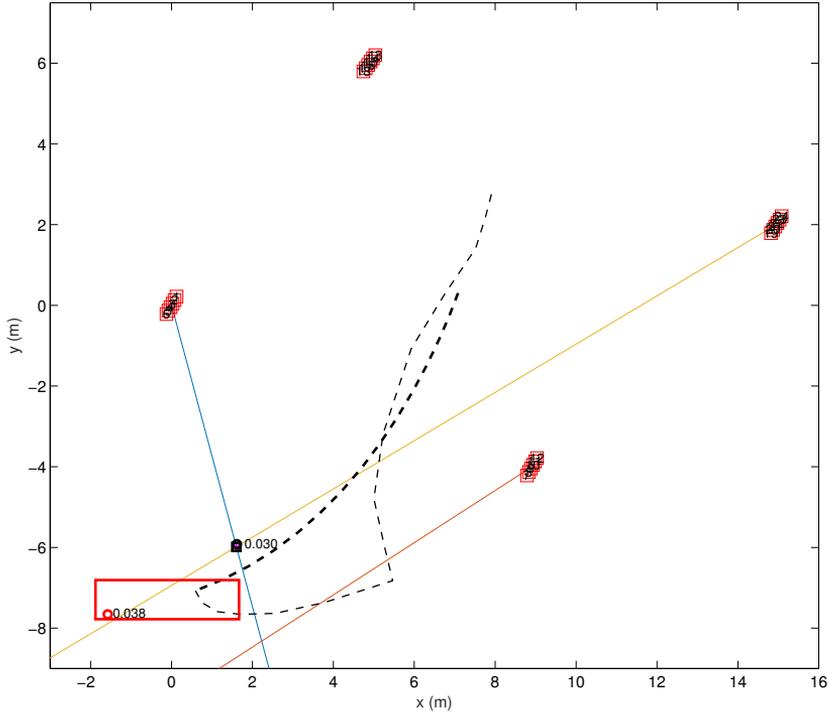


Figure 6.15: Result of vehicle localization and trajectory estimation for UGS cluster 3.

Case Analysis

An example of acoustic localization and trajectory estimation for a single vehicle pass is presented in Figures 6.13–6.15. In the figures straight lines of various colors starting from array centers denote the DOA estimates of each UGS, black rectangles denote the localization regions confined by ISRR, and the red rectangles denote the search regions (5.1) predicted by KF. During offline analysis SRP-PHAT with SRC was applied to each of the regions to estimate the maximal cumulative energy values of the softly synchronous signals. The spatial positions of these maxima and their values are denoted by small black and red circles for ISRR confined regions and KF predicted regions, respectively. Black dotted lines denote the estimated trajectories and thick dotted lines denote smoothed final trajectories.

Figure 6.13 illustrates an example of maintaining the estimated trajectory by the predicted region while the localization estimate goes off course. The higher SRP value of the predicted region in this case indicates that the source is in fact better located by the trajectory estimation procedure than by ISRR alone. The inconsistencies in trajectory estimation at the beginning of vehicle pass result in a trajectory skewed towards UGS S_3 of cluster 1. Applying the RTS smoother resolves the problem, which results in a curve following the vehicle path.

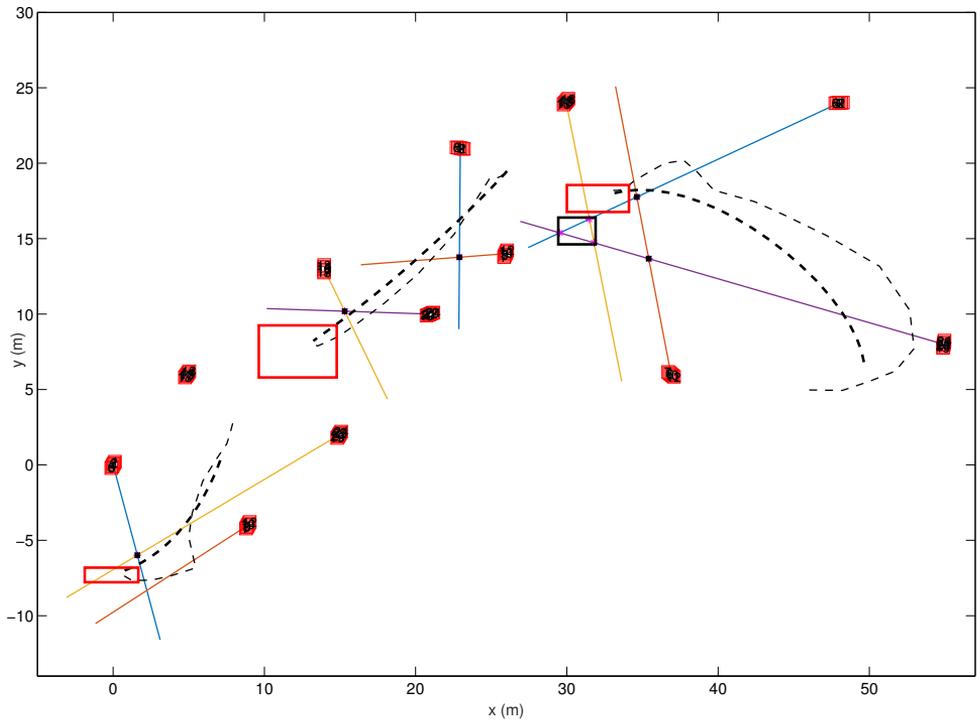


Figure 6.16: Combined result of vehicle localization and trajectory estimation for all three UGS clusters.

UGS cluster 2 has its arrays steered towards the center of the common FOV. Such a configuration proves to perform better than the one of cluster 3, where the UGS are steered parallel to the vehicle path. The result of trajectory estimation by cluster 2 is presented in Figure 6.14. On the very edge of the FOV the source is not detected by UGS S_1 and S_2 , however, the trajectory estimation procedure succeeds in leading the source up to and over the FOV edge. Both the successively estimated and the final smoothed trajectories conform with each other. Cluster 3 performs worse than cluster 2, partly due to UGS S_3 having a technical error and thus not providing DOA estimates. This results in the source not being localized at the FOV edge of vehicle approach. Nevertheless, the trajectory is estimated well for the latter part of the FOV up to its opposite edge.

The combined result of trajectory estimation by all three UGS clusters is presented in Figure 6.16. The resulting smoothed trajectory closely resembles the vehicle path specified in Figure 6.7. As the true vehicle positions at any given moment of driving through the UGS clusters are unknown, precise localization error estimation is not possible. Furthermore, the scale of UGS clusters does not significantly exceed the scale of the localized vehicles. Therefore, localization to a point is both not possible and ultimately unnecessary. The overall trajectory of driving through all three clusters is obtainable at the stage of data fusion through extrapolation and additional smoothing.

6.2.4 Vehicle Identification

Vehicle identification is performed by applying the multistage procedure, discussed in Section 3.2.2. The procedure runs on every UGS parallel to localization and trajectory estimation using the acoustic signal of one of the microphones of the UGS array. The signal RMS energy threshold is chosen such that the classification stages of the multistage algorithm are executed only during vehicle pass at close proximity, i.e., during the Attack Sustain Release (ASR) envelope dynamic. As the vehicles under test produce a significant amount of noise, it is acquired even from a distance of the vehicle parking area (beige regions in Figure 6.7). The acoustic patterns acquired at this distance are distorted and can cause classification errors. The classifiers are trained accordingly on the features extracted from signals frames corresponding to the actual vehicle pass intervals.

Event Concurrence

For temporal validation of vehicle identification results, the spatial distribution of UGS in the clusters must be taken into consideration. Figure 6.17 presents signals acquired by UGS of cluster 2 corresponding to a slow pass

of the medical utility truck. Each subplot of the figure represents a signal acquired by the first microphone of UGS S_i , $i = 1, \dots, 4$. The first portion of the signals (up to the 30th second) corresponds to the vehicle starting up and driving to its starting position, which is denoted in Figure 6.7 by the beginning of the path line. This portion is acquired by the UGS with no visible delays between the signals because the noise originates from outside the cluster. The signal interval between approximately 30 and 62 seconds corresponds to the drive itself. Due to the spatial distribution of sensors the ASR dynamic of the vehicle passing every UGS is registered with a significant delay, which is measured in seconds. UGS S_1 and S_2 are naturally the first to detect the pass and the remaining two UGS register the Sustain portion of the envelope 2–3 seconds later. From the system’s viewpoint at the frame refresh rate of 136.5 ms these events are spread too far apart to be considered simultaneous and corresponding to the same vehicle pass event. To tackle the problem the localization results that show a single object passing at this specific time interval are used to combine the separate identification estimates during data fusion.

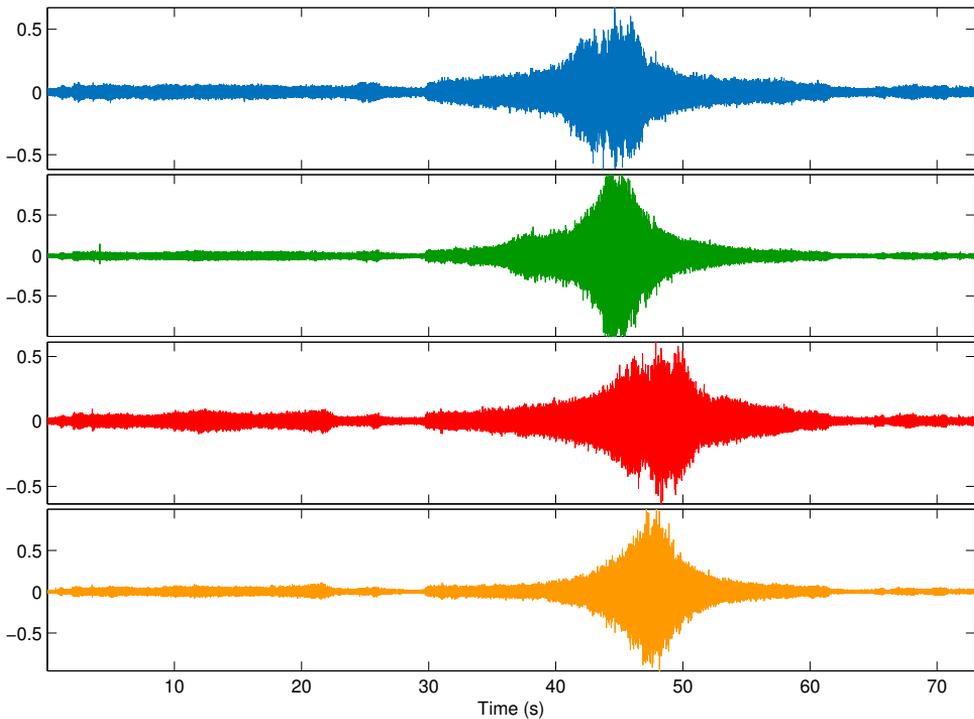


Figure 6.17: Acoustic signals of a medical utility truck pass acquired by UGS cluster 2. From top to bottom: signal of the first microphone of UGS S_1 up to UGS S_4 .

Classifier Training and Validation

The fuzzy logic based classifier was trained on the features extracted from the Sustain portion of the vehicle pass signals. The feature set used for feature extraction corresponds to the one presented in Section 3.2.2: band energies (2.16) with 15 sub-bands chosen in the interval of 10–3000 Hz, spectral centroid (2.10), spectral roll-off (2.14), spectral slope with parameters (2.17), (2.18). Reference vectors for the correlation based classifier (3.2) are generated by averaging several spectra corresponding to the Sustain signal envelope portion of several passes per each vehicle type. For class labeling a separate class is dedicated to each of the vehicles under test regarding all the possible driving speeds of 10–35 km/h. The class labels are set as:

- patrol jeep — class 1,
- personnel truck — class 2,
- medical utility truck — class 3,
- armored personnel carrier — class 4.

Two test datasets were generated for the fuzzy logic based classifier consisting of features corresponding to narrow intervals of the middle portions of the Sustain signal envelope part, and features corresponding to all Attack Sustain and Release signal envelope portions. These two datasets are referred to as the narrow and wide datasets, respectively. The results of fuzzy classifier testing on the narrow dataset are presented in Figure 6.18. The figure shows that class 1 and class 2 are well distinguishable from class 3 and class 4. Several samples are incorrectly classified between class 1 and class 2, however, both are sufficiently separable by the chosen features. Class 3 and class 4, on the other hand, are mistaken for each other for quite a large number of samples. This poor separability of the two classes leads to a lower classification accuracy of the entire set. Close frequency domain analysis of class 3 and 4 vehicle acoustic patterns has shown that the patterns are in fact very similar. Thus the problem of poor separability is not situated with the choice of features.

Testing the fuzzy classifier on the wide dataset results in a larger number of misclassifications, as portrayed in Figure 6.19. Classes 1 and 2 remain well separable with the rare exception of class 2 being mistaken for class 1. The classification accuracy for classes 3 and 4 is worse than in case of the narrow set because additionally to being poorly separable, both are mistaken for class 1 and class 2 for a number of samples. Closer analysis shows that the acoustic patterns of class 3 and class 4 vehicles closely resemble those of class 1 and class 2 at the early stage of Attack and the late stage of Release signal envelope portions. This circumstance further justifies the need of event identification only during the actual vehicle pass and not during its approach.

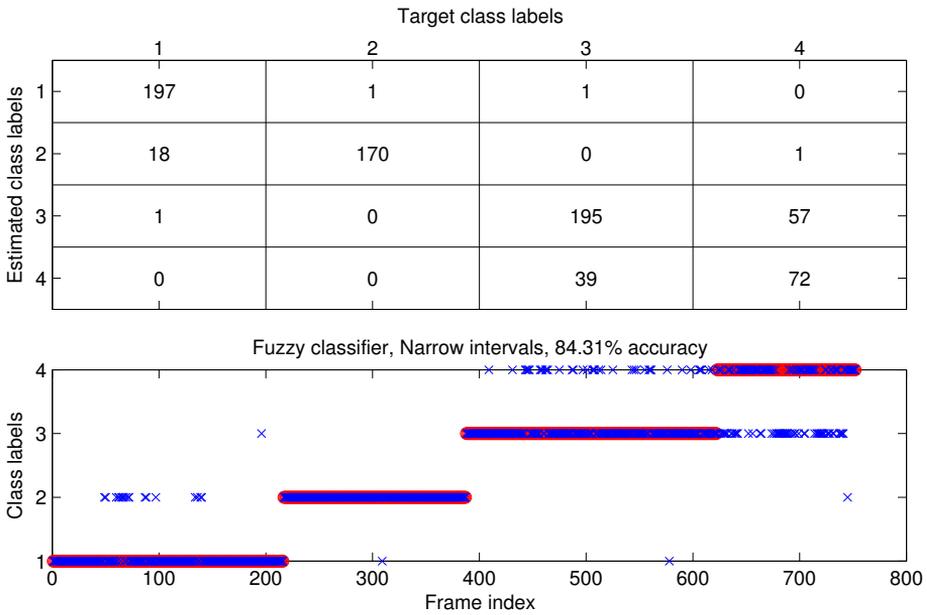


Figure 6.18: Result of fuzzy classifier evaluation on the narrow test dataset.

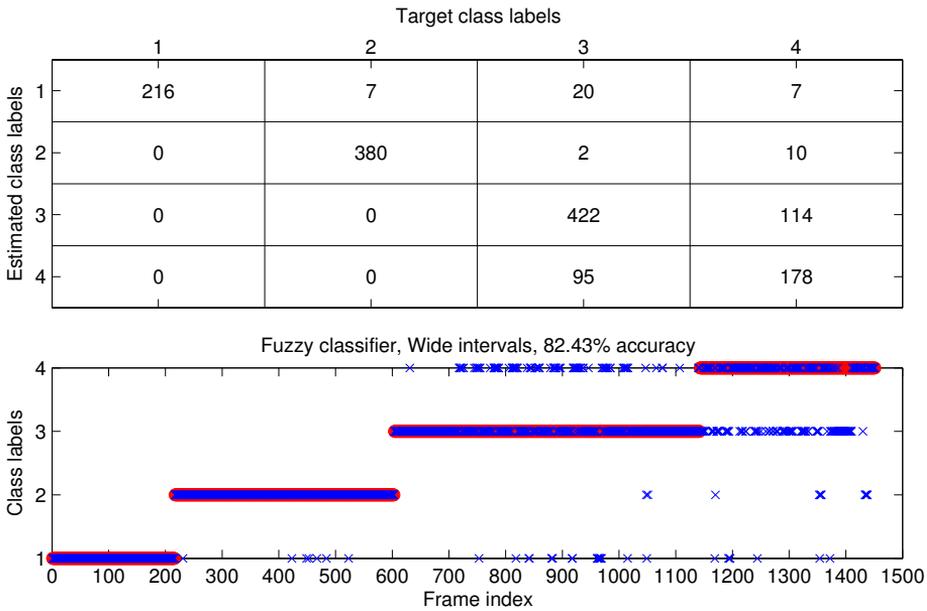


Figure 6.19: Result of fuzzy classifier evaluation on the wide test dataset.

Vehicle Identification Results

The multistage vehicle identification procedure produces a class label for every signal frame that corresponds to the detected vehicle pass interval, i.e., the interval with the RMS energy above the threshold confined by the ASR dynamic. Both the fuzzy logic based and correlation based classifiers produce vehicle type class labels (as opposed to one distinguishing between a vehicle pass event and all other possible events). For combined label generation the fuzzy classifier has greater priority, while the correlation based classifier supports the decision made by the fuzzy classifier. This is achieved by adding a weight coefficient to the class label produced by the fuzzy classifier, which depends on the correlation value corresponding to that class and its distinctiveness from the values corresponding to other classes. The final class label for the entire vehicle pass interval is calculated by identifying the most frequent class label with the largest weight among the individual frame labels of that interval.

Figures 6.20–6.23 present identification results for the pass intervals of vehicles of class 1 up to class 4, respectively. The figures consist of four subplots; from top to bottom these are:

1. Acoustic signal with one instance of vehicle pass (blue); final estimated labels with value 0 corresponding to event non-detection, 0.25 corresponding to class 1, 0.5 — to class 2, 0.75 — to class 3, and 1 — to class 4 (black);
2. RMS energy readings per frame (blue); signal energy threshold exceeded or not (red line);
3. Class labels generated by the fuzzy logic based classifier;
4. Coefficients of correlation with the reference spectral vectors (blue — class 1, green — class 2, red — class 3, magenta — class 4).

Classes 1 and 2 are well identifiable, as shown in Figure 6.20 and Figure 6.21. For both class 1 and class 2 the fuzzy classifier does not produce false class estimates. A number of frames receive the zero label, which means that the features of these signal frames are out of bounds of any class feature subspace. The correlation coefficients also indicate the correct class well. In Figure 6.20 the correlation coefficients of class 1 are significantly higher than the ones corresponding to other classes for the entire vehicle approach and pass interval. This gap is less evident in Figure 6.21 for class 2. There the correlation coefficients corresponding to class 2 are dominant only for the Attack and Sustain portions of the vehicle pass signal, while fuzzy classifier estimates are consistent for the entire pass interval. Classes 3 and 4, on the other hand, are poorly separable, which is evident from Figure 6.22 and Figure 6.23. In case of class 3 (Figure 6.22) a number of frames are incorrectly

classified as class 4 by the fuzzy algorithm, while the correlation method results in a notable dominance of class 3 coefficients along the entire duration of vehicle pass. During class 4 vehicle identification (Figure 6.23) the majority of frames corresponding to the Attack and Decay portions of the signal dynamic are incorrectly classified as class 3 by the fuzzy classifier. The dominant correlation coefficients also vary between classes 3 and 4 during the vehicle pass interval. However, applying the search for the most popular label in the entire pass interval results in correct final estimate generation in both class 3 and class 4 cases. Classes 3 and 4 are also confused with class 2 while the vehicles are far from the measurement point, which is evident from both the fuzzy classifier and correlation results in Figures 6.22–6.23. However, this does not pose a problem because during online operation the classifiers are not invoked unless the signal RMS energy exceeds the threshold and the ASR dynamic is registered.

The problem of strong noise resembling the ASR dynamic of the vehicle pass persists in the identification system. For example, a short acceleration of the patrol jeep at a high engine speed resulted in a false detection at approximately the 40th frame in Figure 6.20 while the vehicle was still at a large distance from the UGS cluster. Wind gusts also often produce the ASR envelope dynamic and can result in high signal RMS energy values. If the classification procedure also fails to reject these false detection instances, a false identification result can be produced. To reduce the number of false detections the results of the identification procedure are fused with the results of acoustic localization and trajectory estimation (and other appropriate system components) at the level of information fusion.

During the entire experiment 84 vehicle pass instances were detected and identified. This number constitutes 18 class 1 vehicle pass, 21 class 2 vehicle pass, 21 class 3 vehicle pass, and 24 class 4 vehicle pass instances. Out of 84 instances 75 were correctly identified, which resulted in a 89.29% estimation accuracy. The confusion matrix for vehicle identification is presented in Figure 6.24. The main source of identification error is again situated with poor separability between class 3 and class 4. The fact that the chosen features separate the four classes well, leaving only one pair of classes poorly separated, suggests the application of a tree structure for class labeling. A set of features specific only to class 3 and class 4 has to be found, then the decision between these two classes can be made at the bottom of the tree structure after separating well distinguishable classes at the top of the tree.

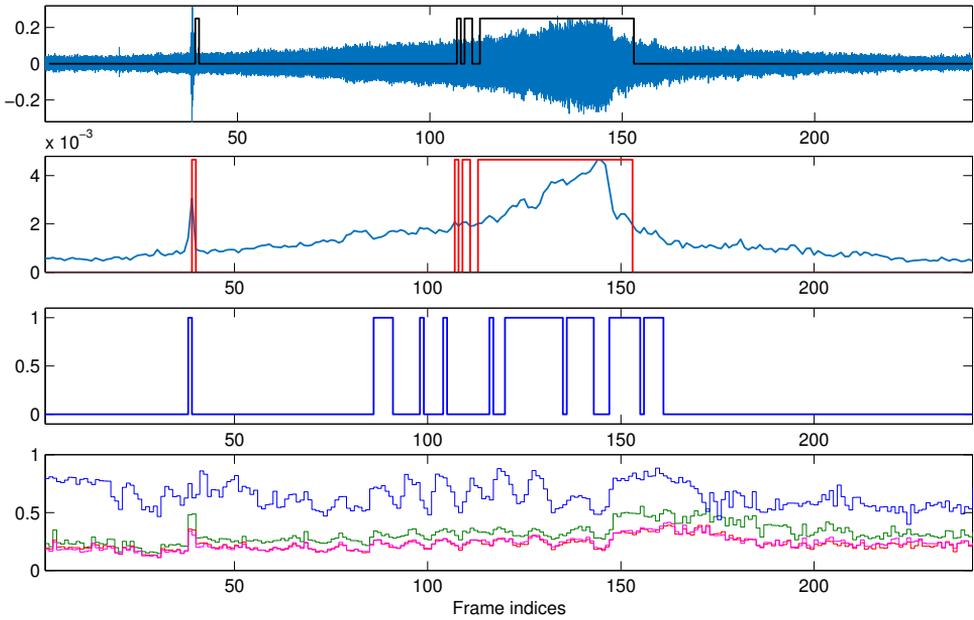


Figure 6.20: Example of patrol jeep (class 1) identification.

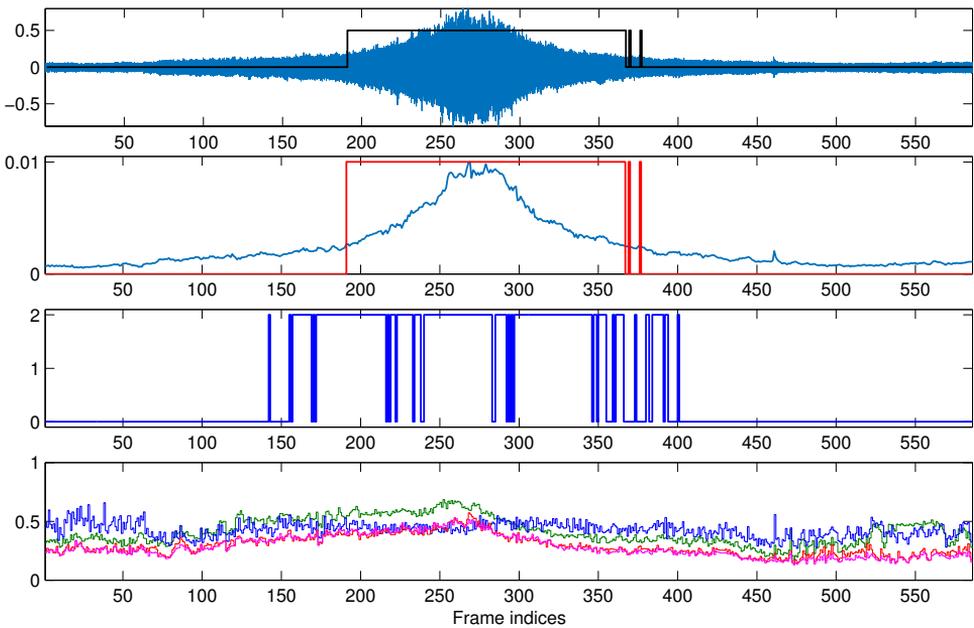


Figure 6.21: Example of personnel truck (class 2) identification.

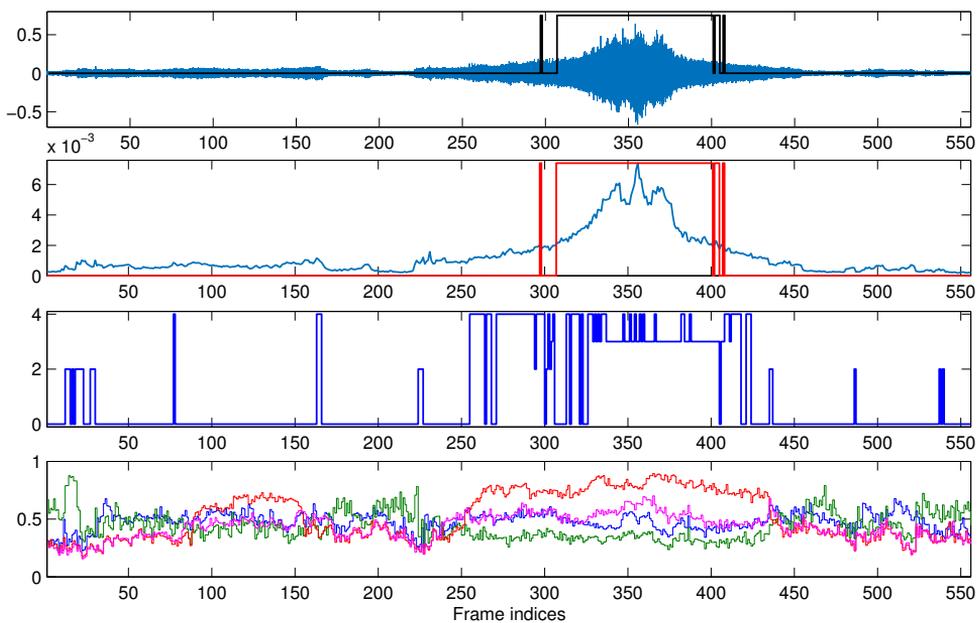


Figure 6.22: Example of medical utility truck (class 3) identification.

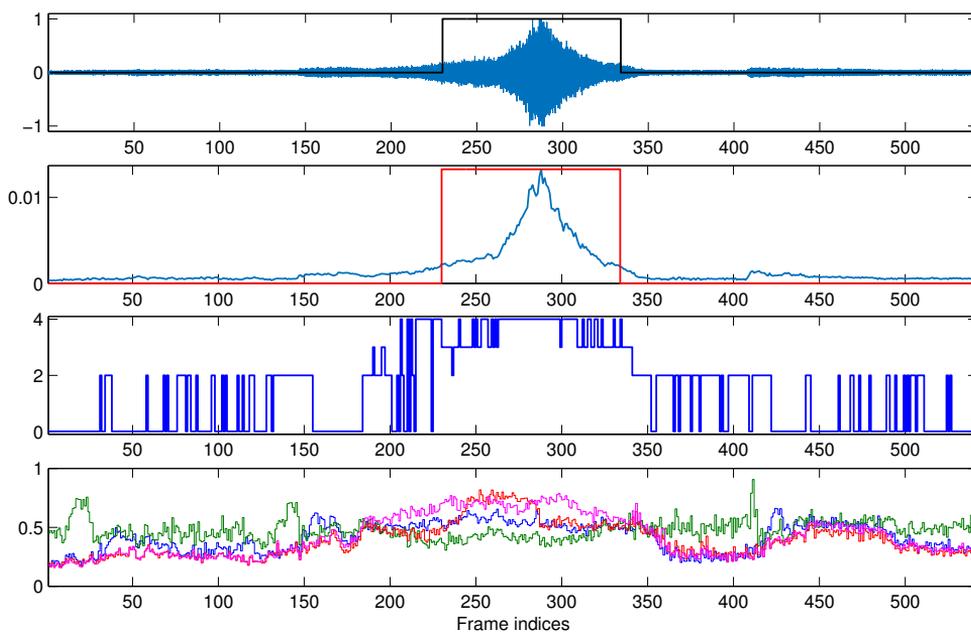


Figure 6.23: Example of armored personnel carrier (class 4) identification.

		Target class labels			
		1	2	3	4
Estimated class labels	1	17	1	1	0
	2	1	20	0	0
	3	0	0	18	4
	4	0	0	2	20

Figure 6.24: Confusion matrix of vehicle identification results. Accuracy equal to 89.29%.

6.3 Detection and Speed Estimation of Ground Troops

Detection and movement speed estimation of ground troops is the second considered topic of research for ISR system development. The considered approach is based on Passive Infrared (PIR) signal analysis. PIR sensors are usually used as binary state sensors for movement detection and are mainly utilized in smart houses and security systems [62]. For the application at hand the possibilities of PIR sensors are explored for movement speed estimation. PIR sensors can serve as less expensive alternatives to visible light and infrared cameras for detection of moving objects [50]. For application in ISR systems the UGS equipped with PIR sensors are placed in areas of potential enemy activity and estimate the speed of passing foot soldier (or other threat) movement. By fusing the movement information gathered from individual distributed UGS allows for estimating the direction of advancement and the number of moving targets.

6.3.1 PIR Signal Analysis for Movement Speed Estimation

A differential PIR sensor utilizes two infrared wavelength sensitive elements made of pyroelectric crystals. The crystals change their polarization when illuminated with thermal radiation. This, in turn, creates an electric potential between the surfaces of the crystals relative to the amount of radiation received. The generated electric charge slowly dissipates if there is no more change in the amount of thermal radiation in the FOV of the sensor. The Fresnel lens splits the sensor's FOV into sections and divides the sections between the two crystals such that two neighboring sections are never focused on the same crystal. This means that as an object moves from section to sec-

tion, one of the elements always outputs a voltage. Additionally the outputs of both the crystals are wired to opposite inputs of a differential amplifier to produce and amplify the difference of the two signals rather than the signals themselves. This eliminates the possibility of false movement detection events when both the sensor elements are illuminated with the same amount of radiation at the same time (e.g., in case of field wide light flashes). The configuration of the Fresnel lens defines the sensor's FOV. For the experiments situated with ground troop detection a lens having a 180° horizontal FOV and a $[0^\circ, 60^\circ]$ vertical FOV is used. The vertical FOV starts at 0° from the horizontal plane (assuming zero tilt of the sensor's aperture) and tends upwards. Such a lens reduces the risk of false detection caused by movement near the ground (e.g., small animals) if the sensor is placed sufficiently high above the ground level. Figure 6.25 presents the PIR sensors used in the experiments.



Figure 6.25: PIR sensors encased in a camouflaged chassis designed by Defendec Inc.

Signal Processing Steps

Signal processing is conducted on the voltage signal of the differential PIR sensor. The principle block-diagram of the proposed processing procedure is presented in Figure 6.26. The acquired signal is passed through initial signal processing (enclosed in green in Figure 6.26), stored in an input buffer and processed in a sliding window manner by the latter parts of the algorithm. The sliding window approach is applied because PIR is a weakly oscillating signal, and thus the frame length has to be considerably larger than, for example, in case of acoustic signals. Non-zero frame overlap ensures higher refresh rates at larger frame lengths.

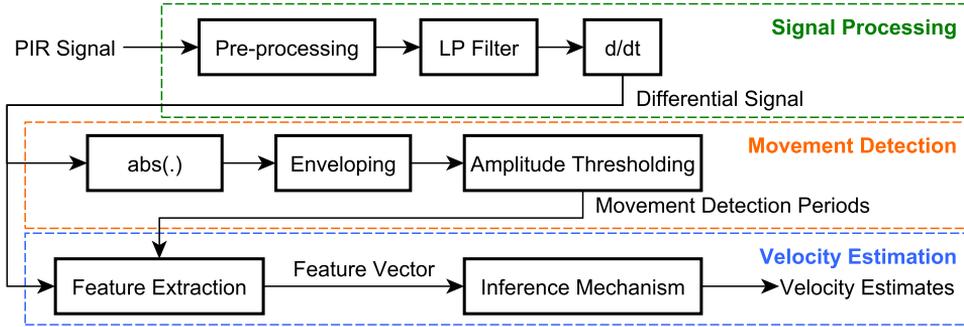


Figure 6.26: Block-diagram of applied PIR signal processing steps.

The first stage of signal processing in Figure 6.26 begins with raw digital signal pre-processing, which consists of eliminating the bias (DC voltage) and scaling the signal. To reduce the high-frequency ambient noise a Low Pass (LP) filter is applied. For the experiments a 5th order Butterworth Infinite Impulse Response (IIR) filter with the cutoff frequency of 10 Hz is used. After filtering the derivative of the signal is calculated as the first order finite difference $\Delta x[k] = x[k] - x[k - 1]$, where $x[k]$ and $x[k - 1]$ are the k -th and previous signal samples, respectively. Further signal analysis is performed on the signal derivative due to the fact that signal settling time after movement end is quite long — not less than 1–2 seconds for the specific sensor at hand. These settling time signal fluctuations are reduced in the derivative.

Movement Detection

The movement patterns of the PIR signal can easily be contaminated by ambient noise or insignificant movements. For this reason target movement detection is performed prior to object velocity estimation. Movement detection consists of computing the envelope of the signal derivative by applying the Hilbert transform (2.2), and applying a preset threshold to the envelope's average value. Low signal dynamics suggest to use a sliding average value of the envelope amplitude

$$\bar{x}[k] = \frac{1}{N_W} \sum_{k-N_W < i \leq k} x[i], \quad (6.1)$$

where N_W is the sliding window length in samples. During a detection event, envelope amplitude may decrease for a short amount of time. Such cases arise when the PIR signal is at its maximum possible amplitude and thus the derivative is close to zero. These cases are handled by performing an additional check of every detection event for short-term drops of envelope amplitude.

Movement Speed Estimation

For pattern recognition a set of time domain features is extracted and an inference mechanism is applied, which maps the feature vectors onto speed estimates. PIR is a weakly oscillating signal with lengthy perturbations (the frequencies of interest are well below 1 Hz). For this reason frequency domain analysis and correlation based pattern identification are inapplicable. The following features are extracted for the periods of detected movement:

1. Zero crossing interval (ZCI) — the time between two successive signal zero crossings

$$\text{ZCI} = k_{\text{ZC}(i)} - k_{\text{ZC}(i-1)}, \quad (6.2)$$

which can be considered a rough estimate of signal half-period. All other features are computed once per every ZCI.

2. Amplitude (AMP) — maximum of half-period absolute values

$$\text{AMP} = \max(|x[k_{\text{ZC}(i-1)}]|, \dots, |x[k_{\text{ZC}(i)}]|). \quad (6.3)$$

3. Gradient (GR) — value of half-period maximal increase rate

$$\text{GR} = \max(|\Delta x[k_{\text{ZC}(i-1)}]|, \dots, |\Delta x[k_{\text{ZC}(i)}]|). \quad (6.4)$$

4. Mean gradient (MGR) — mean value of half-period increase rate

$$\text{MGR} = \frac{1}{\text{ZCI}} \sum_{j=k_{\text{ZC}(i-1)}}^{k_{\text{ZC}(i)}} |\Delta x[j]|. \quad (6.5)$$

5. Mean square energy (MSE) — mean value of signal half-period squared amplitudes

$$\text{MSE} = \frac{1}{\text{ZCI}} \sum_{j=k_{\text{ZC}(i-1)}}^{k_{\text{ZC}(i)}} (x[j])^2. \quad (6.6)$$

For object speed estimation based on the extracted feature vector a fuzzy inference procedure is applied, which provides efficient non-linear mapping for a set of features. For fuzzy inference a Sugeno-type Fuzzy Inference System (FIS) is used, which is composed as an Adaptive Network-based Fuzzy Inference System (ANFIS). The ANFIS neuro-fuzzy architecture of the fuzzy system allows the FIS parameters to be trained using typical Artificial Neural Network (ANN) training algorithms, like backpropagation and conjugate gradient descent [72]. Prior to online operation, the ANFIS structure is trained in a supervised manner on a dataset of reference feature vectors of form (2.23), with the only difference being that instead of integer class labels real values of reference speed are used in the last column of the dataset.

6.3.2 Experimental Results

The experiments of passing person movement speed estimation were performed in a sparse forest environment. The PIR sensor was placed at a height of approximately 2 meters above ground with its aperture slightly tilted downward (approximately 10° – 15°). The sensor was targeted at a forest path situated approximately 5 meters away from the sensor’s position. During the experiment a person was passing the sensor’s FOV along that path, parallel to the sensor’s aperture, at different speeds. Three different speeds were measured: a moderate walking speed of approximately 1 m/s, a brick pace of approximately 2 m/s, and the light jogging speed of approximately 2.5 m/s. The PIR sensor used in the experiment was developed by Defendec Inc. specifically for the military applications. The sensor is powered by an internal battery and encased in a sturdy camouflaged chassis (see Figure 6.25). PIR signals were acquired by the Agilent U2354A DAQ with the sampling rate set to $f_s = 1$ kS/s.

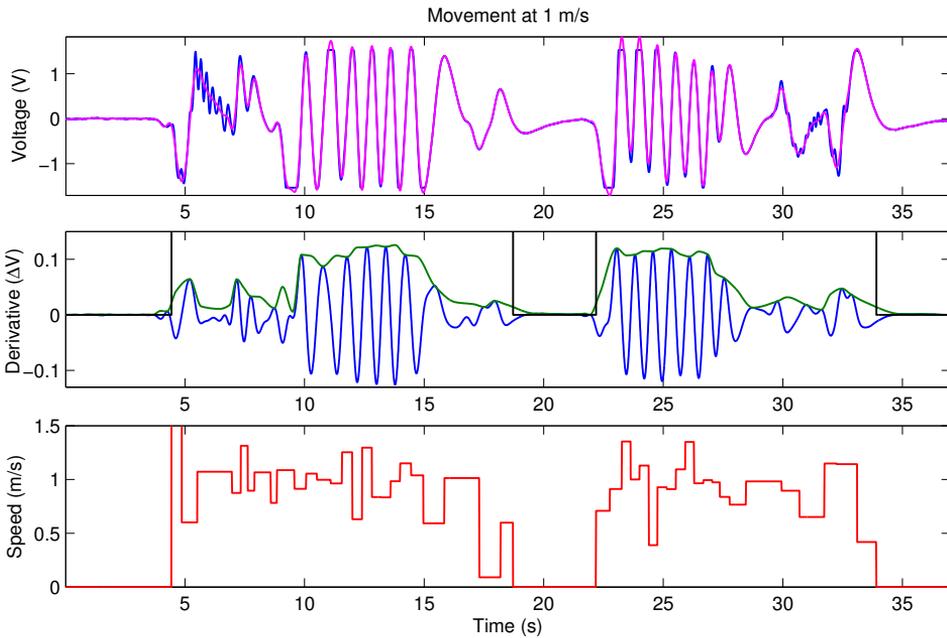


Figure 6.27: Speed estimation results for movement at the reference speed of 1 m/s.

Figures 6.27–6.29 present the results of movement speed estimation for the three considered movement speeds. In the top subplots of the figures blue lines denote the initial, and magenta lines denote the filtered PIR signals, respectively. The middle subplots present the signal derivatives (blue lines), signal derivative envelopes (green lines), and the intervals of movement detection (black lines). The bottom subplots present the movement speed estimates (red lines) produced for every movement detection inter-

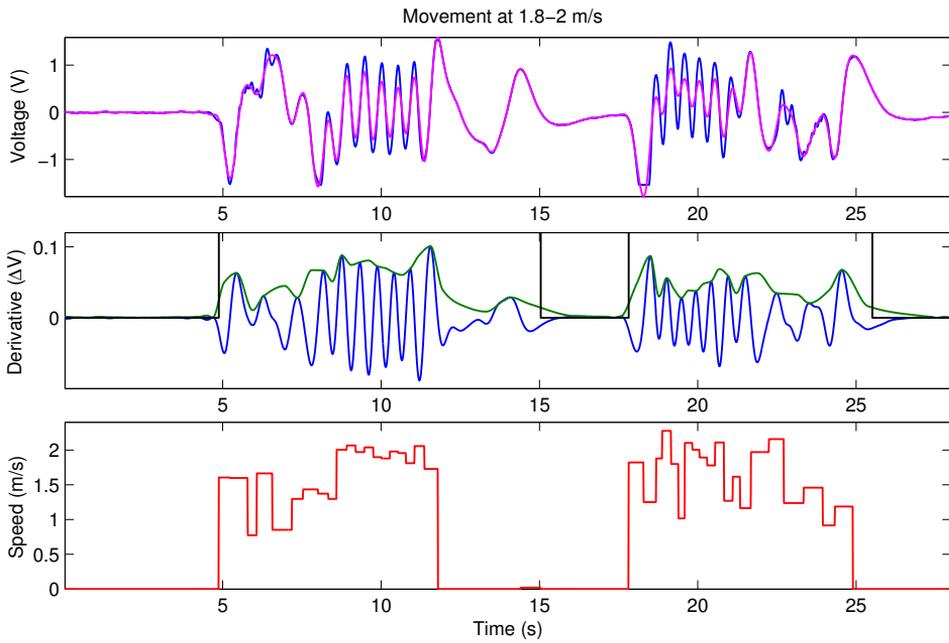


Figure 6.28: Speed estimation results for movement at the reference speed of 1.8–2 m/s.

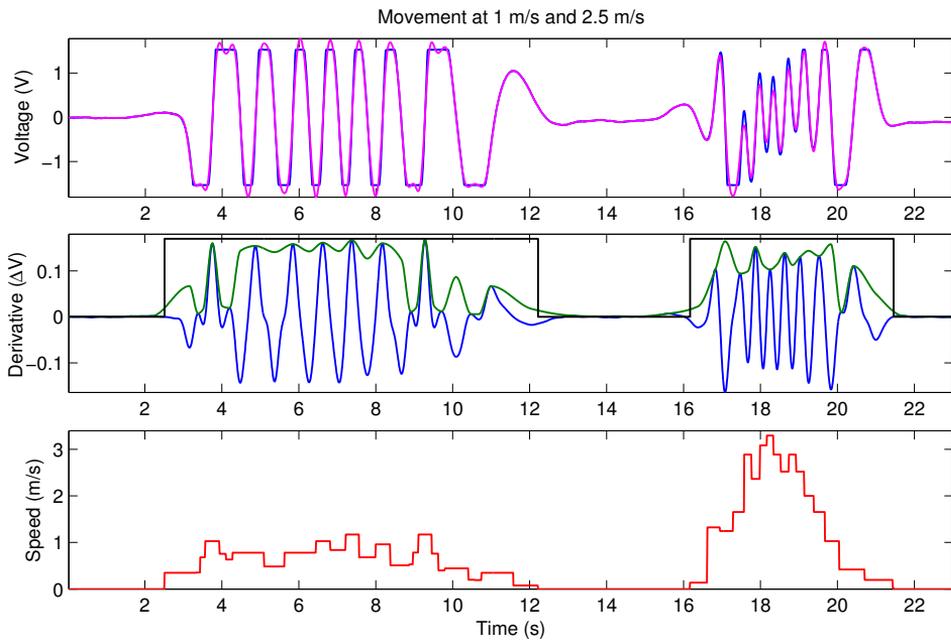


Figure 6.29: Speed estimation results for movement at two reference speeds of 1 m/s and 2.5 m/s.

val. The long signal settling intervals after movement end are presented in, e.g., Figure 6.27 at 16–21 seconds, Figure 6.28 at 12–16 seconds, and Figure 6.29 at 11–13 seconds. The amplitudes of the derivative and its envelope are significantly lower than the signal amplitude at these intervals, and thus the settling time intervals do not influence the procedure based on signal derivative analysis as much as they would if the procedure used the raw signals. The movement estimation procedure succeeds in correctly estimating the three considered movement speeds with most inaccurate estimates being produced at the beginning and at the very end of the movement detection intervals. The reason lies in the specific configuration of the used Fresnel lens. The lens divides the FOV into equal sectors, and when a person enters or leaves the FOV the corner sectors result in irregular oscillations of the PIR signal. Nevertheless, the results for the middle portions of the detection intervals coincide with the approximate reference speeds.

6.4 Gunshot Detection and Shooter Localization

Gunshot detection and shooter acoustic localization is the third considered topic of research for application in ISR systems. The development of shooter acoustic localization systems has continued for more than three decades. Numerous different gunshot detection and direction estimation systems are currently available for military applications of sniper and covert enemy force positioning, and are also used in law enforcement for gun violence reduction and forensics [3]. The devices currently available are generally standalone systems, composed of a single microphone array, e.g., the vehicle-mountable Boomerang system [98]. Individual gunshot detectors developed for military and law enforcement personnel [58, 59, 128] consist of compact shoulder-carried, helmet or uniform mounted sensors. Such individual systems increase local situation awareness, however, for large area coverage a different approach is required.

A distributed ISR system consisting of interconnected UGS is well suited for shooter localization because it expands UGS individual FOV and thus allows for greater area coverage. The state of the art in this area suggests either synchronous [129] or asynchronous [44] gunshot acoustic event detection and subsequent shooter localization based on UGS collective information. The majority of the proposed approaches is based on the analysis of the shockwave (SW) and muzzle blast (MB) produced by supersonic projectiles [102]. Most methods employ single-sensor UGS, identifying the gunshot events and estimating the shot geometry under different initial assumptions, e.g., the known caliber of the fired projectile [129] or a certain ballistic shockwave acoustic model [4]. However, initial assumption inconsistency and the presence of acoustic events not-of-interest (i.e., residual gunshot acoustic events

and various noise produced by other sources) may significantly reduce localization accuracy [8].

Employing multi-channel smart sensors for gunshot localization allows to additionally estimate the DOA of gunshot event acoustic waves. Knowing the DOA aids in acoustic event identification and allows to reduce the number of initial assumptions, which, in turn, makes the localization process more robust. The proposed method of shooter localization is based on gunshot event DOA and TDOA information. Each UGS in the localization WSN independently performs gunshot acoustic event detection, computes the DOA and marks the event occurrence times in its own local time. The fusion node gathers DOA and time information from all UGS, performs identification of SW and MB among events not-of-interest (NOI), calculates the TDOA between SW and MB, and estimates the shooter position based on the known UGS positions. The TDOA are calculated per each UGS and no cross-UGS delays are used, thus node synchronization is not required (however, soft synchronization is still needed to be able to distinguish between shot instances). Uniform Circular Arrays (UCA) are used for this application to ensure 360° horizontal FOV. DOA estimation is performed by the proposed method, discussed in Section 4.2.3. Steered Response Power (SRP-PHAT) applied over a reduced functional (discussed in Section 4.2.2) is used for comparison.

The proposed method was tested on signals acquired during three live shooting experiments. The first experiment was performed at a small outdoor shooting range with a shooter-target distance of 35 m. The second and third experiments were performed at a larger outdoor shooting range with a shooter-target distance of 100 m. Two UCA prototype versions were used in the experiments. Experiment 1 used the implementation employing six condenser microphones and an exterior Data Acquisition Device (DAQ), discussed in Section 4.2.5. Experiments 2 and 3 used the implementation employing six MicroElectroMechanical Systems (MEMS) microphones and a BeagleBone Black (BBB) as a DAQ and processing unit, discussed in Section 4.5.2. These two implementations are further referred to as prototype 1 and prototype 2, respectively.

6.4.1 Shooter Localization Preliminaries

In this work a planar gunshot acoustic event geometry is adopted for the proposed approach to shooter localization (i.e., the sensors and the trajectory of the traveling bullet are situated in the horizontal plane). A straight bullet trajectory is assumed for simplicity, not accounting for various effects considered in exterior ballistics [36].

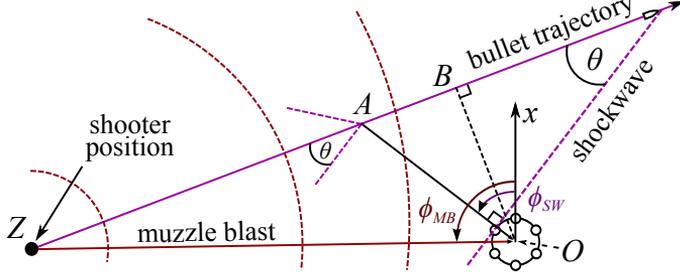


Figure 6.30: Gunshot acoustic event geometry in the horizontal plane.

Gunshot Acoustic Event Geometry

Figure 6.30 portrays the acoustic events produced by a gunshot at point Z as observed at point O . A gunshot is characterized by the shockwave (SW), produced by a supersonic projectile, and the muzzle blast (MB) of the fired weapon. SW produces a conical wavefront at an angle θ to the bullet's trajectory. The angle θ depends on the speed of sound c in air and the bullet velocity v :

$$\theta = \sin^{-1} \frac{c}{v}. \quad (6.7)$$

The waves of MB, on the other hand, propagate spherically at the speed of sound c in all directions.

The initial bullet velocity is equal to the muzzle velocity v_0 (i.e. the velocity at which the bullet leaves the muzzle of a gun), which depends on the bullet caliber and cartridge type and can be approximated for different firearm types [36]. Bullet velocity v decreases with flight distance due to air friction. It can be expressed as a function of traveled distance d_f as

$$v(d_f) = (v_0^\eta - 2\eta C_b^{-1} d_f)^{1/\eta}, \quad (6.8)$$

where C_b is a ballistic constant, which depends on the bullet's type, and η is the exponent value, usually set at 0.5. In this work the function (6.8) is assumed to be unknown and the bullet velocity is rather estimated using the procedure described in Section 6.4.4. For small firearms (e.g. rifles) the decrease in the $v(d_f)$ curve can be considered linear and ultimately insignificant for the travel distance of 100–200 m [36]. Thus the bullet velocity is fixed as a range-invariant parameter v . The speed of sound in air c , on the other hand, depends on the ambient temperature and is calculated using (2.30).

At line-of-sight, the sensor at point O detects MB at the time

$$t_{MB} = t_{shot} + \frac{d_{Z,O}}{c}, \quad (6.9)$$

where t_{shot} is the time of shot, and $d_{Z,O} = \|Z - O\|$ is the Euclidean distance between points Z and O . Acoustic waves of SW originate from the bullet itself

and not from the muzzle. SW travels outwards from the bullet's trajectory and is approximated as a planar wavefront in the horizontal plane. As the bullet has reached point A at speed v , the SW wavefront propagates from point A at speed c and reaches point O at the time

$$t_{SW} = t_{shot} + \frac{d_{Z,A}}{v} + \frac{d_{A,O}}{c}. \quad (6.10)$$

Point A here is such a point on the bullet's trajectory from where SW will travel directly to point O at an angle θ relative to the bullet's trajectory (see Figure 6.30).

The TDOA between SW and MB acoustic events can then be expressed as

$$\Delta t = t_{MB} - t_{SW} = \frac{d_{Z,O}}{c} - \frac{d_{Z,A}}{v} - \frac{d_{A,O}}{c}. \quad (6.11)$$

The shortest distance from the sensor at point O to the bullet's trajectory ($d_{O,B}$ in Figure 6.30) is called the miss distance. Whether Δt is positive depends on the bullet's velocity and the miss distance. If a shot is fired from a rifle (average bullet velocity near or greater than mach 2) in the sensor's direction with the miss distance small enough, then Δt is expected to be positive, as SW will most likely reach the sensor before MB.

The DOA of MB and SW for the sensor at point O are defined in the horizontal plane as azimuth values ϕ_{MB} , ϕ_{SW} , relative to the sensor's local coordinate system (x -axis in Figure 6.30). Here the azimuth ϕ_{SW} is the angle of incidence of a wavefront traveling from point A and ϕ_{MB} is the angle of incidence of a wavefront traveling from point Z .

Problem Statement

Knowing t_{SW} and t_{MB} , gunshot acoustic localization may be performed by estimating the angle θ and the miss distance. Angle θ may be estimated by applying a shockwave acoustic model to the duration of the SW signal [4], or calculated under known bullet caliber assumption [129]. Then, using multiple measurements of t_{SW} and t_{MB} from K synchronous single-sensor UGS, the miss distances can be approximated and point Z located via a bounded search procedure [129]. UGS synchronization plays a crucial role in such approaches and heavily influences the bound parameters of the bounded search procedure, as well as the overall localization accuracy [89]. Alternatively, using multiple measurements from K asynchronous single-sensor UGS and assuming θ to be known, it is possible to iteratively estimate MB DOA, miss distances, the bullet's trajectory and, consequently, point Z via a multistage optimization procedure [44]. If UGS clocks are sufficiently synchronized, a mutual reference moment t_{shot} can be established for all UGS via (6.10), and Z can be estimated by multilateration, using time delays t_{MB} from (6.9).

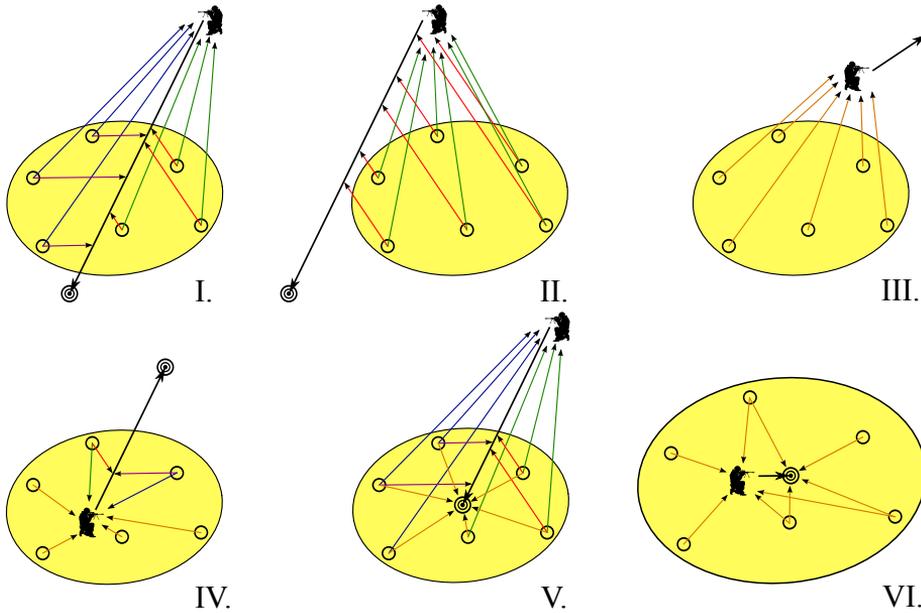


Figure 6.31: Six considered fundamental gunshot scenarios.

Unfortunately, if gunshot acoustic patterns include NOI events, such as reflections and target hit (TH) noise, MB cannot be unambiguously selected from numerous events following SW. Consider, for example, Figure 6.31, which presents six fundamental gunshot scenarios. These scenarios represent the following events:

- the bullet passes through the UGS cluster (I),
- the bullet passes beside the UGS cluster (II),
- a shot is fired away from the UGS cluster (III),
- a shot is fired from inside the UGS cluster (IV),
- the bullet hits the target in the vicinity of the UGS cluster (V),
- a shot is fired and the target is hit inside the UGS cluster (VI).

Scenarios I–III do not contain NOI events and are most commonly considered in the majority of state of the art approaches. In Scenarios I and II the bullet either passes through or beside the UGS cluster and no TH is detected. The localization is then performed using pure SW and MB readings, which are denoted in Figure 6.31 by arrows pointing from one or both sides towards the bullet’s trajectory, and arrows pointing towards the shooter’s position, respectively. Scenario III assumes that only MB are detected. This makes it

a trivial localization problem, which can be solved using conventional localization methods, e.g., multilateration. Scenarios IV–VI, on the other hand, assume the presence of NOI events and the masking effect. Here either SW or MB may be corrupted or masked by TH (Scenario V), or either SW or MB may be corrupted or masked by each other (Scenarios IV and VI). Furthermore, NOI events such as reflections and background noise may be present for all scenarios and must be accounted for accordingly. NOI events can be eliminated by identifying MB and SW by their acoustic properties [87] or applying statistical assignment [112], however, these methods do not solve the masking problem.

The shooter localization algorithm presented in this work assumes Scenario V of Figure 6.31, where the UGS form a look-out perimeter around the potential target, which is very likely to be hit inside or near the UGS cluster. Scenario V implies that either SW or MB may be corrupted or masked by TH, and UGS situated behind the target may not detect SW altogether. As Scenario I is a special case of Scenario V (the bullet passes through the cluster and no TH is detected), the localization rules intended for Scenario V will also be applicable to Scenario I. The study also considers several acoustic event detection problems situated with varying shot range and influence of NOI events. At a sufficient shot range the TDOA between SW and MB acoustic transients makes the events well distinguishable [32]. In one of the considered experiments a short range case is studied, where event separation is not straightforward due to short TDOA. The proposed detection method accounts for all gunshot acoustic events, as the MB signal transient is not guaranteed to strictly follow the one of SW.

6.4.2 Proposed Approach to Shooter Localization

The proposed approach is intended for application in WSN with a dynamic ad-hoc topology, which implies node synchronization complications and a varying number of active nodes at any given time. Thus the focus lies on an asynchronous size-invariant solution. Assuming that the WSN consists of K UGS, equipped with acoustic sensor arrays, and one or several information fusion nodes, the proposed approach consists of the following steps:

1. Each UGS detects a gunshot, separates its acoustic events, marks the time and computes a DOA value per each event.
2. Per each detected shot, each UGS sends an information packet to the fusion node containing its position, steering angle and acoustic event parameters $\{\mathbf{x}, \alpha, \mathbf{t}, \Phi\}$.
3. The fusion node performs event identification and shooter localization based on the information provided by active UGS.

The packet of UGS $k = 1, \dots, K$ contains: UGS coordinates $\mathbf{x}_k = (x_k, y_k)$; UGS steering angle α_k ; gunshot event times $\mathbf{t}_k = [t_1, \dots, t_{E_k}]$; event DOA $\Phi_k = [\phi_1, \dots, \phi_{E_k}]$, where E_k is the number of detected events of the k -th UGS. As each UGS operates in its own coordinate system, the steering angle α_k is used to specify UGS local coordinate system steering from a global zero-rotation angle (which is defined by Earth’s magnetic north).

While receiving packets from the UGS, the fusion node maintains a validity interval beginning at the moment of arrival of the first packet. This way the expired packets or the ones corresponding to another shot are dealt with separately. Data validation is discussed in Section 5.3 along with other communication-related problems. Uniform Circular Arrays (UCA) are chosen for the array configuration because they provide full horizontal FOV with a simple geometry. Each UCA consists of $M = 6$ microphones, which are mounted in circular shells with a radius of $r = 7.5$ cm (prototype 1) and $r = 10$ cm (prototype 2).

6.4.3 Gunshot Acoustic Event Detection and Separation

Gunshot acoustic event detection for a general case (i.e., comprising all scenarios of Figure 6.31) is an intricate task. Amplitude-based methods are well suitable in Scenarios IV and VI, where both SW and MB are detected inside the UGS cluster as high-energy transients and are, therefore, distinguishable from background noise. The same holds for Scenarios I–III and V, if the shot range is short enough for MB to be detected. Otherwise MB can have a small enough amplitude for it not to pass signal analysis or be masked by background noise. Another approach is situated with identifying SW and MB by the shape of their acoustic signals. For example, the N-shaped pattern of the SW transient may be examined in the time domain [4], or SW and MB may be distinguished from reflections by applying classification [5, 87]. This may work well for Scenarios I–III, where no TH or overlapping events occur, and the task lies only in eliminating reflections. For Scenarios IV–VI and specifically Scenario V these methods are not guaranteed to perform well.

Shooter distance plays an important role in acoustic event separation as well. In case of a significantly short shot distance, acoustic event separation poses a challenge due to an extremely short TDOA between SW and MB [56]. Figure 6.32 presents an example of a normalized gunshot signal acquired at 16.2 m away from the shooter. The multi-channel acoustic signal is presented in the top subplot of Figure 6.32. The event detection results in the lower subplot consist of results of peak detection, denoted by red stems, and event establishing peaks, denoted by green stems. Figure 6.32 shows that the TDOA between SW (at 4 ms) and MB (at 11 ms) is only 7 ms in this close range case. Figure 6.34, on the other hand, portrays a normalized gunshot signal acquired at 97.5 m away from the shooter. Here the TDOA between

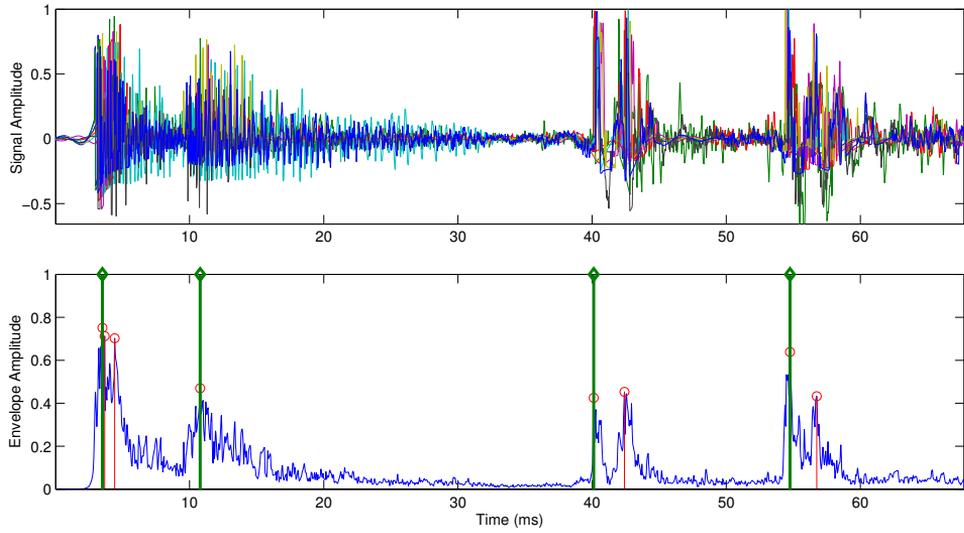


Figure 6.32: Gunshot acoustic components acquired by UGS S_3 during Experiment 1 at $f_s = 48 \text{ kS/s}$ (top). Collective envelope and times of detected events (bottom).

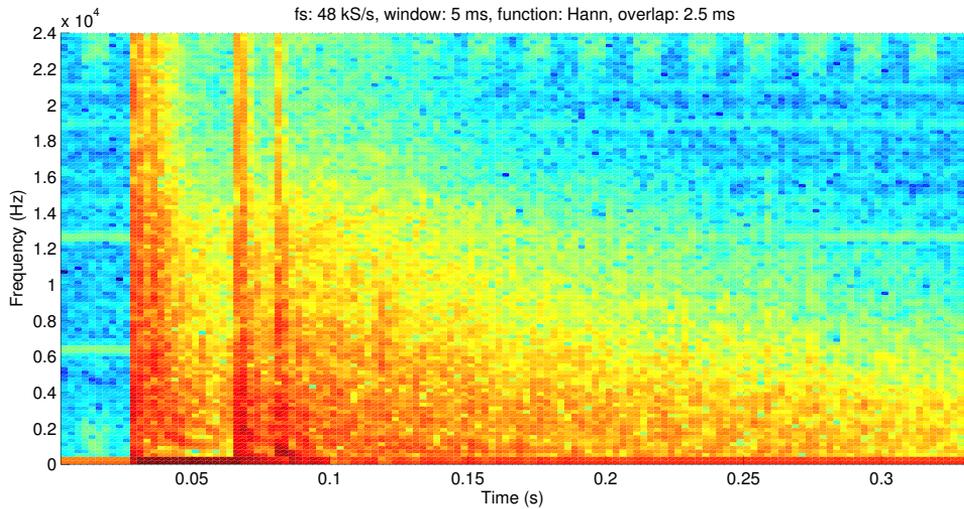


Figure 6.33: Spectrogram of the gunshot signal presented in Figure 6.32. Acoustic components of Figure 6.32 are located at approximately 30–95 ms.

SW (at 25 ms) and MB (at 150 ms) is already 125 ms, twice as long as the whole gunshot signal of Figure 6.32. If the detection algorithm treats the closely spaced events as a single event, MB may be lost in the SW transient. On the other hand, analyzing every closely spaced rising signal will waste computational resources and produce a large number of unwanted results.

Another problem lies in separating gunshot instances in case of burst-mode and automatic fire at close ranges. Consider Figure 6.32, where the TDOA between SW and MB is 7 ms with post-blast events (i.e., TH and reflections) starting to occur at the 40th millisecond. Neglecting these post-blast events may seriously harm the detection process in case of burst-mode fire. For example, an AK-47 in burst mode can fire 600 rounds per minute and an M-4 fires at 950 rpm, which constitutes approximately 1 bullet every 100 ms and 63.2 ms, respectively. In this case consecutive SW and MB may be mistaken for post-blast events, and vice versa for a single shot case.

In the proposed approach to acoustic event detection and separation both short (20–40 meters) and medium (100–200 meters) shooter distances are considered. All arising acoustic events are established by the following procedure. First, a collective envelope is computed using the signals from all microphones. At sampling time k , the envelope of samples $x_1[k], \dots, x_M[k]$ is

$$s_{env}[k] = \max(|x_1[k]|, \dots, |x_M[k]|). \quad (6.12)$$

Event detection is performed on the differential collective envelope

$$\Delta s_{env}[k] = s_{env}[k] - s_{env}[k - 1]. \quad (6.13)$$

The differential envelope $\Delta s_{env}[k]$ is passed through peak detection. Peaks lying within an interval of $t_W/2$ seconds from one another are grouped together, and one (the first) peak per event is chosen. The value t_W here is the predefined length of event window. An example of separation of four events is presented in Figure 6.32 (lower subplot) and the example of separation of eight events — in Figure 6.34 (lower subplot). One frame of duration t_W is retrieved from the multi-channel signal buffer per each event peak such, that event beginning is included in the frame and adjacent events are strictly separated. This means that if the events do not overlap, the event signal is confined from the beginning of its envelope rise for the duration of t_W ; if the events do overlap (i.e., event establishing peaks are less than $t_W/2$ seconds apart), the first event is windowed leftward from the beginning of the second event, and the second event is windowed rightward from its beginning.

Event identification is performed during the data fusion stage. As NOI events can also be transient in nature, they are hard to identify during event detection. Frequency analysis does not offer a straightforward solution either, as NOI events possess highly uniform spectral densities as well as SW

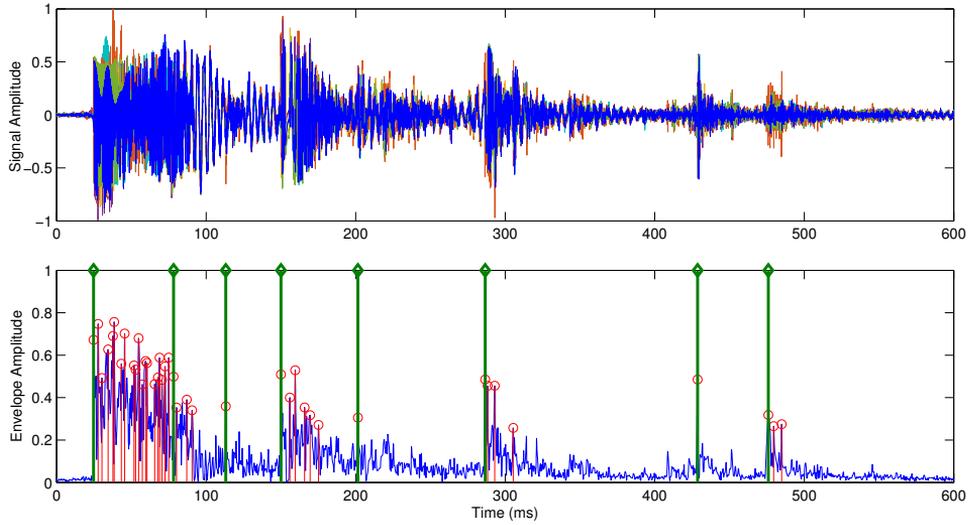


Figure 6.34: Gunshot acoustic components acquired by UGS S_1 during Experiment 3, shooter position 1, at $f_s = 20$ kS/s (top). Collective envelope and times of detected events (bottom).

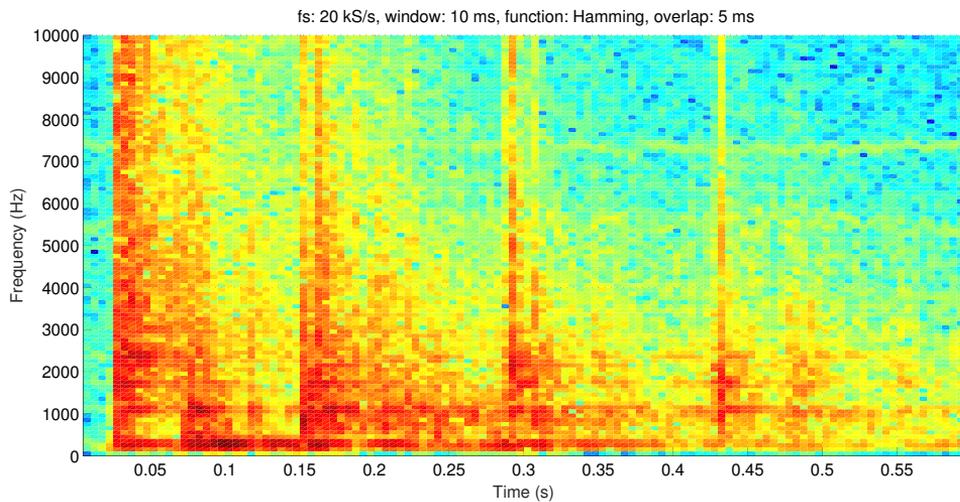


Figure 6.35: Spectrogram of the gunshot signal presented in Figure 6.34.

and MB (see Figure 6.33 and Figure 6.35). Figure 6.34 also portrays overlapping events at 25–110 ms. Here SW is overlapped with its own ground reflection, which results in two additional peaks being detected before MB. In such situations the identification of SW by its shape and duration will likely produce inaccurate results.

Direction of Arrival Estimation

During shot detection and acoustic event separation the k -th UGS produces E_k multi-channel signal frames of length $N = f_s t_W$, where f_s is the sampling frequency. These E_k event frames are ordered by the time of their occurrence $\mathbf{t}_k = [t_1, \dots, t_{E_k}]$, and a separate DOA estimate is then computed per each frame by applying SRP-PHAT (for reference) and the proposed DOA estimation method. Horizontal evaluation of the FOV is performed by SRP-PHAT by applying (4.11) and (4.12), as discussed in Section 4.2.2. For the horizontal evaluation by the proposed DOA estimation method the procedure discussed in Section 4.2.3 is applied only to the set of UCA microphone pairs (4.15). As a result, a set of azimuth estimates $\Phi_k = [\phi_1, \dots, \phi_{E_k}]$ is obtained and transmitted to the fusion node along with other data for further processing.

6.4.4 Information Fusion and Shooter Localization

As a result of shot detection, the fusion node receives K packets $\{\mathbf{x}, \alpha, \mathbf{t}, \Phi\}_k$, $k = 1, \dots, K$, where K is the number of active UGS, which have detected at least one gunshot event. The number of detected events E_k may vary per UGS. The DOA estimates Φ_k are first steered to the global coordinate system, $\Phi_k = \Phi_k - \alpha_k$, and information fusion is then conducted in the following steps:

- identification of DOA corresponding to SW and MB events,
- estimation of shot geometry,
- estimation of miss distance and distance to shooter for each active UGS,
- shooter localization.

DOA Concurrence

Consensual DOA are established by analyzing all $\Phi = \{\Phi_k \mid k = 1, \dots, K\}$ estimates. To locate coherent estimates, the angular values in Φ are clustered in a manner, similar to the one described in Section 4.1.2. If coherent estimates exist, we obtain P clusters Φ_p , $p = 1, \dots, P$, each containing n_p estimates ϕ_i , $i = [1, n_p]$.

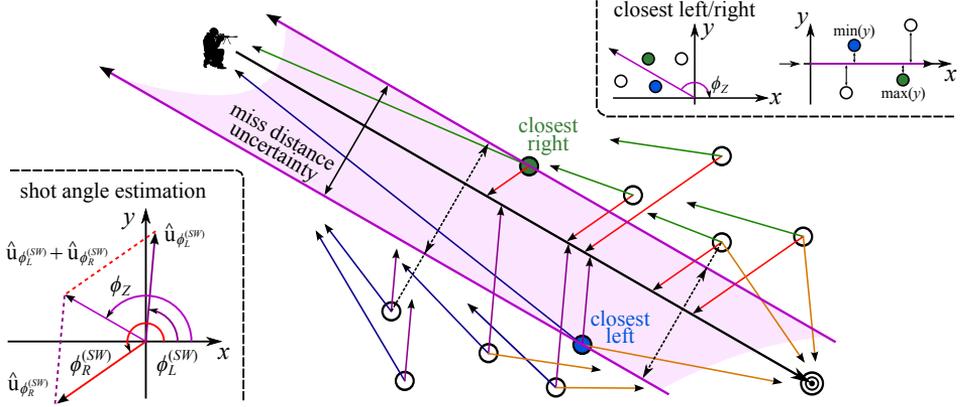


Figure 6.36: Shot angle and miss distance uncertainty interval estimation by UGS groups situated to the left and to the right from the bullet's trajectory.

Assuming Scenario V (Figure 6.31), Φ_p will contain DOA corresponding to the SW events detected by UGS situated to the left and to the right from the bullet's trajectory, DOA corresponding to MB events, and DOA of NOI events, e.g., TH, various reflections and noise. The DOA of SW vary only slightly (due to DOA estimation error and natural variance of angle θ) and do not depend on the distance to shooter. DOA corresponding to MB, on the other hand, depend on the distance to shooter and UGS cluster dimensions. If the distance to shooter is significantly larger than the width of the UGS cluster, MB DOA will be roughly parallel for all UGS. At a closer distance the UGS situated on the opposite sides of the bullet's trajectory will have their MB DOA significantly skewed towards the trajectory in the shooter's direction. A principal diagram of consensual DOA for Scenario V is presented in Figure 6.36.

Event Identification and Shot Geometry Estimation

To reduce the error of individual DOA estimates, event identification is performed on the mean values of clusters Φ_p : $\bar{\phi}_p = \frac{1}{n_p} \sum \Phi_p$, $p = 1, \dots, P$. To identify SW DOA, all $\bar{\phi}_p$ are analyzed pairwise. For each pair $\bar{\phi}_i, \bar{\phi}_j$, $i = [1, P-1]$, $j = [i+1, P]$, a central angle ϕ_Σ is first calculated as the angular component of the sum of their corresponding unit vectors $\hat{u}_{\bar{\phi}_i} + \hat{u}_{\bar{\phi}_j}$ (see Figure 6.36). SW DOA are then identified under the assumptions that SW events are detected first, and at least one SW DOA was detected to the left and to the right from the bullet's trajectory. Thus $\bar{\phi}_p$ are searched for

such $\bar{\phi}_i, \bar{\phi}_j$, that meet all the following conditions:

$$\begin{aligned} \frac{\pi}{2} - \varphi_{\max}^{(SW)} &< |\phi_{\Sigma} - \bar{\phi}_i| < \frac{\pi}{2} - \varphi_{\min}^{(SW)}, \\ \frac{\pi}{2} - \varphi_{\max}^{(SW)} &< |\phi_{\Sigma} - \bar{\phi}_j| < \frac{\pi}{2} - \varphi_{\min}^{(SW)}, \\ \forall \text{ind}_{t_k} (t_{\phi_k} \mid \phi_k \in \Phi_i) = 1, \forall \text{ind}_{t_k} (t_{\phi_k} \mid \phi_k \in \Phi_j) = 1. \end{aligned} \quad (6.14)$$

The term ind is defined as the operation, which determines the index of a specific element in a vector of values. $(\varphi_{\min}^{(SW)}, \varphi_{\max}^{(SW)})$ is the interval of SW propagation angle θ expected values (refer to Section 6.4.1), accounting for variance and measurement error. For example, if $\theta \approx 25^\circ$ and a $\pm 5^\circ$ measurement deviation are expected, this interval is set to $(\frac{\pi}{9}, \frac{\pi}{6})$. If the conditions (6.14) are met, then $\bar{\phi}_i, \bar{\phi}_j$ and $\phi_k \in \Phi_i \cup \Phi_j$ are labeled $\bar{\phi}_i^{(SW)}, \bar{\phi}_j^{(SW)}$ and $\phi_k^{(SW)}$, respectively. For $\bar{\phi}_i^{(SW)}, \bar{\phi}_j^{(SW)}$, conditions (6.14) also imply that they were measured on the opposite sides of the bullet's trajectory. Consequently, their central angle ϕ_{Σ} is adopted as the shot angle ϕ_Z estimate (i.e., the angle, at which the bullet travels towards the UGS cluster in Figure 6.36).

Having estimated ϕ_Z , the UGS S_k that have detected SW are placed into either the "left" or "right" groups G_L, G_R :

$$\begin{aligned} \phi_k^{(SW)} < \phi_Z &\Rightarrow S_k \in G_L, \\ \phi_k^{(SW)} > \phi_Z &\Rightarrow S_k \in G_R. \end{aligned} \quad (6.15)$$

To estimate the miss distance, $S_k \in G_L \cup G_R$ closest to the bullet's trajectory are first located. This is done by steering the S_k coordinates \mathbf{x}_k by ϕ_Z towards the x -axis around the UGS common spatial centroid $\bar{\mathbf{x}} = \frac{1}{K} \sum \mathbf{x}_k$ as

$$\begin{bmatrix} x'_k \\ y'_k \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} + \begin{bmatrix} \cos(\phi_Z) & \sin(\phi_Z) \\ -\sin(\phi_Z) & \cos(\phi_Z) \end{bmatrix} \begin{bmatrix} x_k - \bar{x} \\ y_k - \bar{y} \end{bmatrix}. \quad (6.16)$$

Then, as portrayed in Figure 6.36, "closest left" and "closest right" UGS \check{S}_L, \check{S}_R are defined as

$$\begin{aligned} \check{S}_L &= S_i, i = \text{ind min}(y'_k), S_k \in G_L, \\ \check{S}_R &= S_j, j = \text{ind max}(y'_k), S_k \in G_R, \end{aligned} \quad (6.17)$$

and the distance between them, perpendicular to the shot angle, $\phi_Z - \frac{\pi}{2}$, is referred to as the miss distance uncertainty interval. Inside this interval the exact miss distance cannot yet be established at this point. It is approximated at a later stage of shooter localization.

To identify the DOA corresponding to MB events, $\bar{\phi}_p$ are searched for such $\bar{\phi}_i, i = [1, P]$, that meet the following condition:

$$|\phi_Z - \bar{\phi}_i| < \varphi_{\max}^{(MB)}, \bar{\phi}_i \neq \bar{\phi}_i^{(SW)}. \quad (6.18)$$

During MB DOA identification preference is given to $S_k \in G_L \cup G_R$, because SW detection implies that the bullet has passed the UGS, and thus TH will likely not come from the same direction as MB. This way TH DOA will most certainly be avoided. NOI events caused by different noise, on the other hand, are seldom acquired with consistent DOA by a significant number of UGS, and thus their corresponding clusters Φ_p are significantly smaller, and the estimates are more dispersed. At this stage they are easily separable from the estimates considered for the MB label. Incidental acoustic sources arising in the FOV can be identified and excluded from analysis by general acoustic monitoring and source tracking techniques previously discussed in this work. As a result of MB DOA identification, $\phi_k \in \Phi_i$ meeting condition (6.18) are labeled as $\phi_k^{(MB)}$.

Estimation of Distance to Shooter and Shooter Localization

Having identified $\phi_k^{(SW)}$ and $\phi_k^{(MB)}$, $k = 1, \dots, K$, where K is now the number of UGS with both detected events, it is possible to accurately compute the TDOA between MB and SW, Δt_k as

$$\Delta t_k = t_{k,i} - t_{k,j}, \quad (6.19)$$

$$i = \underset{\Phi_k}{\text{ind}} \left(\phi_k^{(MB)} \right), \quad j = \underset{\Phi_k}{\text{ind}} \left(\phi_k^{(SW)} \right).$$

Based on Δt_k and the k -th UGS miss distance estimate $\hat{d}_{miss}^{(k)}$, it is possible to assess the distance to shooter from the k -th UGS using a closed form solution, proposed by Sallai et al. [129]:

$$d_{S_k, Z} = \frac{1}{2(c^4 - v^4)} \left(A - 2\sqrt{B} \right), \quad (6.20)$$

where

$$\begin{aligned} A &= -2v^3 \hat{d}_{miss}^{(k)} \sqrt{v^2 + c^2} - 2\Delta t_k c^3 v^2 \\ &\quad + 2c^2 \hat{d}_{miss}^{(k)} v \sqrt{v^2 + c^2} - 2\Delta t_k c v^4, \\ B &= -2c^4 v^4 \left(\hat{d}_{miss}^{(k)} \right)^2 + 2(\Delta t_k)^2 c^6 v^4 \\ &\quad + 2(\Delta t_k)^2 c^4 v^6 - 2c^7 \hat{d}_{miss}^{(k)} \Delta t_k v \sqrt{v^2 + c^2} + c^8 (\Delta t_k)^2 v^2 \\ &\quad + 2c^8 \left(\hat{d}_{miss}^{(k)} \right)^2 + 2v^5 \hat{d}_{miss}^{(k)} \sqrt{v^2 + c^2} \Delta t_k c^3. \end{aligned}$$

Projectile velocity can be empirically estimated by inverting (6.7) as $\hat{v} = c / \sin(\hat{\theta})$ and applying it to $\hat{\theta}$, which is computed as $\hat{\theta} = \bar{\phi}_L^{(SW)} - (\pi - \phi_Z)$, where $\bar{\phi}_L^{(SW)}$ is the mean value of the set of estimates, labeled as SW and belonging to the left group. For $\hat{d}_{miss}^{(k)}$ estimation a minimal and maximal miss

distance interval $\left[d_{\min}^{(k)}, d_{\max}^{(k)} \right]$ is first established. For every S_k , its minimal miss distance $d_{\min}^{(k)}$ spans from its coordinates \mathbf{x}_k in the direction towards the bullet's trajectory (perpendicularly to ϕ_Z) up to the point, where miss distance ambiguity starts. The maximal distance $d_{\max}^{(k)}$ spans further, up to the point, where miss distance ambiguity ends (dashed line spanning from an UGS of the right group in Figure 6.36).

Equation (6.20) suggests that $d_{S_k, Z}$ rises with $\hat{d}_{miss}^{(k)}$, therefore, $S_k \in G_L$ will give larger, and $S_k \in G_R$ — smaller estimates if $\hat{d}_{miss}^{(k)}$ is at the ambiguity start of group G_R , and vice versa if it is at the ambiguity start of G_L . So, the ambiguity interval is iteratively passed from $d_{\min}^{(k)}$ to $d_{\max}^{(k)}$ with a step of d_{step} , the miss distances for K UGS are estimated as $\hat{d}_{miss}^{(k)} = d_{\min}^{(k)} + i \cdot d_{step}$, and distance estimates to shooter $\hat{d}_{S_k, Z}(i)$ at each step are obtained using (6.20). A shooter position estimate $\hat{Z}_k(i)$ is computed per each UGS using \mathbf{x}_k , $\phi_k^{(MB)}$ and $\hat{d}_{S_k, Z}(i)$. The fitness of $\hat{Z}_k(i)$ point estimates is measured by their average distance from their common centroid $\bar{Z}(i)$:

$$f_{fit}(i) = \frac{1}{K} \sum_{k=1}^K \left\| \bar{Z}(i) - \hat{Z}_k(i) \right\|. \quad (6.21)$$

The minimum of the fitness function f_{fit} indicates the miss distance estimates, which are closest to the actual value, $\hat{d}_{miss}^{(k)} \simeq d_{miss}^{(k)}$, and the final shooter's position estimate is selected as $\hat{Z} = \bar{Z}(i)$, where $i = \arg \min (f_{fit}(i))$.

6.4.5 Experimental Results

The proposed shooter localization approach is tested on real gunshot signals acquired during three separate live experiments at two different outdoor shooting ranges. Experiment 1 was performed at a small shooting range with the shooter-target distance of 35 meters. The shooter took one position for the entire experiment. The signals were acquired by 4 UGS. The layout of Experiment 1 is presented in Figure 6.37. Experiments 2 and 3 were performed at a larger shooting range with the shooter-target distance of 100 meters (for the central shooting position). The shooter took three firing positions during both experiments. The signals were acquired by 6 UGS. In Experiment 2 the UGS were placed in a tight hexagon-shaped cluster, equidistantly positioned 5 meters away from the cluster's center. The layout of Experiment 2 is presented in Figure 6.38 (left image). In Experiment 3, on the other hand, the UGS were distributed more sparsely. The layout of Experiment 3 is presented in Figure 6.38 (right image). The firearm used in all three experiments was the Husqvarna 8x57JS rifle with the cartridge muzzle velocity equal to $v_0 = 780$ m/s. The shockwave is thus expected to spread at approximately $\theta \simeq 25.8^\circ$ relative to the bullet's trajectory.

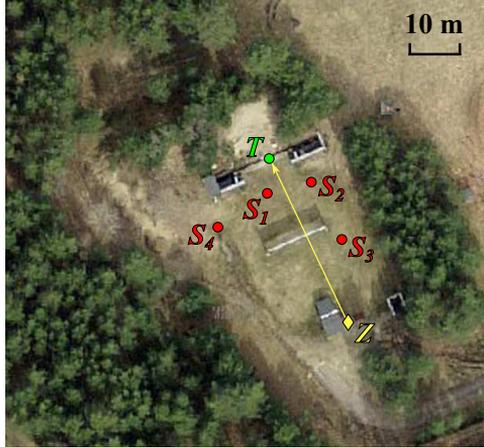


Figure 6.37: Layout of Experiment 1. T — target position; Z — shooter position; S_k — UGS positions.

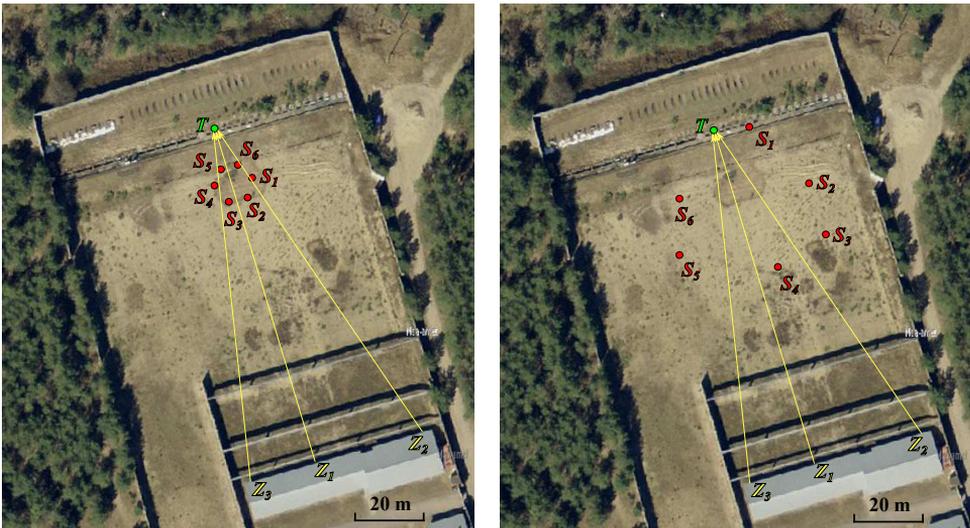


Figure 6.38: Layout of Experiments 2 (left) and 3 (right). T — target position; Z_i — shooter positions; S_k — UGS positions.

UGS latitude/longitude coordinates were measured using a standalone GPS device (Trimble R8 GNSS) since none of the UGS prototypes had GPS locators on board. For offline data analysis the GPS coordinates were converted into a local planar coordinate system with the target being set as the zeroth coordinate. The steering angle α_k is defined for each UGS as the heading measured with a high-precision compass. The presented experimental results are already brought to zero steering and the influence of α_k measurement error is not discussed.

The acoustic data was acquired with prototype 1 UGS during Experiment 1 and with prototype 2 UGS during Experiments 2 and 3. The shooter localization procedure was then evaluated offline in the MATLAB environment. The signals acquired by prototype 1 UGS were sampled at $f_s = 48$ kS/s per channel using the MATLAB Data Acquisition Toolbox. Prototype 1 UGS operate independently from one another. No inter-UGS communication is performed, and only rough synchronization is achieved by scheduling the starting moment of data acquisition in MATLAB. Prototype 2 UGS sample the signals at $f_s = 20$ kS/s per channel, and are softly synchronized by the WSN middleware component, discussed in Section 5.3. Mutual start time is achieved by broadcasting a sequence of specially timed messages from a control node (six messages counting down from 100 ms with 20 ms intervals), which are used to trigger the concurrent start of signal sampling within the sensor cluster.

Experiment Conditions

Experiment 1 was conducted at a shooting range surrounded by scattered trees. A bullet-catching sand mound is situated approximately 5 meters behind the target. The shooter’s position is situated beside a small concrete safety bunker, which obstructed direct line of sight of UGS S_4 . An overhead horizontal barrier is situated in the middle of the shooting range. The shooter fired 30 shots from a standing position. As the target and all UGS were raised by approximately 1 meter from the ground, each bullet passed the cluster at UGS level of elevation or slightly higher. Layout coordinates in meters are presented in Table 6.2. Weather conditions were the following: temperature $t^\circ \simeq 2$ °C, cloudiness 10%, no precipitation, wind speed 1–2 m/s. Parameters for all steps of the localization process are presented in Table 6.3.

Experiment 2 was conducted at a shooting range, which is entirely fenced by tall concrete walls. A bullet-catching sand mound is situated approximately 15–20 meters behind the target. The firing points are situated just outside the shooting range hall. Three overhead horizontal barriers are placed along the first 25 meters of the range (see Figure 6.39 top image). The shooter fired 6 shots from each of the three firing points from a standing position. As the target was elevated from the ground level by 3 meters, and all UGS were

Table 6.2: Target \mathbf{x}_T , firing point \mathbf{x}_{Z_i} and UGS \mathbf{x}_{S_k} coordinates in meters

Type	Experiment 1	Experiment 2	Experiment 3
\mathbf{x}_T	(0, 0)	(0, 0)	(0, 0)
\mathbf{x}_{Z_1}	(0, 35)	(0, 100)	(0, 100)
\mathbf{x}_{Z_2}	-	(-28.5, 100)	(-28.5, 100)
\mathbf{x}_{Z_3}	-	(20, 100)	(20, 100)
\mathbf{x}_{S_1}	(4, 6)	(-5, 16)	(-10, 3)
\mathbf{x}_{S_2}	(-5.5, 7)	(-2.5, 20.3)	(-20, 20)
\mathbf{x}_{S_3}	(-6, 20)	(2.5, 20.3)	(-20, 35)
\mathbf{x}_{S_4}	(14, 7.5)	(5, 16)	(-5, 40)
\mathbf{x}_{S_5}	-	(2.5, 11.7)	(20, 30)
\mathbf{x}_{S_6}	-	(-2.5, 11.7)	(15, 15)

raised by slightly more than 1 meter from the ground, the bullets traveled above the UGS cluster (see Figure 6.39 bottom image). Layout coordinates in meters are presented in Table 6.2. Weather conditions were the following: temperature $t^\circ \simeq 8^\circ\text{C}$, cloudiness 50%, no precipitation, wind speed 5–10 m/s. Parameters for all steps of the localization process are presented in Table 6.3.

Experiment 3 was conducted at the same shooting range as Experiment 2. The same firing points and target position were used. The shooter fired 6 shots from points 1 and 2, and 7 shots from point 3 from a standing position. The UGS were more widely distributed. UGS S_1 was placed at the target’s elevation level, as portrayed in Figure 6.40. Layout coordinates in meters are presented in Table 6.2. Weather conditions were the following: temperature $t^\circ \simeq 6^\circ\text{C}$, cloudiness 100%, light rain, wind speed 9–12 m/s with gusts up to 20 m/s. Parameters for all steps of the localization process are presented in Table 6.3.

Table 6.3: Shot detection, DOA estimation and shooter localization parameters

Parameter	Unit	Experiment 1	Experiment 2	Experiment 3
f_s	kS/s	48	20	20
t_W	ms	10	20	20
n_h	-	500	500	500
r_{FOV}	m	0.5	0.5	0.5
σ, n_{\min}	-	0.8, 3	0.8, 3	0.8, 3
$(\varphi_{\min}^{(SW)}, \varphi_{\max}^{(SW)})$	deg.	(21, 31)	(21, 31)	(21, 31)
$\varphi_{\max}^{(MB)}$	deg.	60	40	40
d_{step}	m	0.5	0.5	0.5



Figure 6.39: View of the shooting range from the shooter's position, 100 meters away from the target (top). UGS placement for Experiment 2 (bottom). The span of the bottom image is highlighted on the top image with a red rectangle.



Figure 6.40: UGS placement for Experiment 3. Shooting range front view is presented in the top image of Figure 6.39.

Results of Experiment 1

An example of gunshot event detection by UGS S_3 was presented in Figure 6.32. Results show that the applied detection procedure succeeds in detecting gunshot events even with a significantly short TDOA between SW and MB events. During the experiment all 30 shots were detected by all UGS, however, UGS S_4 failed to provide the DOA of seven MB events. Close analysis of signals acquired by UGS S_4 shows that the number of detected events was equal to the number of signal transients per shot. Since the direct line of sight from the shooter to UGS S_4 was obstructed by the safety bunker, the intermediate azimuth estimates did not have sufficient quality to pass the criteria of Algorithm 4.1 and no final estimates were thus made. Other UGS detected both SW and MB for every shot; TH was detected in the majority of cases. There were also 13 cases of detection of TH before MB by UGS S_1 and S_2 , the reason being their close disposition to the target. These results clearly indicate the need of gunshot event identification prior to shooter localization.

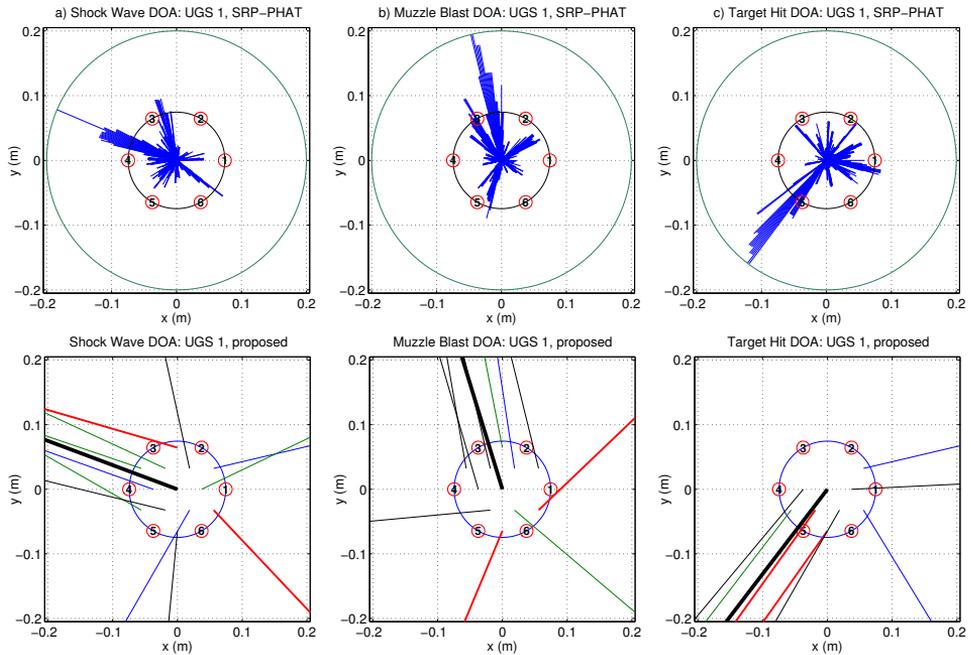


Figure 6.41: Intermediate results of DOA estimation of SW, MB and TH events made by UGS S_1 during Experiment 1.

The two considered DOA estimation methods succeed in establishing a single distinct direction in the majority of cases. A visualization of DOA estimation intermediate results for UGS S_1 is presented in Figure 6.41. In the figure top subplots represent DOA estimation results of SRP-PHAT with

blue lines denoting SRP values of points defined in Section 4.2.2, their length having been normalized to the radius of the green circle. Bottom subplots represent DOA estimation results of the proposed method. The individual pair-wise estimates for the microphone pairs defined by (4.15) are ordered by their cross-correlation peak distinctness from the least to the most sharp and depicted as black, blue, green and red lines, respectively. The thick black lines denote the final estimates. Figure 6.41 shows that that both methods produce one distinct beam and several lesser beams, corresponding to DOA of NOI events. The subplots corresponding to SW events both show a minor beam in the MB direction. This evidently happens due to short TDOA between the two events and their partial overlapping. The DOA of MB itself is very evident in the middle pair of subplots. Figure 6.41 clearly shows that the proposed method produces results highly similar to the ones of SRP-PHAT.

The DOA estimates of four consecutive shots computed by SRP-PHAT are presented in Figure 6.42a, and the estimates produced by the proposed method — in Figure 6.42b. In the figures several estimate values are equal and thus fully overlap with each other. Red diamonds denote shooter true positions, green circle denotes the target position, and blue circles denote UGS positions. DOA estimates of UGS S_1 – S_4 are denoted by blue, green, purple and red lines, respectively. It can be seen in Figure 6.42 that SRP-PHAT estimates are more dispersed for UGS S_2 and S_3 . SW, MB and TH events are well distinguishable for both methods, however, results for UGS S_4 are significantly worse due to its larger miss distance and the obstructed line of sight to the shooter.

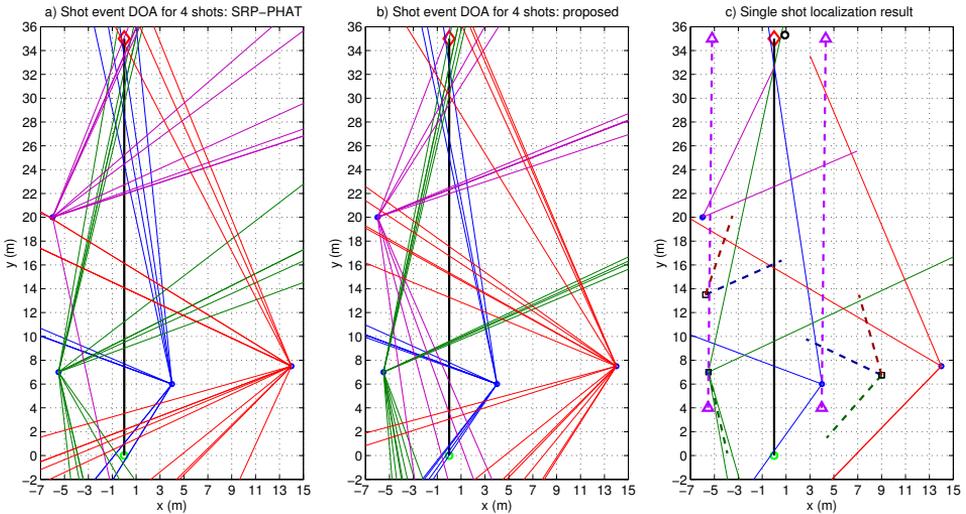


Figure 6.42: DOA estimates for four consecutive shots of Experiment 1 produced by a) SRP-PHAT and b) the proposed method. c) Localization result for a single shot.

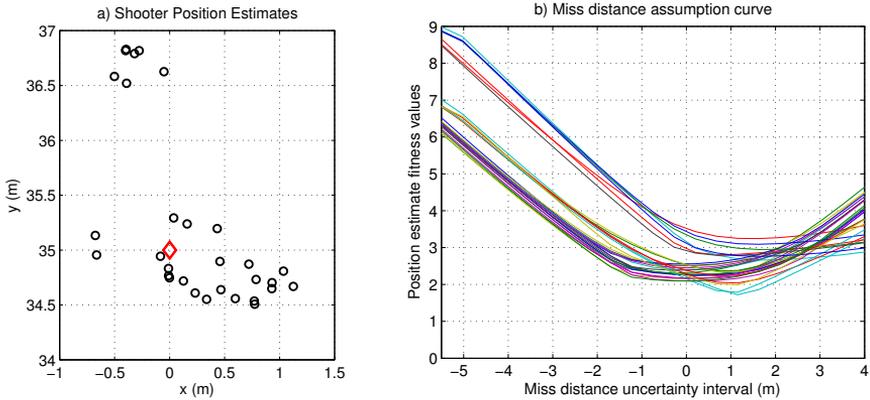


Figure 6.43: Experiment 1 results for 30 shots. a) Estimated shooter positions. b) Values of f_{fit} (6.21) for the miss distance uncertainty interval.

The intermediate results of localization and the final shooter location estimate for a single shot are presented in Figure 6.42c. In the figure red, blue and green dotted lines denote the mean values of SW DOA $\bar{\phi}^{(SW)}$, MB DOA $\bar{\phi}^{(MB)}$ and NOI event DOA of clusters Φ_p . The purple dotted arrows denote the shot angle ϕ_Z and the miss distance uncertainty interval; the black circle denotes the final estimated shooter position. UGS $\{S_2, S_3\}$ and $\{S_1, S_4\}$, as expected, form clusters of consistent DOA estimates and group into G_L and G_R , respectively. Mean estimates of clustered DOA values are presented in Figure 6.42c as dotted lines starting from the spatial centroids of these clusters. The shot angle $\phi_Z \simeq 90^\circ$ is estimated with high accuracy; the closest left and right UGS $\check{S}_L = S_2$, $\check{S}_R = S_1$ are correctly assigned, and thus the miss distance uncertainty interval is properly determined.

Final shooter position estimates obtained using the proposed DOA estimation method are presented in Figure 6.43a. In the figure the red diamond denotes the shooter's true position. To quantify localization accuracy the mean error (ME) metric is used, which is calculated as the mean distance between the estimated and the known shooter positions as

$$\text{ME} = \frac{1}{N_s} \sum_{i=1}^{N_s} \left((x_{\hat{Z}}(i) - x_Z(i))^2 + (y_{\hat{Z}}(i) - y_Z(i))^2 \right)^{1/2}, \quad (6.22)$$

where N_s is the total number of shots. ME for 30 shots along with its Standard Deviation (SD) is presented in Table 6.4. The table shows that using the proposed DOA method results in a slightly lower ME. Generally the localization quality for both DOA estimation methods is notably high for Experiment 1. In Figure 6.43a a congestion of remote points in the top left corner results from several MB not being detected by UGS S_4 . Instantaneous bullet velocity estimation (see Section 6.4.4) resulted in $\hat{v} \simeq 740$ m/s, which

is consistent with the cartridge specification parameters (velocity of 753 m/s for ranges under 50 m). The values of the fitness function f_{fit} are presented in Figure 6.43b. The function’s minimum is situated at ± 1 m from the actual miss distance, and one global minimum of f_{fit} exists for every shot. Thus, miss distance estimation in this case can be performed by a gradient descent method rather than by iterative search.

Table 6.4: Shooter position estimate mean error (ME) and standard deviation (SD) in meters

Method	Parameter	Experiment 1	Experiment 2	Experiment 3
SRP-PHAT	\hat{Z} ME	1.12	6.65	8.92
	\hat{Z} SD	0.73	3.53	6.80
Proposed	\hat{Z} ME	0.87	7.08	7.32
	\hat{Z} SD	0.56	3.86	6.15

Results of Experiment 2

The gunshot acoustic event detection procedure succeeded in detecting every shot instance on every UGS with 5–6 acoustic events per shot on average, occasionally reaching 8–9 events. The distribution of acoustic events of Experiment 2 is very similar to the one of Experiment 3, a single shot signal example of which was presented in Figure 6.34. The large number of NOI events is caused by numerous reflections of SW, MB, as well as TH off the concrete walls surrounding the shooting range (see Figure 6.38). An elevated bullet trajectory causes ground reflections of SW, as it was explained in Section 6.4.3. Consequently, the SW signal pattern resembles a transient combined with several weaker disturbances. This results in MB being detected as the 3rd or 4th event peak for every shot instance.

The DOA estimates for all 18 shots calculated by SRP-PHAT are presented in Figure 6.44. The results of the proposed method are presented in Figure 6.45. In the figures red diamonds denote shooter true positions, green circle denotes the target position, and blue circles denote UGS positions. Both methods succeed in estimating the SW and MB DOA with a sufficient accuracy for the majority of cases. Generally it can be noticed that the MB DOA produced by the proposed method are less dispersed than the ones produced by SRP-PHAT. The DOA of several SW and MB instances were not correctly determined either due to low quality of intermediate estimates, or severe overlapping with NOI events. Both methods completely fail to estimate the MB DOA for UGS S_6 . Close analysis of UGS S_6 signal and DOA intermediate results has shown that, in fact, both methods do not produce a beam in the shooter’s direction even for slightly overlapping or non-overlapping MB events. The cause presumably lies in some peculiarity

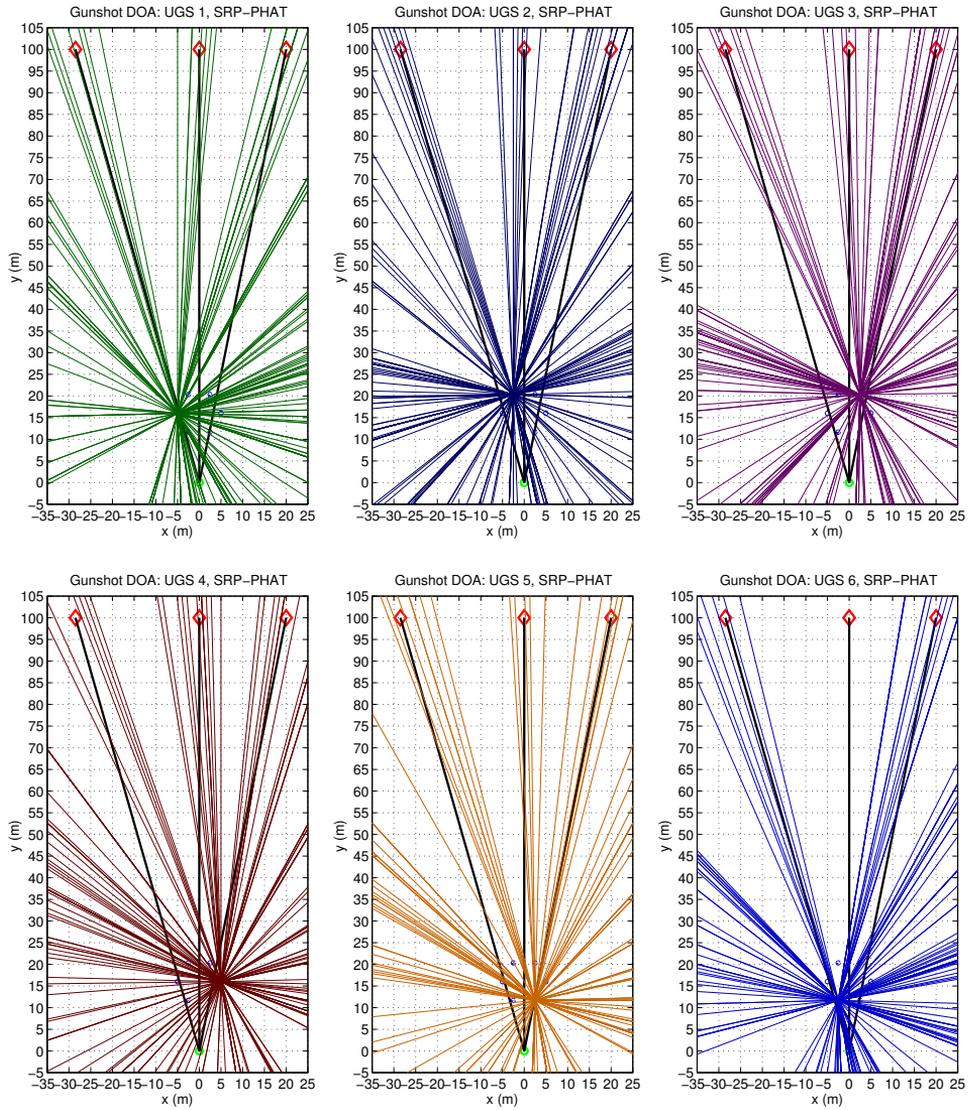


Figure 6.44: Experiment 2 DOA estimates for 18 consecutive shots using SRP-PHAT.

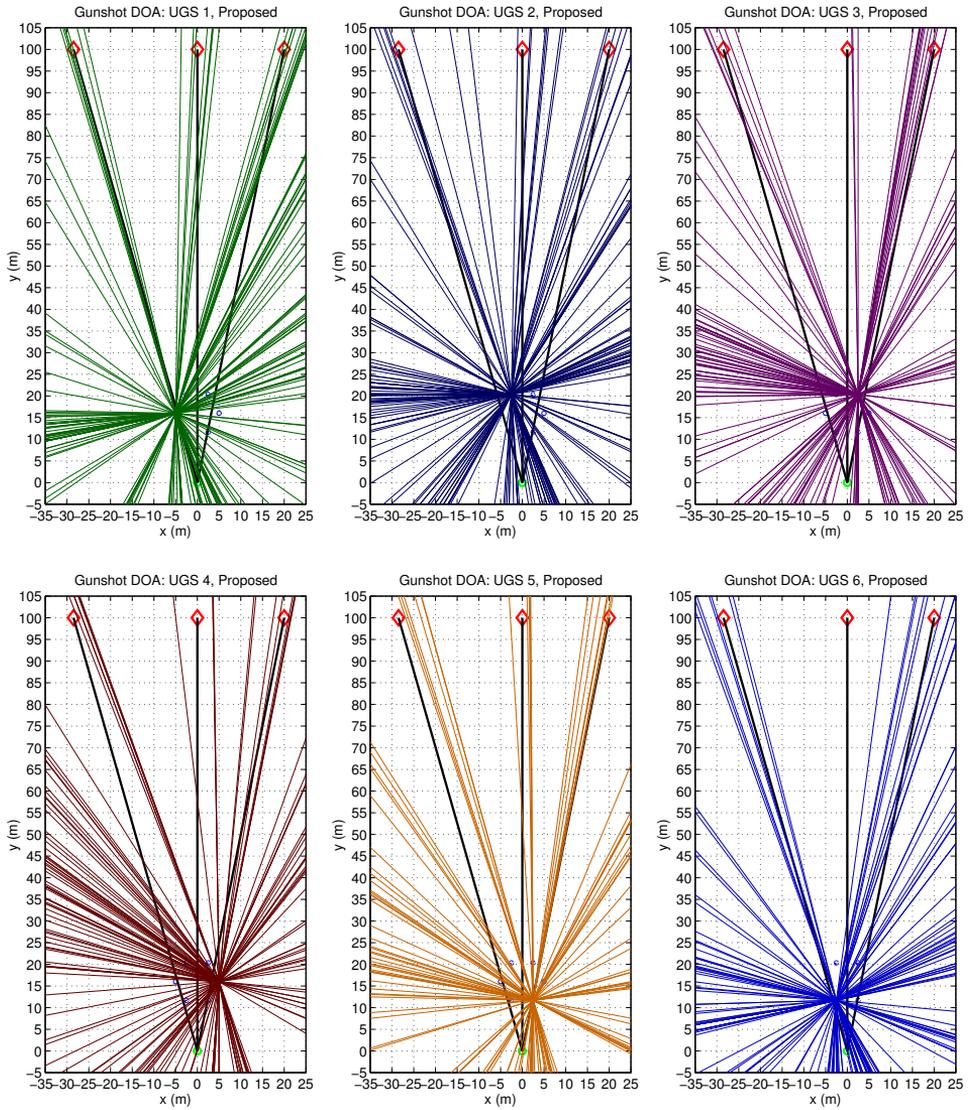


Figure 6.45: Experiment 2 DOA estimates for 18 shots using the proposed method.

of UGS S_6 position. As seen in Figure 6.39, UGS S_6 is situated next to a corner-like barrier gap between target points 9 and 10. It is possible that this gap produces a NOI event, which totally overlaps with the MB signal and corrupts its DOA estimation. This effect once again indicates the need for acoustic event identification prior to data fusion. The overall DOA patterns of Figure 6.44 and Figure 6.45 clearly show the nature of the NOI events: the DOA skewed left and right from the shooter's position are caused by reflections of SW (produced by the bullet at the beginning of its flight path) and MB off the side walls; the DOA pointing in the target's direction are caused by TH noise, the bullet penetrating the sand mound and reflections off the back wall.

The intermediate results of localization and the final shooter location estimate for a single shot case from each of the three firing points are presented in Figure 6.46. In the figure red and blue dotted lines denote $\bar{\phi}^{(SW)}$ and $\bar{\phi}^{(MB)}$ of clusters Φ_p , purple dotted arrows denote the shot angle ϕ_Z and the miss distance uncertainty interval, and black circles denote final estimated shooter positions. The DOA of NOI events are removed from the plots for presentation clarity. For firing point Z_1 all UGS form a single cluster of MB DOA concurrent estimates. UGS $\{S_1, S_2, S_6\}$ and $\{S_3, S_4, S_5\}$ form clusters of DOA corresponding to SW, detected to the left and right of the bullet's trajectory, and group into G_L and G_R , respectively. For firing points Z_2 and Z_3 MB DOA clusters are also formed from all UGS because the cluster dimensions are significantly smaller compared to the distance between the cluster and the shooter positions, which results in MB DOA being roughly equal. The clusters of concurrent SW DOA estimates are formed for Z_2 from UGS $\{S_1\}$ in the left group and $\{S_2, S_3, S_4, S_5, S_6\}$ in the right group. For point Z_3 the left group consists of UGS $\{S_1, S_2, S_3, S_6\}$ and the right group — of UGS $\{S_4, S_5\}$. As UGS S_1, S_6 and UGS S_3, S_5 are situated nearly along the bullet's trajectory for points Z_2 and Z_3 , respectively, their belonging to either the left or right group changes from shot to shot. This does not influence the overall localization accuracy, as the considered UGS cluster is quite dense, and thus the miss distance ambiguity interval does not change significantly. The shot angles $\phi_{Z_1} \simeq 90^\circ$, $\phi_{Z_2} \simeq 106^\circ$ and $\phi_{Z_3} \simeq 79^\circ$ are estimated with high accuracy.

Final shooter position estimates for all three firing points are presented in Figure 6.47. In the figure black circles denote the estimated shooter positions and red diamonds denote true shooter positions. It can be seen that the estimates are significantly more scattered, when compared to the estimates of Experiment 1. Table 6.4 shows that the ME for Experiment 2 is approximately 7 m, which is notably higher than a ME of approximately 1 m of Experiment 1. However, taking into consideration that the range set for Experiment 2 is almost three times larger, and prototype 2 UGS use an inferior embedded ADC at $f_s = 20$ kS/s, compared to a standalone DAQ

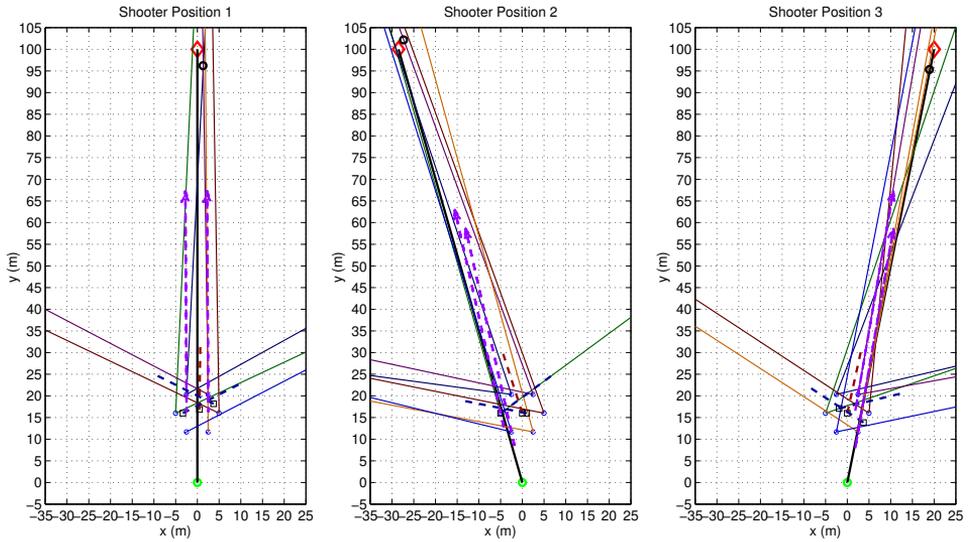


Figure 6.46: Experiment 2 localization results for one shot per shooter position.

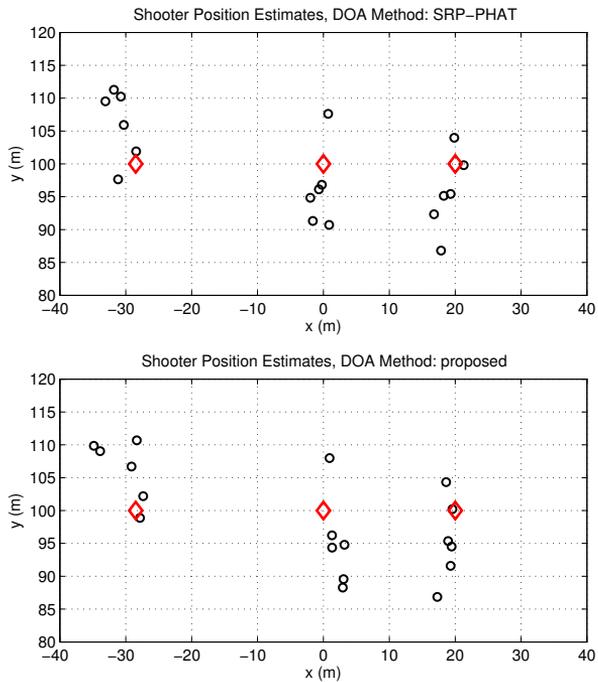


Figure 6.47: Experiment 2 localization results for 18 shots with SRP-PHAT (top) and the proposed method (bottom) used for DOA estimation.

of prototype 1 with a larger bit depth and sampling at $f_s = 48$ kS/s, the decrease in localization quality is quite expected and justified. Generally, applying both SRP-PHAT and the proposed method of DOA estimation in the localization procedure yields similar localization quality with SRP-PHAT resulting in slightly more accurate estimates.

Bullet velocity estimation resulted in $\hat{v} \simeq 720$ m/s, which is consistent with the cartridge specification parameters (velocity of 727 m/s for a range of 100 m). Miss distance estimation via the fitness function f_{fit} is less trustworthy for Experiment 2 due to UGS being very closely positioned to each other, which results in very narrow miss distance ambiguity intervals, especially for firing points Z_2 and Z_3 . As a result, if ϕ_Z estimation produces even a slightly inaccurate result, the bullet's trajectory will not fall into the ambiguity interval, which obstructs miss distance estimation. In the considered case ϕ_Z estimation performed accurately enough for the bullet's trajectory to be at an edge of the ambiguity interval or very close to it, e.g., for firing point Z_2 result in Figure 6.46. This means that in the minimal value of f_{fit} appears close to the edge of the ambiguity interval. A more spatially distributed UGS cluster would solve this problem.

Results of Experiment 3

The number of detected gunshot acoustic events is similar to the one of Experiment 2, i.e., 5–6 events per shot on average. The situation with reflections off the surrounding walls is worse for UGS S_1 , S_2 and S_3 , as they are situated closer to the left and back walls of the shooting range. On the other hand, the effect of SW overlapping with its ground reflection is less evident for the UGS with larger miss distances. Nevertheless, MB is detected as the 3rd peak for 18 out of 19 shot instances.

The DOA estimates for all 19 shots calculated by SRP-PHAT are presented in Figure 6.48. The results of the proposed method are presented in Figure 6.49. The DOA corresponding to reflections off the side and back walls is clearly evident in these figures. Both methods perform well with the results of the proposed method being less disperse, as in the two previously discussed experiments. The proposed method performs worse for UGS S_2 , firing point Z_3 , however, it succeeds in calculating the DOA for UGS S_6 , firing point Z_3 , where SRP-PHAT fails. The problem with estimating MB DOA for point Z_3 by UGS S_6 is presumably situated with the deterioration of MB wavefronts caused by reflections off the wall right next to point Z_3 (see Figure 6.38). Generally, the UGS positioning scheme of Experiment 3 results in the DOA corresponding to NOI events being more consistent, making SW and MB DOA more distinguishable.

The intermediate results of localization and the final shooter location estimate for a single shot from each of the three firing points are presented in

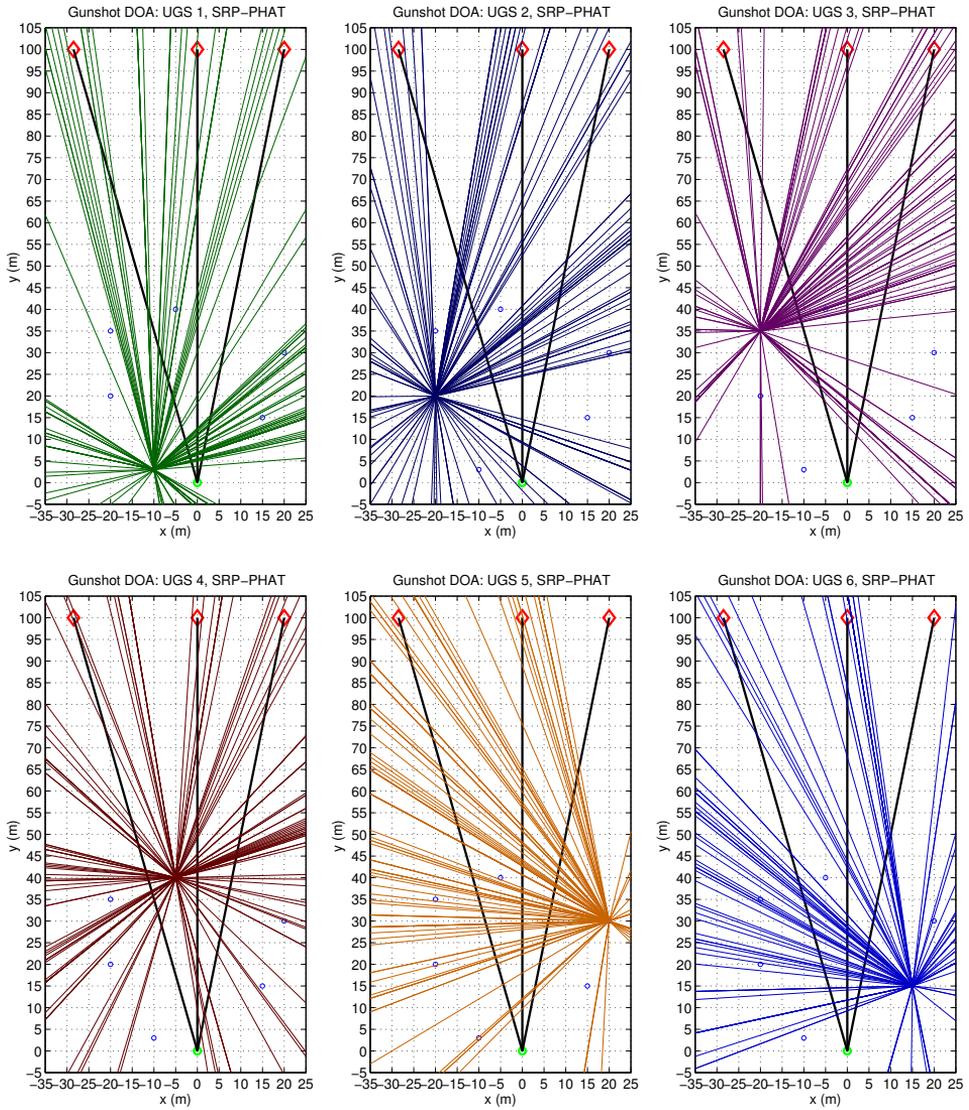


Figure 6.48: Experiment 3 DOA estimates for 19 consecutive shots using SRP-PHAT.

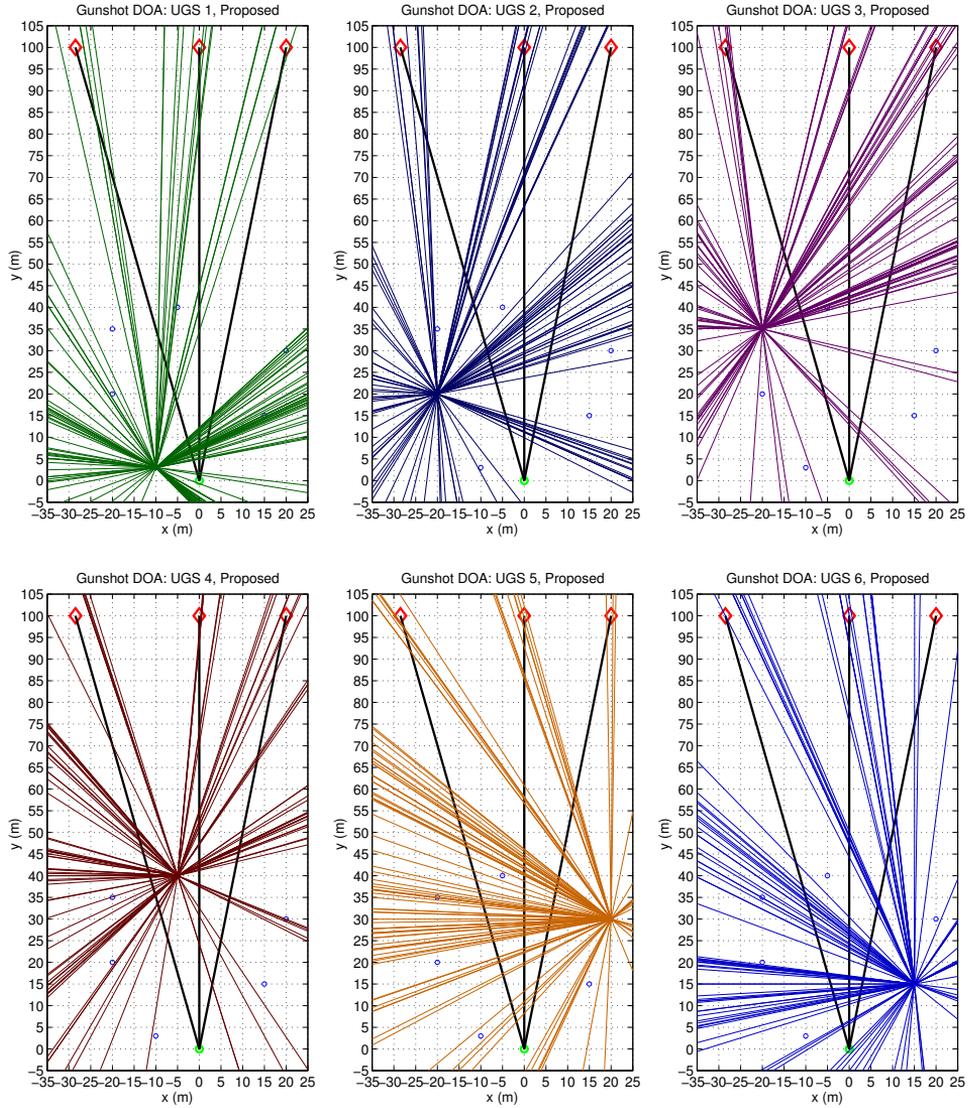


Figure 6.49: Experiment 3 DOA estimates for 19 shots using the proposed method.

Figure 6.50. DOA corresponding to NOI events are removed from the plots for presentation clarity. For point Z_1 UGS $\{S_1, S_2, S_3, S_4\}$ and $\{S_5, S_6\}$ form MB and SW DOA concurrent estimate clusters, corresponding to the left and right groups G_L and G_R , respectively. For point Z_2 the UGS belonging to G_L are $\{S_1, S_2, S_3\}$, and the UGS belonging to G_R are $\{S_4, S_5, S_6\}$. For point Z_3 the UGS are partitioned as $\{S_1, S_2, S_3, S_4\}$ into G_L and $\{S_5, S_6\}$ are partitioned into G_R . As the dimensions of the UGS cluster are large enough to be comparable with the distance from the cluster to the shooter, MB DOA do not form a single consistent direction, as was the case in Experiment 2. Rather, coherent estimates are formed by UGS situated to the left and right of the bullet's trajectory and are skewed towards the shooter's position. Ultimately this experiment can be perceived as a scaled-up version of Experiment 1. The shot angles $\phi_{Z_1} \simeq 90^\circ$, $\phi_{Z_2} \simeq 106^\circ$ and $\phi_{Z_3} \simeq 79^\circ$ are estimated with high accuracy.

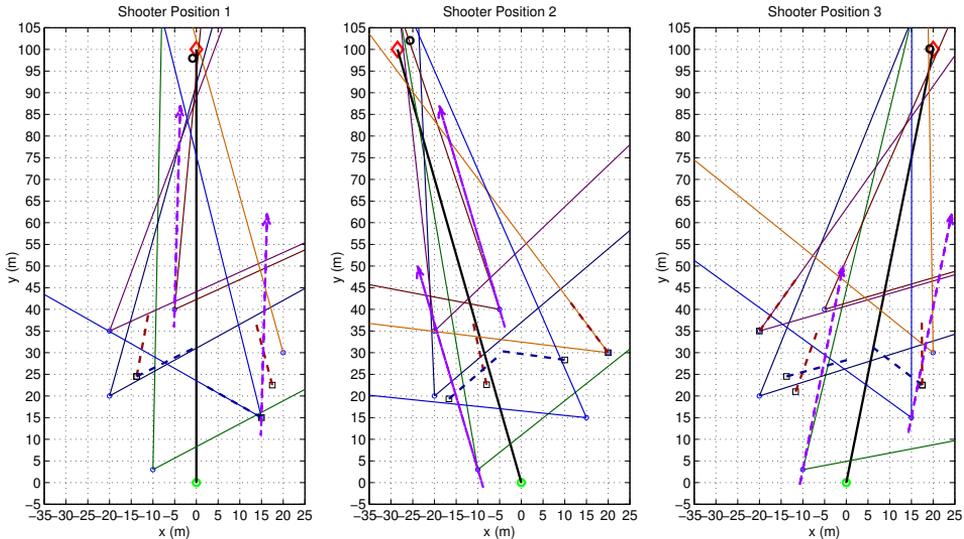


Figure 6.50: Experiment 3 localization results for one shot per shooter position.

Final shooter position estimates for all three firing points are presented in Figure 6.51. The estimates are also significantly more scattered, compared to the estimates of Experiment 1. Table 6.4 presents the ME of localization calculated using (6.22). The ME for both Experiments 2 and 3 using the proposed method for DOA estimation is approximately 7 m. The ME of Experiment 3 with SRP-PHAT used as a DOA method is larger, which indicates the supremacy of the proposed method over SRP-PHAT in this case. It can be also noticed from Figure 6.51 that Z_2 has only 5 estimates around its true position. This is due to one shot being localized incorrectly and the point residing outside the figure bounds for both DOA methods. This is a single example of gunshot event identification failure by DOA. If a NOI event

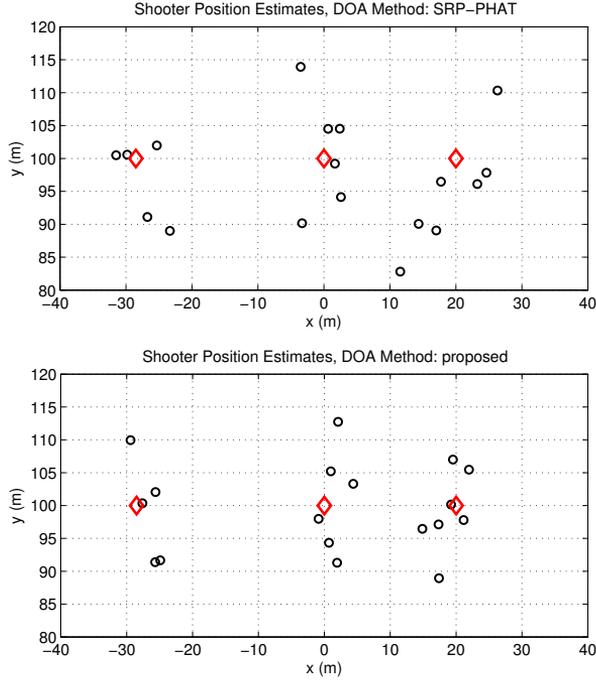


Figure 6.51: Experiment 3 localization results for 19 shots with SRP-PHAT (top) and the proposed method (bottom) used for DOA estimation.

has a DOA resembling that of MB and satisfies all the temporal and spatial bounds of the MB check, it can be falsely labeled as MB. Consequently, the TDOA Δt is computed incorrectly and the whole localization procedure can fail. However, this requires the NOI event to corrupt the DOA estimates of several UGS, which is highly unlikely. In our case UGS S_2 and S_3 mistook a NOI event for MB, and their incorrect estimates of distance to shooter steered the cluster’s global estimate farther from shooter’s true position.

Bullet velocity estimation resulted in $\hat{v} \simeq 725$ m/s, which closely corresponds to the result of Experiment 2. Miss distance estimation via the fitness function f_{fit} operates well for this experiment, as the miss distances for all UGS are sufficient and f_{fit} forms curves, similar to the ones portrayed in Figure 6.43 with a single global minimum for the majority of shot instances.

6.4.6 Discussion

The main differences between the state of the art and the proposed method of shooter acoustic localization are presented in Table 6.5. Although the proposed method of gunshot acoustic component identification using DOA information increases shooter localization robustness, accounting for the destructive influence of various types of NOI events, it has several weak points that yet require attention.

The instantaneous bullet velocity estimation via the shot angle needs to be developed into a more general procedure that also accounts for the decrease in bullet velocity with traveled distance. The bullet velocity was approximately estimated to be 720–725 m/s, which is significantly less than the 780 m/s muzzle velocity claimed in the cartridge specification. Such velocity reduction even for a 100 meter range case can influence localization results. Thus the degree of this influence needs to be quantified and accounted for in the future.

The event identification and shooter localization approach needs to be tested in a burst-mode shooting scenario, the peculiarities of which were reviewed in Section 6.4.3. In such a scenario shot instance separation will likely pose a serious problem, so the acoustic event detection procedure will have to be developed further to account for extremely closely spaced shot instances. Also the procedure of sending the shot information to the fusion node is to be reviewed for this case, as sending a large number of packets through the WSN in a very short period of time tends to be problematic.

The boundaries of application of the gunshot planar geometry model, where either the shooter’s or target’s elevation above the UGS cluster starts to influence localization accuracy have to be identified. If the bullet’s trajectory does not lie in the same plane as the UGS cluster, the shot geometry cannot be estimated by a planar model, since the conical wavefront of SW cannot be modeled as a planar wavefront. Consequently, the distance to shooter cannot be estimated by the horizontal projection of the bullet’s trajectory. As the results of Experiment 2 have shown, slight elevation of the target does not influence the localization procedure, however, larger elevation levels were not considered.

Table 6.5: Main differences between state of the art methods and the proposed approach to shooter acoustic localization

State-of-the-art	Proposed
MB strictly follows SW; NOI events not accounted for	Accounts for NOI events, SW and/or MB may be corrupted or masked
Some require high precision of synchronization	Soft synchronization only to distinguish between shot instances
Assume known bullet caliber $\rightarrow v_0 \rightarrow \approx v \rightarrow \theta$	θ estimated from measured $\phi_k^{(SW)}$ and calculated shot angle ϕ_Z
Estimate miss distance $\hat{d}_{miss}^{(k)}$ using a shockwave acoustic model: unreliable in presence of NOI events	Estimates $\hat{d}_{miss}^{(k)}$ by establishing the ambiguity interval $[d_{min}^{(k)}, d_{max}^{(k)}]$ and finding a best fit solution
Generally more than one assumption	Only assumption: SW event is detected first, if not masked

6.5 Conclusions

The applications of the signal processing methods discussed in this work are not by all means limited to the ones presented in this chapter. There exists a great variety of applications of CPS in many areas of life, both civil and military, where human-machine interaction and computer system integration into the environment occur. The presented applications merely serve as an example of the vast possibilities of distributed multimodal sensor systems.

The considered applications for ISR systems are discussed separately, however, the individual components of acoustic, PIR, vibration, etc. smart sensors are combined during real life operation of the entire CPS, and their individual situation assessments are fused at higher levels of information fusion in order to obtain global situation estimates. The vehicle identification system, for example, is to operate in collaboration with the foot soldier detection system in order to obtain estimates on both motorized and infantry party advancements. The roles of UGS equipped with different types of sensors can be reassigned during system online operation, or they can perform multiple tasks at once if the computational capabilities of their hardware components allow for that. For the civil application of industrial machinery monitoring this means that machine operation state identification, fault and malfunction detection, and localization of noise emitting regions are all performed simultaneously, and the roles of smart sensors can be reassigned between the tasks depending on any specific situation.

Regarding the experimental results of each individual application, the identification, localization, etc. accuracy of the proposed methods is assessed for a relatively small number of registered events. During real life system operation, which can be uninterrupted for months and even years, unusual and unexpected events, which will influence estimation accuracy, will undoubtedly arise. The high uncertainty of the real world environments does not allow to consider all the possible process states and outcomes, and thus the system's true potential can be assessed only during long-term operation.

Conclusions

We can only see a short distance ahead, but we can see plenty there that needs to be done.

Alan Turing

This final chapter formulates concluding remarks according to the problem statement and the solutions presented in the thesis. At the end of the chapter the topics of future research are discussed.

The solutions of single sensor signal processing discussed in Chapter 3 are both verified on real life test signals in the same chapter, and later used for specific applications discussed in Chapter 6. The presented pattern recognition procedure comprising feature extraction and classification stages is regarded as the most common approach to identification of process states with well established dynamics. The main contribution of this chapter is situated with the proposed multistage procedure of non-stationary process state estimation, applied to moving vehicle detection and identification. The procedure reflects the need of process detection prior to classification due to sophisticated process dynamics and the harmful influence of background noise. The signal analysis stages adopted in the procedure are not strictly fixed and additional stages can be added or removed for application-specific purposes. The procedures for stationary and non-stationary process state estimation are applied to industrial machinery state monitoring and identification of military vehicles for Intelligence Surveillance and Reconnaissance (ISR) purposes, respectively. The experimental results account for background noise arising in the real unconfined environments, however, to ensure increased reliability of situation assessments made by the entire Cyber Physical System (CPS) the results of signal processing of single-sensor components have to be reassured by other components through data fusion. Using sensors employing different signal modalities is beneficial in this regard. As discussed in Chapter 3, for example, the information of acoustic sensors with wide Fields of View (FOV), though prone to background noise influence, can be reassured by highly localized readings of acceleration sensors, not

susceptible to noise originating from elsewhere besides the specific region of vibration. Noise tolerance and pattern recognition accuracy can also be increased by analyzing different signal properties independently, as it is done in the proposed multistage procedure. The application of multi-sensor solutions further expands the possibilities of distributed systems.

The methods of multi-channel signal analysis presented in Chapter 4 are applied to acoustic signals, however, they can also be theoretically applied to band-limited signals of other modalities propagating in media other than air (where the same geometry of wave propagation applies). The contributions of Chapter 4 are most prevalent in the work. The reduced functional of Direction of Arrival (DOA) search for the Steered Response Power with Phase transform (SRP-PHAT) is a minor contribution, however, analogous approaches are seldom encountered in state of the art DOA estimation approaches. The main contributions constitute the planar (1D) DOA estimation approach for azimuth estimation in linear microphone arrays and volumetric (2D) DOA estimation approach for azimuth and elevation Angles of Arrival (AOA) in conical arrays. The azimuth AOA estimation is also used for planar DOA estimation in Uniform Circular Arrays (UCA), which constitute the base of the conical array. Another major contribution is the proposed distributed acoustic localization method. The ideas of approximate localization of the region confining the acoustic source and operation on asynchronous wireless multi-channel nodes without raw signal interchange are almost unique to state of the art distributed acoustic localization Wireless Sensor Network (WSN) systems. Most modern approaches assume strict WSN node synchronization or apply single sensor nodes with raw signal aggregation on the fusion node for Time Difference of Arrival (TDOA) estimation. In the proposed approach both TDOA and DOA are estimated on each of the individual WSN nodes and only short messages containing node position coordinates, DOA estimates and the time labels are transmitted. The DOA estimation and distributed localization approaches are verified in both controlled environment and field experiments. Initial verification results are presented in Chapter 4 for indoor experiments, and the results of application-specific field experiments are presented in Chapter 6. The proposed localization method is applied to localization of noise emitting regions of industrial machinery and to military vehicle localization for ISR purposes. The proposed DOA estimation method is additionally applied to shooter acoustic localization. Throughout the investigation the proposed DOA estimation method is compared with SRP-PHAT, which is considered to be one of the most (if not the most) robust and reverberation tolerant methods of DOA estimation.

The operation of the distributed CPS consisting of a WSN of smart sensors is briefly discussed in Chapter 5. As spatial and temporal data validation and WSN management falls out of the scope of this work, a solution of align-

ing and verifying local situation assessments for data fusion developed by the Research Laboratory for Proactive Technologies (Tallinn University of Technology) is presented as an example. The solution is also partly employed in the developed smart sensor prototypes presented throughout the thesis. Data fusion concepts are briefly discussed in Chapter 5, and application specific examples of data fusion are presented in Chapters 4–6. The contributions to data fusion algorithms comprise:

- local data fusion on a single smart sensor (as part of the decision making logic of the multistage procedure presented in Chapter 3),
- data fusion of DOA estimates for localization (as part of the distributed localization procedure presented in Chapter 4),
- fusion of localization estimates and object trajectory estimates presented in Chapter 5,
- application-specific data fusion of DOA estimates for shooter localization presented in Chapter 6.

Data fusion for localization and trajectory estimation is part of the proposed localization procedure and is thus novel. DOA estimate fusion for shooter localization is part of the proposed shooter localization approach, which has not been discussed in previous research on the topic.

Several applications of the signal processing methods presented in the thesis are considered in Chapter 6. The ideas of industrial machinery monitoring by distributed wireless sensors and noise emission source localization for diagnostics are not novel, however, the combination of multi-channel acoustic and other types of sensors for continuous operation state monitoring has not yet been widely discussed in previous research. The ISR application of vehicle detection and tracking has been previously discussed, and various approaches employing acoustic and other types of sensors (e.g., radar, lidar, hyperspectral) have been proposed. This application serves as a demonstration of both proposed vehicle identification and trajectory estimation approaches. The possibilities of using Passive Infrared (PIR) sensors for object speed estimation, on the other hand, have not been sufficiently discussed in previous research. Thus the proposed method of PIR signal analysis for speed estimation can be considered a novel contribution to ISR systems and other monitoring applications. The proposed shooter localization procedure is also a novel approach for the considered shot geometry. It is advantageous to the state of the art approaches based on acoustic information in that the proposed approach accounts for noise produced by events not-of-interest and generally makes less assumptions on specific properties of gunshots. The application scope of the presented signal processing methods is not by all

means limited to the specific applications considered in Chapter 6 and will be further explored in the future.

Future Research

Several topics situated with the signal processing methods presented in this thesis are considered for future research. First of all, the proposed methods of DOA estimation need to be complemented with reliable methods of acoustic source number estimation. Though the proposed DOA estimation methods are capable of finding several DOA corresponding to multiple sources residing in the FOV, there was little research conducted on the topic of source number estimation algorithms during the work. Reliable estimation of the number of sources in the FOV will enable more reliable acoustic localization as well. The incidental localized regions, eliminated during trajectory estimation in this work, will be identified more precisely with this additional knowledge. Using aggregate estimates of the number of sources made by all smart sensor nodes in the same FOV will yield more precise estimates for consecutive signal frames of individual smart sensors.

The localization procedure presented in the thesis is applied only for planar localization using only azimuth AOA of the estimated DOA. Future research plans include the expansion of the localization procedure for operation in the volumetric space. Although planar localization is sufficient for ground applications where slight elevation levels are concerned, for localization of acoustic sources situated well above the plane in which the sensors reside (e.g., localization of aerial vehicles or targets on building upper floors) requires the use of both AOA components and localization to a volumetric confined space.

Further testing of the presented methods in real life conditions for a prolonged periods of time is required in order to verify their tolerance to a large variety of possible noise sources and special case situations. During this testing period the methods can be refined to ensure increased reliability and robustness of the situation assessment process.

Generally, CPS are a fairly new concept, and thus their full potential is still to be explored. The topic of situation assessment in CPS will undoubtedly be rising in popularity over the coming years, and bringing with itself a large variety of new research fields. Great possibilities are arising before the science of signal processing, bringing the image of a fully connected world with sophisticated interaction between people and autonomous systems closer with every passing year. In this regard, the topics of future research greatly exceed the ones considered for the proposed methods alone, with almost endless possibilities lying ahead.

Bibliography

- [1] T. Adali and S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*, ser. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2010.
- [2] C. Aggarwal and C. Reddy, *Data Clustering: Algorithms and Applications*, ser. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Taylor & Francis, 2013.
- [3] J. Aguilar, “Gunshot location systems the transfer of the sniper detection technology from military to civilian applications,” in *Proc. 47th Int. Carnahan Conf. on Security Technology (ICCST)*, Medellin, Colombia, Oct 2013, pp. 1–6.
- [4] J. R. Aguilar, R. A. Salinas, and M. A. Abidi, “Acoustical model of small calibre ballistic shock waves in air for automatic sniper localization applications,” *Proc. SPIE 6538, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VI*, vol. 6538, pp. 65 381D–65 381D–8, 2007.
- [5] T. Ahmed, M. Uppal, and A. Muhammad, “Improving efficiency and reliability of gunshot detection systems,” in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 513–517.
- [6] R. Alur, *Principles of Cyber-Physical Systems*. MIT Press, 2015.
- [7] D. Arney, S. Fischmeister, J. M. Goldman, I. Lee, and R. Trausmuth, “Plug-and-play for medical devices: Experiences from a case study,” *Biomedical Instrumentation & Technology*, vol. 43, no. 4, pp. 313–317, 2009.
- [8] J. Ash, G. Whipps, and R. Kozick, “Performance of shockwave-based shooter localization under model misspecification,” in *Proc. 2010 IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, Texas, USA, Mar 2010, pp. 2694–2697.

- [9] S. Astapov, J. Berdnikova, J. Ehala, and J. S. Preden, "Shooter localization by networked multichannel acoustic ground sensors," in *Proc. IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, Oct 2015, pp. 332–337.
- [10] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [11] S. Astapov, J. Berdnikova, and J.-S. Preden, "Predictive acoustic localization and speaker tracking for distributed sensor networks," in *Proc. 13th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV)*, Singapore, Dec 2014, pp. 833–838.
- [12] S. Astapov, J. Ehala, and P. J.-S., "Performing acoustic localization in a network of embedded smart sensors," *International Journal of Microelectronics and Computer Science (IJMCS)*, vol. 6, no. 3, pp. 86–95, 2015.
- [13] S. Astapov, J. Ehala, and J.-S. Preden, "Collective acoustic localization in a network of dual channel low power devices," in *Proc. 21st Int. Conf. Mixed Design of Integrated Circuits Systems (MIXDES)*, Lublin, Poland, Jun 2014, pp. 430–435.
- [14] S. Astapov, J. S. Preden, T. Aruvali, and B. Gordon, "Production machinery utilization monitoring based on acoustic and vibration signal analysis," in *Proc. 8th Int. DAAAM Baltic Industrial Engineering Conf.*, Tallinn, Estonia, Apr 2012, pp. 268–273.
- [15] S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, "Object detection for military surveillance using distributed multimodal smart sensors," in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP 2014)*, Hong Kong, China, Aug 2014, pp. 366–371.
- [16] S. Astapov, J. S. Preden, and E. Suurjaak, "A method of real-time mobile vehicle identification by means of acoustic noise analysis implemented on an embedded device," in *Proc. 13th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2012, pp. 283–286.
- [17] S. Astapov and A. Riid, "A hierarchical algorithm for moving vehicle identification based on acoustic noise analysis," in *Proc. 19th Int. Conf. Mixed Design of Integrated Circuits and Systems*, Warsaw, Poland, May 2012, pp. 467–472.

- [18] S. Astapov and A. Riid, “A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device,” *Int. Journal of Electronics and Telecommunications*, vol. 59, no. 2, pp. 151–160, 2013.
- [19] S. Astapov, A. Riid, J. S. Preden, and T. Aruvali, “Industrial process monitoring by multi-channel acoustic signal analysis,” in *Proc. 14th Biennial Baltic Electronic Conference (BEC)*, Tallinn, Estonia, Oct 2014, pp. 209–212.
- [20] S. Astapov, J. Berdnikova, and J.-S. Preden, “A method of initial search region reduction for acoustic localization in distributed systems,” in *Proc. 20th Int. Conf. Mixed Design of Integrated Circuits and Systems (MIXDES)*, Gdynia, Poland, Jun 2013, pp. 451–456.
- [21] S. Astapov, J. Berdnikova, and J.-S. Preden, “A two-stage approach to 2D DOA estimation for a compact circular microphone array,” in *Proc. 4th Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, Kitakyushu, Japan, Jun 2015, pp. 1–6.
- [22] S. Astapov, J. Ehala, J. Berdnikova, and J.-S. Preden, “Gunshot acoustic component localization with distributed circular microphone arrays,” in *Proc. 2015 IEEE Int. Conf. on Digital Signal Processing (DSP)*, Singapore, Jul 2015, pp. 1186–1190.
- [23] S. Astapov, J. S. Preden, and J. Berdnikova, “Simplified acoustic localization by linear arrays for wireless sensor networks,” in *Proc. 18th Int. Conf. on Digital Signal Processing (DSP)*, Fira, Santorini, Greece, Jul 2013, pp. 1–6.
- [24] B. Atmaja and D. Arifianto, “Machinery fault diagnosis using independent component analysis (ICA) and instantaneous frequency (IF),” in *Proc. Int. Conf. on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, Bandung, Indonesia, Nov 2009, pp. 1–5.
- [25] S. Bancroft, “An algebraic solution of the gps equations,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-21, no. 1, pp. 56–59, Jan 1985.
- [26] Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2002.
- [27] M. Bedworth and J. O’Brien, “The omnibus model: a new model of data fusion?” *IEEE Aerospace and Electronic Systems Magazine*, vol. 15, no. 4, pp. 30–36, Apr 2000.

- [28] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, ser. Springer Topics in Signal Processing. Springer Berlin Heidelberg, 2008.
- [29] J. Berdnikova, T. Ruuben, V. Kozevnikov, and S. Astapov, “Acoustic noise pattern detection and identification method in Doppler system,” *Electronics and Electrical Engineering*, vol. 18, no. 2, pp. 65–68, 2012.
- [30] D. Blatt and A. O. Hero, “APOCS: a rapidly convergent source localization algorithm for sensor networks,” in *Proc. IEEE/SP 13th Workshop Statistical Signal Processing*, Bordeaux, France, Jul 2005, pp. 1214–1219.
- [31] P. A. Bonnefoy, “Scalability of the air transportation system and development of multi-airport systems: A worldwide perspective,” Ph.D. dissertation, Massachusetts Institute of Technology, 2008.
- [32] A. Borzino, J. Apolinario, and M. de Campos, “Estimating direction of arrival of long range gunshot signals,” in *Proc. Int. Telecommunications Symp. (ITS)*, Sao Paulo, Brazil, Aug 2014, pp. 1–5.
- [33] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, ser. Digital Signal Processing. Springer Berlin Heidelberg, 2013.
- [34] R. Bucher and D. Misra, “A synthesizable VHDL model of the exact solution for three-dimensional hyperbolic positioning system,” *VLSI Design*, vol. 15, no. 2, pp. 507–520, 2002.
- [35] D. Cai, C. Zhang, and X. He, “Unsupervised feature selection for multi-cluster data,” in *Proc. 16th Int. Conf. on Knowledge Discovery and Data Mining*. Washington, DC, USA: ACM, Jul 2010, pp. 333–342.
- [36] D. Carlucci and S. Jacobson, *Ballistics: Theory and Design of Guns and Ammunition*. Taylor & Francis, 2010.
- [37] V. Cevher, R. Chellappa, and J. H. McClellan, “Vehicle speed estimation using acoustic wave patterns,” *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 30–47, Jan. 2009.
- [38] I. Chatzigiannakis, J. P. Drude, H. Hasemann, and A. Kröller, *Developing Smart Homes Using the Internet of Things: How to demonstrate Your System*. Springer International Publishing, 2014, pp. 415–426.
- [39] C.-E. Chen, H. Wang, A. Ali, F. Lorenzelli, R. Hudson, and K. Yao, “Particle filtering approach to localization and tracking of a moving acoustic source in a reverberant room,” in *Proc. IEEE Int. Conf. on*

- Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Toulouse, France, May 2006, pp. IV–849–IV–852.
- [40] X. Chen, S. Xiang, C. L. Liu, and C. H. Pan, “Vehicle detection in satellite images by hybrid deep convolutional neural networks,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1797–1801, Oct 2014.
- [41] M. Cobos, A. Marti, and J. J. Lopez, “A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling,” *Signal Processing Letters, IEEE*, vol. 18, no. 1, pp. 71–74, 2011.
- [42] A. Colombo, T. Bangemann, S. Karnouskos, J. Delsing, P. Stluka, R. Harrison, F. Jammes, and J. Lastra, *Industrial Cloud-Based Cyber-Physical Systems: The IMC-AESOP Approach*, ser. SpringerLink : Bücher. Springer International Publishing, 2014.
- [43] P. Conn, *Imaging and Spectroscopic Analysis of Living Cells: Optical and Spectroscopic Techniques*, ser. Methods in Enzymology. Elsevier Science, 2012.
- [44] T. Damarla, L. Kaplan, and G. Whipps, “Sniper localization using acoustic asynchronous sensors,” *Sensors Journal, IEEE*, vol. 10, no. 9, pp. 1469–1478, Sept 2010.
- [45] B. V. Dasarathy, *Decision Fusion*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1994.
- [46] A. Dhawan, R. Balasubramanian, and V. Vokkarane, “A framework for real-time monitoring of acoustic events using a wireless sensor network,” in *Proc. IEEE Int. Technologies for Homeland Security (HST) Conf.*, Waltham, MA, USA, Nov 2011, pp. 254–261.
- [47] J. DiBiase, “A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays,” Ph.D. dissertation, Brown University, Providence, RI, USA, 2000.
- [48] H. Do and H. F. Silverman, “SRP-PHAT methods of locating simultaneous multiple talkers using a frame of microphone array data,” in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, Dallas, Texas, USA, Mar 2010, pp. 125–128.
- [49] H. Do, H. F. Silverman, and Y. Yu, “A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array,” in *Proc. IEEE Int. Conf. Acoustics,*

Speech and Signal Processing (ICASSP), vol. 1, Honolulu, HI, USA, Apr 2007, pp. 121–124.

- [50] B. Donovan, Y. Li, R. Stern, J. Jiang, C. Claudel, and D. Work, “Vehicle detection and speed estimation with PIR sensors,” in *Proc. 14th Int. Conf. on Information Processing in Sensor Networks*, ser. IPSN '15. Seattle, WA, USA: ACM, Apr 2015, pp. 370–371.
- [51] M. A. A. Elmaleeh, N. Saad, and M. Awan, “Incipient fault detection of industrial pilot plant machinery via acoustic emission,” in *Proc. 8th Int. Mechatronics and its Applications (ISMA) Symp.*, Sharjah, United Arab Emirates, Apr 2012, pp. 1–6.
- [52] P. Eret and C. Meskell, “Compressed air leak detection using microphone array techniques,” in *Leveraging Technology for a Sustainable World*. Springer, 2012, pp. 347–352.
- [53] H. Farhangi, “The path of the smart grid,” *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18–28, January 2010.
- [54] *Crisis Response and Disaster Resilience 2030: Forging Strategic Action in an Age of Uncertainty*, The Federal Emergency Management Agency (FEMA) Std., 2012.
- [55] G. Fink, *Markov Models for Pattern Recognition: From Theory to Applications*, ser. SpringerLink: Springer e-Books. Springer Berlin Heidelberg, 2007.
- [56] I. Freire and J. Apolinario, “GCC-based DoA estimation of overlapping muzzleblast and shockwave components of gunshot signals,” in *Proc. IEEE Second Latin American Symp. on Circuits and Systems (LASCAS)*, Bogota, Colombia, Feb 2011, pp. 1–4.
- [57] M. Frigo and S. G. Johnson, “FFTW: an adaptive software architecture for the FFT,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, Seattle, Washington, USA, May 1998, pp. 1381–1384.
- [58] J. George and L. Kaplan, “Shooter localization using soldier-worn gunfire detection systems,” in *Proc. 14th Int. Conf. on Information Fusion (FUSION)*, Chicago, Illinois, USA, Jul 2011, pp. 1–8.
- [59] J. George, L. Kaplan, S. Deligeorges, and G. Cakiades, “Multi-shooter localization using finite point process,” in *Proc. 17th Int. Conf. on Information Fusion (FUSION)*, Salamanca, Spain, Jul 2014, pp. 1–7.

- [60] A. GhaffarianHoseini, N. D. Dahlan, U. Berardi, A. GhaffarianHoseini, and N. Makaremi, “The essence of future smart houses: From embedding ICT to adapting to sustainability principles,” *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 593 – 607, 2013.
- [61] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics: a foundation for computer science*. Addison-Wesley Reading, MA, 1994, vol. 2.
- [62] L. Gu, D. Jia, P. Vicaire, T. Yan, L. Luo, A. Tirumala, Q. Cao, T. He, J. A. Stankovic, T. Abdelzaher, and B. H. Krogh, “Lightweight detection and classification for wireless sensor networks in realistic environments,” in *Proc. 3rd int. Conf. on Embedded networked sensor systems*, ser. SenSys ’05. San Diego, CA, USA: ACM, Nov 2005, pp. 205–217.
- [63] V. Gunes, S. Peter, T. Givargis, and F. Vahid, “A survey on concepts, applications, and challenges in cyber-physical systems,” *KSII Transactions on Internet and Information Systems*, vol. 8, no. 12, pp. 4242–4268, 2014.
- [64] D. L. Hall and S. A. H. McMullen, *Mathematical Techniques in Multi-sensor Data Fusion*, 2nd ed. Norwood, MA, USA: Artech House, Inc., 2004.
- [65] S. A. Haque, S. M. Aziz, and M. Rahman, “Review of cyber-physical system in healthcare,” *International Journal of Distributed Sensor Networks*, vol. 2014, pp. 1–19, 2014.
- [66] D. J. Henry, “ISR systems: Past, present, and future,” in *Proc. SPIE 9828, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications XIII*, 2016, pp. 982 802–982 802–12.
- [67] J. A. G. Herrera, J. F. Lopez, and H. F. Q. Riaza, “Nostic method based on the analysis of vibration and acoustic emissions for internal combustion engines faults,” in *Proc. XIX Symp. on Image, Signal Processing and Artificial Vision*, Armenia, Colombia, Sept 2014, pp. 1–5.
- [68] H. Hongsen, L. Jing, and G. Yang, “Acoustic direction of arrival estimation based on spatial circular prediction,” in *Proc. Int. Forum on Information Technology and Applications, IFITA ’09*, vol. 3, Chengdu, China, May 2009, pp. 177–180.
- [69] M. Ilyas and I. Mahgoub, *Smart Dust: Sensor Network Applications, Architecture and Design*. CRC Press, 2016.

- [70] U. Iqbal, S. W. Zamir, M. H. Shahid, K. Parwaiz, M. Yasin, and M. S. Sarfraz, "Image based vehicle type identification," in *Proc. Int. Conf. on Information and Emerging Technologies (ICIET)*, Karachi, Pakistan, Jun 2010, pp. 1–5.
- [71] C. T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS*, St. Louis, MO, USA, Oct 2009, pp. 2027–2032.
- [72] J.-S. R. Jang, C. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, 1997.
- [73] H. Kagermann, W. Wahlster, and J. Helbig, "Recommendations for implementing the strategic initiative INDUSTRIE 4.0," Final report of the Industrie 4.0 Working Group, Tech. Rep., 2013.
- [74] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," *Journal of Machine Learning Research*, vol. 10, pp. 1391–1445, 2009.
- [75] O. Kessler, "Functional description of the data fusion process," Office of Naval Technology, Naval Air Development Center, Warminster, PA, Tech. Rep., 1992.
- [76] O. Kessler and B. Fabian, "Estimation and ISR process integration," Defense Advanced Projects Research Agency, Washington D.C., Tech. Rep., 2001.
- [77] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, June 2015.
- [78] Z. Khan, A. Ali, and Z. Riaz, *Computational Intelligence for Decision Support in Cyber-Physical Systems*, ser. Studies in Computational Intelligence. Springer Singapore, 2014.
- [79] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Feature subset selection using differential evolution and a statistical repair mechanism," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11 515–11 526, 2011.
- [80] M. Kuorilehto, M. Kohvakka, J. Suhonen, P. Hämäläinen, M. Hännikäinen, and T. Hamalainen, *Ultra-Low Energy Wireless Sensor Networks in Practice: Theory, Realization and Deployment*. Wiley, 2008.

- [81] P. Langley, "Induction of recursive bayesian classifiers," in *Proc. European Conf. on Machine Learning*, ser. ECML '93. Vienna, Austria: Springer-Verlag, Apr 1993, pp. 153–164.
- [82] J. Lanslots, F. Deblauwe, and K. Janssens, "Selecting sound source localization techniques for industrial applications," *Sound and Vibration*, vol. 44, no. 6, pp. 6–10, 2010.
- [83] E. Lee and S. Seshia, *Introduction to Embedded Systems - A Cyber Physical Systems Approach*, 2nd ed. LULU Press, 2015.
- [84] I. Lee, J. Leung, and S. Son, *Handbook of Real-Time and Embedded Systems*, ser. Chapman & Hall/CRC Computer and Information Science Series. CRC Press, 2007.
- [85] I. Lee and O. Sokolsky, "Medical cyber physical systems," in *Proc. 47th Design Automation Conf.*, ser. DAC '10. Anaheim, CA, USA: ACM, Jun 2010, pp. 743–748.
- [86] S. Li and A. Jain, *Handbook of Face Recognition*, ser. SpringerLink : Bücher. Springer London, 2011.
- [87] U. Libal and K. Spyra, "Wavelet based shock wave and muzzle blast classification for different supersonic projectiles," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5097–5104, 2014.
- [88] M. Liggins, D. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed., ser. Electrical Engineering & Applied Signal Processing Series. CRC Press, 2008.
- [89] D. Lindgren, O. Wilsson, F. Gustafsson, and H. Habberstad, "Shooter localization in wireless sensor networks," in *Proc. 12th Int. Conf. on Information Fusion (FUSION)*, Seattle, WA, USA, Jul 2009, pp. 404–411.
- [90] T. Liu, Y. Liu, X. Cui, G. Xu, and D. Qian, "MOLTS: Mobile object localization and tracking system based on wireless sensor networks," in *Proc. IEEE 7th Int. Networking, Architecture and Storage (NAS) Conf.*, Xiamen, China, Jun 2012, pp. 245–251.
- [91] Y. Liu and Z. Yang, *Location, localization, and localizability: location-awareness technology for wireless networks*. Springer Science & Business Media, 2010.
- [92] E. Lupu, Z. Feher, and P. G. Pop, "On the speaker verification using the TESPARG coding method," in *Proc. Int. Symp. on Signals, Circuits and Systems*, vol. 1, Bangkok, Thailand, May 2003, pp. 173–176.

- [93] C. Mahlander and D. Rosenholm, *Rule-based Identification of Revision Objects in Satellite Images*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 219–228.
- [94] D. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, 1st ed. Springer Publishing Company, Inc., 2008.
- [95] E. Mangas and A. Bilas, “FLASH: Fine-grained localization in wireless sensor networks using acoustic sound transmissions and high precision clock synchronization,” in *Proc. 29th IEEE Int. Conf. Distributed Computing Systems (ICDCS)*, Montreal, Canada, Jun 2009, pp. 289–298.
- [96] P. Manolitzas, *Effective Methods for Modern Healthcare Service Quality and Evaluation*, ser. Advances in Medical Diagnosis, Treatment, and Care. IGI Global, 2016.
- [97] F. Masson, D. Puschini, P. Julian, P. Croce, L. Arlenghi, P. Mandolesi, and A. G. Andreou, “Hybrid sensor network and fusion algorithm for sound source localization,” in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Kobe, Japan, May 2005, pp. 2763–2766.
- [98] J. A. Mazurek, J. E. Barger, M. Brinn, R. J. Mullen, D. Price, S. E. Ritter, and D. Schmitt, “Boomerang mobile counter shooter detection system,” in *Proc. SPIE 5778, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV*, vol. 5778, 2005, pp. 264–282.
- [99] W. Meng, R. Ma, and H. H. Chen, “Smart grid neighborhood area networks: a survey,” *IEEE Network*, vol. 28, no. 1, pp. 24–32, January 2014.
- [100] Z. Merhi, M. Elgamel, and M. Bayoumi, “Acoustic target localization in sensor networks with FUZZYART,” in *Proc. 50th Midwest Symp. Circuits and Systems (MWSCAS)*, Montreal, QC, Canada, Aug 2007, pp. 1536–1539.
- [101] Z. Merhi, M. Elgamel, and M. Bayoumi, “A lightweight collaborative fault tolerant target localization system for wireless sensor networks,” *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1690–1704, 2009.
- [102] J. Millet and B. Baligand, “Latest achievements in gunfire detection systems,” DTIC Document, Tech. Rep., 2006.

- [103] O. Minck, N. Binder, O. Cherrier, L. Lamotte, and V. Pommier-Budinger, “Fan noise analysis using a microphone array,” in *Proc. Int. Conf. on Fan Noise, Technology, and Numerical Methods*, Senlis, France, Apr 2012.
- [104] J. A. Momoh, “Fundamentals of analysis and computation for the smart grid,” in *Proc. IEEE Power and Energy Society General Meeting*, Minneapolis, MN, USA, Jul 2010, pp. 1–5.
- [105] K. Murphy and S. Russell, “Rao-blackwellised particle filtering for dynamic bayesian networks,” in *Sequential Monte Carlo methods in practice*. Springer, 2001, pp. 499–515.
- [106] O. Niggemann and J. Beyerer, *Machine Learning for Cyber Physical Systems: Selected papers from the International Conference ML4CPS 2015*, ser. Technologien für die intelligente Automation. Springer Berlin Heidelberg, 2016.
- [107] I. Nikolaidis and K. Iniewski, *Building Sensor Networks: From Design to Applications*, ser. Devices, Circuits, and Systems. Taylor & Francis, 2013.
- [108] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer verlag, 1999.
- [109] M. Noda, T. Takahashi, D. Deguchi, I. Ide, H. Murase, Y. Kojima, and T. Naito, “Road image update using in-vehicle camera images and aerial image,” in *Proc. Intelligent Vehicles Symp. (IV), 2011 IEEE*, Baden-Baden, Germany, Jun 2011, pp. 460–465.
- [110] S. Nof, *Springer Handbook of Automation*, ser. Springer Handbook of Automation. Springer Berlin Heidelberg, 2009.
- [111] H. J. Ortiz-Pena, R. Nagi, M. Sudit, M. D. Moskal, M. Dawson, J. Fink, T. Hanratty, E. Heilman, and D. Tuttle, “From information needs to information gathering: a system optimization perspective to ISR synchronization,” in *Proc. SPIE 8389, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III*, 2012, pp. 838 912–838 912–12.
- [112] R. Osborne, Y. Bar-Shalom, J. George, and L. Kaplan, “Data fusion from multiple passive sensors for multiple shooter localization via assignment,” in *Proc. 17th Int. Conf. on Information Fusion (FUSION)*, Salamanca, Spain, Jul 2014, pp. 1–7.

- [113] D. W. Patterson, *Artificial Neural Networks: Theory and Applications*, ser. Prentice-Hall Series in Advanced Communications. Prentice Hall, 1996.
- [114] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, “Real-time multiple sound source localization and counting using a circular microphone array,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2193–2206, Oct 2013.
- [115] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the CUIDADO project,” CUIDADO I.S.T. Project Report, Tech. Rep., 2004.
- [116] J. Preden, J. Kaugerand, E. Suurjaak, S. Astapov, L. Motus, and R. Pahtma, “Data to decision: pushing situational information needs to the edge of the network,” in *Proc. IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision (CogSIMA)*, Orlando, USA, Mar 2015, pp. 158–164.
- [117] J.-S. Preden, J. Kaugerand, E. Suurjaak, R. Pahtma, S. Astapov, and L. Motus, “Uav as a generic information provider in an isr system of systems,” in *Proc. SPIE, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV: SPIE Defense, Security and Sensing*, 2015, in press.
- [118] J. S. Preden, L. Motus, R. Pahtma, and M. Meriste, “Data exchange for shared situation awareness,” in *Proc. IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, New Orleans, LA, USA, Mar 2012, pp. 198–201.
- [119] J. Preden, L. Motus, R. Pahtma, and M. Meriste, “Reducing bandwidth requirements and optimizing data flow in distributed data acquisition and processing,” in *Proc. 2013 IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, San Diego, CA, USA, Feb 2013, pp. 175–182.
- [120] J. Preden, R. Pahtma, S. Astapov, J. Ehala, A. Riid, and L. Motus, “Distributed fusion and automated sensor tasking in ISR systems,” in *Proc. SPIE 9079, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR V, 90790M*, vol. 9079, 2014, pp. 90 790M–90 790M–10.
- [121] J. S. Preden, J. Llinas, G. Rogova, R. Pahtma, and L. Motus, “On-line data validation in distributed data fusion,” in *Proc. SPIE Defense, Security and Sensing, Ground/Air Multisensor Interoperability, Integra-*

- tion, and Networking for Persistent ISR IV, vol. 8742. International Society for Optics and Photonics, 2013, pp. 1–12.
- [122] F. Qu, F. Y. Wang, and L. Yang, “Intelligent transportation spaces: vehicles, traffic, communications, and beyond,” *IEEE Communications Magazine*, vol. 48, no. 11, pp. 136–142, November 2010.
- [123] D. Rawat, J. Rodrigues, and I. Stojmenovic, *Cyber-Physical Systems: From Theory to Practice*. Taylor & Francis, 2015.
- [124] A. Riid and J. S. Preden, “A combined method for error and complexity reduction in fuzzy rule-based classification,” in *Proc. 2015 IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, Istanbul, Turkey, Aug 2015, pp. 1–6.
- [125] A. Riid and E. Rustern, “An integrated approach for the identification of compact, interpretable and accurate fuzzy rule-based classifiers from data,” in *Proc. 15th IEEE Int. Intelligent Engineering Systems (INES) Conf.*, Poprad, Slovakia, Jun 2011, pp. 101–107.
- [126] A. Riid and J. Preden, “Interpretability improvement of fuzzy rule-based classifiers via rule compression,” in *Proc. 2015 Conf. of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology*, ser. Advances in Intelligent Systems Research. Gijon, Spain: Atlantis Press, Jun 2015, pp. 162–169.
- [127] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications*. World Scientific Publishing Company, 2014.
- [128] J. Sallai, P. Volgyesi, A. Ledeczki, K. Pence, T. Bapty, S. Neema, and J. Davis, “Acoustic shockwave-based bearing estimation,” in *Proc. ACM/IEEE Int. Conf. on Information Processing in Sensor Networks (IPSN)*, Philadelphia, PA, USA, Apr 2013, pp. 217–228.
- [129] J. Sallai, P. Volgyesi, K. Pence, and A. Ledeczki, “Fusing distributed muzzle blast and shockwave detections,” in *Proc. 14th Int. Conf. on Information Fusion (FUSION)*, Chicago, Illinois, USA, Jul 2011, pp. 1–8.
- [130] A. Sansan, L. Zhen, Z. Nan, B. Xianzheng, and L. Yang, “Microphone arrays for acoustic detection during laser welding of cold-rolled steel strip,” in *Proc. 2010 Int. Conf. on Intelligent System Design and Engineering Application*, vol. 1. Changsha, Hunan, China: IEEE, Oct 2010, pp. 935–939.

- [131] V. Saxena, *The Amazing Growth and Journey of UAV's and Ballistic Missile Defence Capabilities: Where the Technology is Leading to?*, ser. 1. Vij Books India Private Limited, 2013.
- [132] C. Schweitzer, K. Stein, and N. Wendelstein, "Evaluation of appropriate sensor specifications for space based ballistic missile detection," in *Proc. SPIE 8541, Electro-Optical and Infrared Systems: Technology and Applications IX, 85410M*, vol. 8541, 2012, pp. 85 410M–85 410M–11.
- [133] B. Sheno, *Introduction to Digital Signal Processing and Filter Design*. Wiley, 2005.
- [134] G. Siddesh, G. Deka, K. Srinivasa, and L. Patnaik, *Cyber-Physical Systems: A Computational Perspective*. CRC Press, 2015.
- [135] *Implementing 21st Century Smart Manufacturing*, Smart Manufacturing Leadership Coalition (SMLC) Std., 2011.
- [136] M. Smith, K. Kim, and D. Thompson, "Noise source identification using microphone arrays," in *Proc. Institute of Acoustics*, vol. 29, Oxford, UK, Oct 2007.
- [137] H. Song, D. Rawat, S. Jeschke, and C. Brecher, *Cyber-Physical Systems: Foundations, Principles and Applications*, ser. Intelligent Data-Centric Systems: Sensor Collected Intelligence. Elsevier Science, 2016.
- [138] Y. Song, S. Rahardja, and R. Yu, "Beamforming performance of circular microphone array on spherical platform near bottom boundary," in *Proc. 2010 IEEE Int. Conf. on Multimedia and Expo (ICME)*, Singapore, Jul 2010, pp. 743–747.
- [139] J. A. Stankovic, I. Lee, A. Mok, and R. Rajkumar, "Opportunities and obligations for physical computing systems," *IEEE Computer Society*, vol. 38, no. 11, pp. 23–31, Nov 2005.
- [140] S. Suh, U. Tanik, J. Carbone, and A. Eroglu, *Applied Cyber-Physical Systems*. Springer New York, 2013.
- [141] A. Sutin, B. Bunin, A. Sedunov, N. Sedunov, L. Fillinger, M. Tsion-skiy, and M. Bruno, "Stevens passive acoustic system for underwater surveillance," in *Proc. 2010 Int. WaterSide Security Conf.*, Carrara, Italy, Nov 2010, pp. 1–6.
- [142] L. Tan and J. Jiang, *Fundamentals of Analog and Digital Signal Processing*. AuthorHouse, 2007.

- [143] P. Thenkabail, *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*, ser. Remote Sensing Handbook. CRC Press, 2015.
- [144] P. J. Tremblay, “Shaping and adapting: Unlocking the power of Colonel John Boyd’s OODA loop,” Ph.D. dissertation, United States Marine Corps, Command and Staff College, Marine Corps University, 2015.
- [145] E. Trentin and M. Gori, *Combining Neural Networks and Hidden Markov Models for Speech Recognition*. London: Springer London, 1999, pp. 63–79.
- [146] G. Vakulya and G. Simon, “Fast adaptive acoustic localization for sensor networks,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1820–1829, 2011.
- [147] S. Vaseghi, *Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications*. Wiley, 2007.
- [148] M. Walpola, H. Zhu, and J. Zhang, “Self organization algorithm for unattended acoustic sensor networks in ground target tracking,” in *Proc. IEEE Conf. on Wireless Communications and Networking*, Hong Kong, China, Mar 2007, pp. 2350–2354.
- [149] Q. Wang, R. Zheng, A. Tirumala, X. Liu, and L. Sha, “Lightning: A hard real-time, fast, and lightweight low-end wireless sensor election protocol for acoustic event localization,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 570–584, 2008.
- [150] P. E. William and M. W. Hoffman, “Classification of military ground vehicles using time domain harmonics’ amplitudes,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3720–3731, Nov 2011.
- [151] L. Yan, Y. Zhang, L. Yang, and H. Ning, *The Internet of Things: From RFID to the Next-Generation Pervasive Networked Systems*, ser. Wireless Networks and Mobile Communications. CRC Press, 2008.
- [152] S.-H. Yang, *Wireless Sensor Networks: Principles, Design and Applications*. Springer Publishing Company, Incorporated, 2013.
- [153] J. Yick, B. Mukherjee, and D. Ghosal, “Analysis of a prediction-based mobility adaptive tracking algorithm,” in *Proc. 2nd Int. Conf. on Broadband Networks (BroadNets)*, vol. 1, Boston, MA, USA, Oct 2005, pp. 753–760.

- [154] D. Yu and L. Deng, *Automatic Speech Recognition: A Deep Learning Approach*, ser. Signals and Communication Technology. Springer London, 2014.
- [155] Z.-J. Yu, G.-X. You, J.-M. Wei, and H.-T. Liu, “Multi-model rao-blackwellised particle filter for maneuvering target tracking in distributed acoustic sensor networks,” in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, Honolulu, HI, USA, Apr 2007, pp. III–1213–III–1216.
- [156] S. Zeadally and N. Jabeur, *Cyber-Physical System Design with Sensor Networking Technologies*. Institution of Engineering & Technology, 2015.
- [157] S. Zhang, G. Li, W. Wei, and B. Yang, “A novel iterative multilateral localization algorithm for wireless sensor networks,” *Journal of Networks*, vol. 5, no. 1, pp. 112–119, 2010.
- [158] X. Zhao, J. Tang, L. Zhou, and Z. Wu, “Accelerated steered response power method for sound source localization via clustering search,” *Science China: Physics, Mechanics and Astronomy*, vol. 56, no. 7, pp. 1329–1338, 2013.
- [159] F. Zheng, G. Zhang, and Z. Song, “Comparison of different implementations of MFCC,” *Journal of Computer Science and Technology*, vol. 16, no. 6, pp. 582–589, 2001.
- [160] Y. Zheng and E. Blasch, “Multispectral image fusion for vehicle identification and threat analysis,” in *Proc. SPIE 9871, Sensing and Analysis Technologies for Biomedical and Cognitive Applications*, 2016, pp. 98 710G–98 710G–12.
- [161] X. Zhong and J. R. Hopgood, “A time-frequency masking based random finite set particle filtering method for multiple acoustic source detection and tracking,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2356–2370, Dec 2015.
- [162] H. Zhou, *The Internet of Things in the Cloud: A Middleware Perspective*. CRC Press, 2015.
- [163] M. Zivanovic, A. Röbel, and X. Rodet, “Adaptive threshold determination for spectral peak classification,” *Computer Music Journal*, vol. 32, no. 2, pp. 57–67, Jun. 2008.

Abstract

The thesis is devoted to the development of signal processing methods applicable in distributed Cyber-Physical Systems (CPS). CPS are combinations of computational, communication and physical processes, integrated into the physical environment of the system operation domain. CPS are primarily feedback systems, consisting of many distributed interconnected components, which gather information about the state of the environment and the objects of interaction in order to determine the appropriate system response. For this purpose the distributed CPS sensor components perform situation assessment of the observed area, produce local assessments independently and transmit gathered data through a Wireless Sensor Network (WSN) to information aggregation and fusion components, where a global situation assessment is made.

The thesis discusses methods of signal processing that allow for both local and global situation assessment, while accounting for limitations situated with reduced computational power of sensor component embedded hardware and problems of data communication through WSN with ad-hoc topologies. The embedded hardware limitations are overcome by developing optimized signal processing procedures, which require less computational resources, compared to the state-of-the-art methods. WSN communication problems are overcome by adopting an autonomous data fusion architecture, where signal processing and initial situation assessment is performed on individual sensor components, and no raw signal transmission is performed between WSN nodes. The developed data fusion methods are aimed at operation in an asynchronous WSN with an undetermined number of active nodes at any given time instance.

The thesis focuses on the analysis of one-dimensional band-limited signals. The proposed methods are tested primarily on acoustic and vibration signals, however, they are also applicable to signals of other modalities. The discussed tasks of situation assessment include process state identification and moving object detection, classification, localization and trajectory estimation. The chapter devoted to single sensor solutions discusses the approaches to state identification of stationary and non-stationary pro-

cesses, where several feature selection and extraction methods are presented, and pattern recognition is performed by correlation analysis and fuzzy logic based classification. For non-stationary process state identification, a multi-stage procedure, comprising several time and frequency domain signal feature analysis steps, is proposed. Multi-channel solutions include methods for Direction of Arrival (DOA) estimation and distributed acoustic localization of objects in the horizontal plane. For trajectory estimation an approach based on recursive Bayesian estimators, such as the Kalman Filter, is proposed. The application of trajectory estimation reassures consecutive localization estimates and allows to predict the search region of the localized object.

The proposed single and multi-channel signal processing methods are implemented on embedded hardware and proven to operate in real-time. The implemented standalone smart sensors are used for both civil and military applications during *in situ* experiments. The methods of stationary process state identification are applied to industrial machinery monitoring and malfunction detection. Vehicle identification, DOA estimation, object localization and trajectory estimation methods are applied to military vehicle tracking and shooter acoustic localization. An application-specific method of foot soldier speed estimation using passive infrared sensors is also presented.

Experimental results confirm high accuracy of the proposed methods and their applicability in real life conditions. The results also indicate that precise synchronization between WSN nodes is not required for the proposed methods to operate with high precision. Though the harmful influence of background noise on the detection and identification accuracy of methods based on acoustic signal analysis is not eliminated to full extent, the results of these methods can be reassured and corrected by sensor components of other modalities in the future.

Kokkuvõte

Käesoleva väitekirja teemaks on signaalitöötlemise meetodite arendus hajutatud küberfüüsikalistes süsteemides (KFS). KFSd koosnevad omavahel ühendatud arvutuslikest, kommunikatsiooni- ja füüsilistest protsessidest, mis on integreeritud süsteemi füüsilisse keskkonda. KFSd on peamiselt tagasidesetatud süsteemid, mis koosnevad mitmetest hajutatud ja sidestatud komponentidest. Need komponendid koguvad keskkonna oleku ja interaktsioonis olevate objektide kohta informatsiooni, mille alusel omakorda määratakse kindlaks süsteemi sobilik reageering. Hajutatud sensorikomponendid tegelevad nende poolt jälgitud ala lokaalse olukorra hindamisega ning edastavad informatsiooni läbi traadita sensorvõrgu (WSN) andmekogumise ja ühildamise komponentidele, kus viiakse läbi globaalse olukorra hindamine.

Väitekiri käsitleb signaalitöötlemise meetodeid, mis võimaldavad lokaalse ja globaalse olukorra hindamist, võttes arvesse piiranguid sardsüsteemide arvutusvõimsusele ja spontaanvõrkude topoloogiatega seotud probleeme. Piirangud sardsüsteemide arvutusvõimsusele on leevendatavad optimeeritud signaalitöötlemise protseduuridega, mis tarbivad vähem arvutuslike ressursse võrreldes standardmeetoditega. Infoedastusprobleeme spontaanvõrkudes leevendab autonoomne arhitektuur, kus signaalitöötlus ja esialgse olukorra hindamine on teostatud iseseisvatel sensorikomponentidel ning toorsignaali edastamine läbi WSNi sõlmede vajalik ei ole. Väljaarendatud andmete ühildamise meetodid on mõeldud kasutamiseks ette kindlaks tegemata või muutuva arvu sõlmedega asünkroonsetes spontaanvõrkudes.

Väitekiri keskendub ühemõõtmeliste piiratud sagedusribaga signaalide analüüsile. Väljapakutud signaalitöötlemise meetodeid on katsetatud peamiselt akustilistel ja vibratsioonisignaalidel, kuid need meetodid on samuti kohaldatavad muud tüüpi signaalide töötlemiseks. Käsitletud olukorrahindamise ülesanded hõlmavad protsessi olekute identifitseerimist ning liikuvate objektide avastamist, klassifitseerimist, lokaliseerimist ja trajektoori jälgimist. Ühe anduriga lahendustele pühendatud peatükk käsitleb statsionaarsete ja mittestatsionaarsete protsesside olekute identifitseerimismeetodeid. Lähemalt on vaadeldud mitmeid eriomaste tunnusjoonte valimise ja eraldamise meetodeid ning korrelatsioonianalüüsil ning hädakõrgel loogikal baseeruvaid mustrite

tuvastamise protseduure. Mittestatsionaarsete protsesside olekute identifitseerimiseks on välja pakutud mitmeetapiline protseduur, mille hierarhiline struktuur koosneb mitmetest aja- ning sagedusdomeeni signaali eriomaste tunnusjoonte analüüsi etappidest. Mitme anduriga lahendused hõlmavad heliobjektide suunatu vastuse ja hajutatud akustilise lokaliseerimise meetodeid horisontaaltasandil. Lokaliseeritud objektide trajektoori määramiseks on välja pakutud rekursiivsel Bayesi hindajal (nt. Kalman filter) baseeruv meetod. Trajektoori määramine võimaldab valideerida järjestikusi lokaliseerimisi ning ennustada objekti paiknemispiirkonda.

Väljapakutud ühe ja mitme anduriga signaalitöötuse meetodid on realiseeritud sardsüsteemides ning nende töövõime reaalses on tõestatud praktikas. Väljatöötatud iseseisvaid nutisensoreid on välikatsete käigus testitud tsiviil- ja militaarrakendustes. Statsionaarsete protsesside olekute identifitseerimismeetodeid on kasutatud tööstuslike seadmete monitooringul ja rikete avastamisel. Sõidukite identifitseerimise, objektide suunamääramise, lokaliseerimise ja trajektoori jälgimise meetodeid on rakendatud militaarsõidukite äratundmisel ja jälgimisel ning tulistaja asukoha määramisel. Samuti on välja töötatud jalaväe liikumiskiiruse määramiseks passiivseid infrapunaandureid kasutatav meetod.

Katsetulemused demonstreerivad väljapakutud meetodite kõrget hindamistäpsust ning töövõimet reaalses tingimustes. Samuti näitavad katsetulemused, et täpne sünkroniseerimine WSN sõlmede vahel ei ole vajalik väljapakutud meetodite kõrge hindamistäpsuse tagamiseks. Kuigi taustamüra kahjulikku mõju akustiliste signaalide analüüsil baseeruvate meetodite avastamis- ja identifitseerimistäpsusele ei ole võimalik täies ulatuses kõrvaldada, saab nende meetodite hindamistulemusi garanteerida ja parandada muid modaalsusi kasutavate sensorikomponentide abil.

Acknowledgments

The author wishes to express his deep gratitude to his supervisors, Dr. Jürigo-Sören Preden and Dr. Andri Riid, for their support and guidance throughout the thesis work and beyond that; all his work and research colleagues at the Laboratory for Proactive Technologies, and especially Mr. Johannes Ehala, Mr. Jaanus Kaugerand, and Mr. Erki Suurjaak, for collaboration in device prototype implementation, assistance in conducting in-field experiments, and fruitful scientific discussions; his research colleagues Mrs. Julia Berdnikova, Dr. Aleksei Tepljakov and Dr. Eduard Petlenkov for scientific discussions and insightful comments related to this work.

The author would like to thank his family and especially his mother for their lifelong support.

This work was partially supported by the European Defense Agency project IN4STARS, by the Estonian Doctoral School in Information and Communication Technology, and by the Estonian IT Academy program.

Elulookirjeldus

1. Isikuandmed

Ees- ja perekonnanimi	Sergei Astapov
Sünniaeg ja -koht	20.05.1988, Tallinn, Eesti
Kodakondsus	Eesti
E-posti aadress	sergei.astapov@ttu.ee

2. Hariduskäik

Õppeasutus (nimetus lõpetamise ajal)	Lõpetamise aeg	Haridus (eriala/kraad)
Tallinna Tehnikaülikool	2011	Arvuti- ja süsteemitehnika, M.Sc., Cum Laude
Tallinna Tehnikaülikool	2009	Arvuti- ja süsteemitehnika, B.Sc.

3. Keelteoskus (alg-, kesk- või kõrgtase)

Keel	Tase
Eesti	kõrgtase
Inglise	kõrgtase
Vene	emakeel
Saksa	algtase

4. Teenistuskäik

Töötamise aeg	Tööandja nimetus	Ametikoht
2012 – ...	Automaatikainstituut, TTÜ	Insener
2010 – 2013	IMECC OÜ	Insener

5. Projektid

Projekt	Kirjeldus
VA598 Informatsiooni ühilduvus ja luure ühilduvus, kasutades statistikat, agente, arutlemist ja semantikat (15.09.2013 – 15.09.2016)	Signaalitöötlusmeetodite arendamine, akustiliste signaalide analüüs, akustiliste mustrite leidmine ja nende tunnusjoonte eraldamine klassifitseerimise jaoks. Akustiliste massiivide arendamine, akustilise signaali allika hajutatud lokaliseerimisprotseduuri loomine. Jälgimise ja trajektoormääramise ülesannete lahendamine.
AR12139 Targad komposiitmaterjalid: projekteerimine ja valmistamine (01.07.2012 – 30.06.2015)	Andmehõive ja signaalitötluse ülesanded.
SF0140113As08 Proaktiivsus ja situatsiooniteadlikkus (01.01.2008 – 31.12.2013)	Akustiliste ja vibratsiooni signaalide andmehõive ja töötlus. Mustrite leidmine ja klassifitseerimine.

6. Kaitstud lõputööd

Rotatsioonpöördpendli mittelineaarse mudeli juhtimine hägusa regulaatoriga, B.Sc., Tallinna Tehnikaülikool, 2009.

Piiratud sagedusribaga signaalide eriomaste tunnusjoonte eraldamine ja klassifitseerimine, M.Sc., Tallinna Tehnikaülikool, 2011.

7. Teadustöö põhisuunad

Digitaalse signaalitötluse algoritmide arendamine ja realiseerimine sard-süsteemides. Hajutatud signaalitötluse ja otsuste tegemise protseduuride loomine küberfüüsikaliste süsteemide kontekstis (nutisensorid, traadita sensorvõrgud). Mitme anduriga süsteemid, akustiline lokaliseerimine, objektide trajektoori jälgimine ja ennustamine.

8. Teadustegevus

Ajakirja- ja konverentsiartiklite loetelu on toodud ingliskeelses elulookirjelduses.

Curriculum Vitae

1. Personal Data

Name	Sergei Astapov
Date and place of birth	20.05.1988, Tallinn, Estonia
E-mail address	sergei.astapov@ttu.ee

2. Education

Educational institution	Graduation year	Education (field of study/degree)
Tallinn University of Technology	2011	Computer and Systems Engineering, M.Sc., Cum Laude
Tallinn University of Technology	2009	Computer and Systems Engineering, B.Sc.

3. Language competence/skills (fluent, average, basic skills)

Language	Level
Estonian	fluent
English	fluent
Russian	native
German	basic skills

4. Professional employment

Period	Organization	Position
2012 – ...	Department of Computer Control, TUT	Engineer
2010 – 2013	IMECC Ltd.	Engineer

5. Projects

Project	Description
VA598 INformation INteroperability & INtelligence Interoperability by STatistics, Agents, Reasoning and Semantics (15.09.2013 – 15.09.2016)	Signal processing method development, acoustic signal analysis, acoustic pattern discovery and feature extraction for the purposes of classification. Development of acoustic sensor arrays and a distributed acoustic localization and trajectory estimation procedure.
AR12139 Smart composites: design and manufacturing (1.07.2012 – 30.06.2015)	Solving of data acquisition and signal processing problems.
SF0140113As08 Proactivity and situation-awareness (1.01.2008 – 31.12.2013)	Acoustic and vibration signal acquisition and processing. Pattern discovery and classification.

6. Defended theses

Fuzzy Control of a Nonlinear Single Rotational Inverted Pendulum Model,
B.Sc., Tallinn University of Technology, 2009.

*Feature Extraction from Band-Limited Signals and Classification of
Features*, M.Sc., Tallinn University of Technology, 2011.

7. Main areas of scientific work

Development of digital signal processing algorithms and implementa-
tion in embedded systems. Distributed signal processing and coopera-
tive decision-making procedures through signal and information fusion
in the context of cyber-physical systems (smart sensors, Wireless Sen-
sor Networks). Multi-sensor solutions and multi-modal signal analysis.
Acoustic localization, object trajectory estimation and prediction.

8. Scientific work

1. S. Astapov and A. Riid, “A hierarchical algorithm for moving
vehicle identification based on acoustic noise analysis,” in *Proc.
19th Int. Conf. Mixed Design of Integrated Circuits and Systems
(MIXDES)*, Warsaw, Poland, May 2012, pp. 467–472.
2. S. Astapov, J.-S. Preden, T. Aruväli, and B. Gordon, “Production
machinery utilization monitoring based on acoustic and vibration

- signal analysis,” in *Proc. 8th International DAAAM Baltic Industrial Engineering Conference*, Tallinn, Estonia, Apr 2012, pp. 268–273.
3. J. Berdnikova, T. Ruuben, V. Kozevnikov, and S. Astapov, “Acoustic noise pattern detection and identification method in Doppler system,” *Electronics and Electrical Engineering*, vol. 18, no. 2, pp. 65–68, 2012.
 4. S. Astapov, J.-S. Preden, and E. Suurjaak, “A method of real-time mobile vehicle identification by means of acoustic noise analysis implemented on an embedded device,” in *Proc. 13th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2012, pp. 283–286.
 5. S. Astapov and A. Riid, “A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 2, pp. 151–160, 2013.
 6. S. Astapov, J. Berdnikova, and J.-S. Preden, “A method of initial search region reduction for acoustic localization in distributed systems,” in *Proc. 20th Int. Conf. Mixed Design of Integrated Circuits and Systems (MIXDES)*, Gdynia, Poland, Jun 2013, pp. 451–456.
 7. A. Tepljakov, E. Petlenkov, J. Belikov, and S. Astapov, “A digital fractional-order control of a position servo,” in *Proc. 20th Int. Conf. Mixed Design of Integrated Circuits and Systems (MIXDES)*, Gdynia, Poland, Jun 2013, pp. 462–467.
 8. S. Astapov, J.-S. Preden, and J. Berdnikova, “Simplified acoustic localization by linear arrays in wireless sensor networks,” in *Proc. 18th Int. Conf. on Digital Signal Processing (DSP)*, Fira, Santorini, Greece, Jul 2013, pp. 1–6.
 9. C. Ibala, S. Astapov, F. Bettens, F. Escobar, X. Chang, C. Valderama, and A. Riid, “Combining multiple sound sources localization hybrid algorithm and fuzzy rule based classification for real-time speaker tracking application,” *International Journal of Microelectronics and Computer Science*, vol. 4, no. 1, pp. 12–25, 2013.
 10. S. Astapov, J. Berdnikova, and J.-S. Preden, “Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
 11. A. Tepljakov, E. Petlenkov, J. Belikov, and S. Astapov, “Tuning and digital implementation of a fractional-order PD controller for

- a position servo,” *International Journal of Microelectronics and Computer Science*, vol. 4, no. 3, pp. 116–123, 2013.
12. J.-S. Preden, L. Motus, J. Llinas, R. Pahtma, R. Savimaa, M. Meriste, and S. Astapov, “Improvised explosive devices in asymmetric conflicts: multisource data fusion for providing situational information,” A. Gorod, B. E. White, V. Ireland, S. J. Gandhi, B. Sauser, Eds., *Case Studies in System of Systems, Enterprise Systems, and Complex Systems Engineering*, pp. 407–443, Taylor & Francis, 2014.
 13. S. Astapov, J. Ehala, and J.-S. Preden, “Collective acoustic localization in a network of dual channel low power devices,” in *Proc. 21st Int. Conf. Mixed Design of Integrated Circuits and Systems (MIXDES)*, Lublin, Poland, Jun 2014, pp. 430–435.
 14. S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, “Object detection for military surveillance using distributed multimodal smart sensors,” in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP)*, Hong Kong, China, Aug 2014, pp. 366–371.
 15. A. Riid, J.-S. Preden, and S. Astapov, “Detection, identification and tracking of mobile objects with distributed system of systems,” in *Proc. 9th Int. Conf. on System of Systems Engineering (SOSE)*, Adelaide, SA, Australia, Jun 2014, pp. 224–229.
 16. J.-S. Preden, R. Pahtma, S. Astapov, A. Riid, E. Suurjaak, J. Ehala, and L. Motus, “Distributed fusion and automated sensor tasking in ISR systems,” in *Proc. SPIE 9079, Ground/Air Multi-sensor Interoperability, Integration, and Networking for Persistent ISR IV: SPIE Defense, Security and Sensing*, 2014, pp. 90790M–90790M-10.
 17. S. Astapov, A. Riid, J.-S. Preden, and T. Aruvali, “Industrial process monitoring by multi-channel acoustic signal analysis,” in *Proc. 14th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2014, pp. 209–212.
 18. S. Astapov, J. Berdnikova, and J.-S. Preden, “Predictive acoustic localization and speaker tracking for distributed sensor networks,” in *Proc. 13th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV)*, Singapore, Dec 2014, pp. 833–838.
 19. J.-S. Preden, J. Kaugerand, E. Suurjaak, S. Astapov, L. Motus, and R. Pahtma, “Data to decision: pushing situational information needs to the edge of the network,” in *Proc. Int. Inter-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, Orlando, USA, Mar 2015, pp. 158–164.

20. S. Astapov, J. Berdnikova, and J.-S. Preden, "A two-stage approach to 2D DOA estimation for a compact circular microphone array," in *Proc. Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, Kitakyushu, Japan, Jun 2015, pp. 1–6.
21. S. Astapov, J. Ehala, J. Berdnikova, and J.-S. Preden, "Gunshot acoustic component localization with distributed circular microphone arrays," in *Proc. Int. Conf. on Digital Signal Processing (DSP)*, Singapore, Jul 2015, pp. 1186–1190.
22. S. Astapov, J. Berdnikova, J. Ehala, and J. S. Preden, "Shooter localization by networked multichannel acoustic ground sensors," in *Proc. Int. Conf. on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, Oct 2015, pp. 332–337.
23. J. Kaugerand, J.-S. Preden, E. Suurjaak, S. Astopov, L. Motus, and R. Pahtmaa, "A system of systems solution for perimeter control: combining unmanned aerial system with unattended ground sensor network," in *Proc. 9th Annual IEEE International Systems Conference (SysCon)*, Vancouver, BC, Canada, Apr 2015, pp. 317–323.
24. J.-S. Preden, J. Kaugerand, E. Suurjaak, R. Pahtma, S. Astapov, and L. Motus, "UAV as a generic information provider in an ISR system of systems," in *Proc. SPIE, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV: SPIE Defense, Security and Sensing*, 2015, in press.
25. S. Astapov, J. Ehala, and J.-S. Preden, "Performing acoustic localization in a network of embedded smart sensors," *International Journal of Microelectronics and Computer Science (IJMCS)*, vol. 6, no. 3, pp. 86–95, 2015.
26. S. Astapov, J. Berdnikova, J. Ehala, J. Kaugerand, and J.-S. Preden, "Gunshot acoustic event identification and shooter localization in a WSN of asynchronous multichannel acoustic ground sensors," *Multidimensional Systems and Signal Processing*, 2016, in press.
27. S. Astapov, A. Riid, and J.-S. Preden, "Military vehicle acoustic pattern identification by distributed ground sensors," in *Proc. 15th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2016, in press.
28. A. Tepļakov, S. Astapov, and D. Draheim, "Sound localization and processing for inducing synesthetic experiences in virtual reality," in *Proc. 15th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2016, in press.

Publications

Publication 1

Reference

S. Astapov and A. Riid, “A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 2, pp. 151–160, 2013.

Abstract

Mobile vehicle identification has a wide application field for both civilian and military uses. Vehicle identification may be achieved by incorporating single or multiple sensor solutions and through data fusion. This paper considers a single-sensor multistage hierarchical algorithm of acoustic signal analysis and pattern recognition for the identification of mobile vehicles in an open environment. The algorithm applies several standalone techniques to enable complex decision-making during event identification. Computationally inexpensive procedures are specifically chosen in order to provide real-time operation capability. The algorithm is tested on pre-recorded audio signals of civilian vehicles passing the measurement point and shows promising classification accuracy. Implementation on a specific embedded device is also presented and the capability of real-time operation on this device is demonstrated.

A Multistage Procedure of Mobile Vehicle Acoustic Identification for Single-Sensor Embedded Device

Sergei Astapov and Andri Riid

Abstract—Mobile vehicle identification has a wide application field for both civilian and military uses. Vehicle identification may be achieved by incorporating single or multiple sensor solutions and through data fusion. This paper considers a single-sensor multistage hierarchical algorithm of acoustic signal analysis and pattern recognition for the identification of mobile vehicles in an open environment. The algorithm applies several standalone techniques to enable complex decision-making during event identification. Computationally inexpensive procedures are specifically chosen in order to provide real-time operation capability. The algorithm is tested on pre-recorded audio signals of civilian vehicles passing the measurement point and shows promising classification accuracy. Implementation on a specific embedded device is also presented and the capability of real-time operation on this device is demonstrated.

Keywords—vehicle identification, acoustic signal analysis, feature extraction, classification, fuzzy logic

I. INTRODUCTION

MOVING object identification is one of many tasks of environment monitoring systems. It finds its uses in civilian and military applications. The civilian applications of moving motor vehicle identification vary from speed limit control to traffic density analysis and traffic behavior prediction. Military uses involve reconnaissance and identification of friendly over enemy craft [1]. The most important aspect of such monitoring systems is real-time computation and timely result processing as the nature of the problem most often implies time-critical operation. Most state of the art systems typically rely on single sensor ultrasonic, acoustic, video, infrared, radar, microwave, magnetic, laser, vibration based, etc. signal analysis, otherwise they employ combinational multisensory detectors [2], [3]. The main advantage of acoustic [4], [5], [6], [7], [8] and video [9] methods lies in the ease of data signal interpretability, i.e., the acquired data is perceptual without additional manipulations.

Video based methods of vehicle identification are generally more effective and robust in changing weather conditions if provided sufficient visibility and illumination. However, the large amounts of video data and significantly more complex pattern search algorithms, if compared to algorithms for one-dimensional data streams, put significant constraints on the

This research was supported by the Innovative Manufacturing Engineering Systems Competence Centre IMECC, co-financed by European Union Regional Development Fund (project EU30006).

S. Astapov is with the Laboratory for Proactive Technologies, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mail: sergei.astapov@dcc.ttu.ee).

A. Riid is with the Laboratory for Proactive Technologies, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mail: andri@dcc.ttu.ee).

possibilities of real-time system implementation. Acoustic systems on the other hand do not rely on visibility factors, yet are sensitive to background acoustic noise variation. Unlike the classical task of distinguishing the incident signal from uniform ambient noise, the task of vehicle identification lies in distinguishing one type of noise (i.e. vehicle-produced sound) from other noises that occur in the environment. Acoustic noise analysis provides the possibility to distinguish well separable classes of motor vehicles, such as passenger cars from large trucks. The acoustic noise patterns of mobile vehicles consist of multiple components [10]. The harmonic nature of the motor noise is, however, seldom present in the civilian vehicle sound pattern due to the fact that motor sounds are well dampened in modern cars. This fact complemented by the Doppler Effect renders the spectral analysis based on fundamental frequency detection (e.g. [11]) ineffective. Instead, parameters of the spectrum overall shape and energy distribution resembling the vehicle noise patterns may be adopted.

This paper considers different methods of digital audio signal analysis, namely the estimation of spectral energy levels and energy envelope, the analysis of several frequency spectrum instantaneous features and spectral pattern matching. The proposed algorithm possesses a hierarchical structure, beginning with the detection of signal perturbation, continuing with the estimation of noise resemblance to those produced by vehicles, and ending with the classification of the detected vehicle. The algorithm is computationally inexpensive and thus is well implementable on embedded devices. We focus on a single-sensor approach in order to reduce the computational load of the algorithm. However, the procedure can be integrated into a more complex system through data fusion for more sophisticated decision-making.

The paper is organized as follows. In Section II, the applied methods of audio signal analysis and audio feature extraction are reviewed in detail. Section III handles the proposed algorithm's procedures and multistage decision-making. In Section IV several computational simplifications are discussed for optimization. In Section V we use two real test signals for experimental verification of the algorithm's identification accuracy and present intermediate and final results of detection and classification. This section of the paper shows that the algorithm is well applicable to the task of identifying motorized vehicles under varying weather conditions. Additionally, in Section VI we present one option of procedure implementation on specific embedded hardware and demonstrate the real-time operation capability on this specific device.

II. SIGNAL ANALYSIS METHODS

The audio signal is analyzed in the frequency domain. The frequency domain representation of the signal is achieved by applying a temporal signal decomposing operation, namely the Fourier Transform (FT). Frequency features are less affected by noise than temporal features; also most of the temporal features may be approximated in the frequency domain. Furthermore, the frequency spectrum of a temporal signal frame consists of half as many points as there are in the temporal frame, which is relevant in computation complexity critical systems.

A. The Fast Fourier Transform

The discrete temporal signal is decomposed by the Discrete Fourier Transform (DFT). For a finite duration discrete signal $x(m)$ of length N , the DFT function is

$$X(k) = \sum_{m=0}^{N-1} x(m) \cdot e^{-j\frac{2\pi}{N}mk}, \quad k = 0, \dots, N-1. \quad (1)$$

In this manner the transform is performed along two integer dimensions: m and k , i.e. it can be presented as a linear system transformation of complexity $\mathcal{O}(N^2)$. In order to reduce its computation the Fast Fourier Transforms were developed. The proposed system applies a specific implementation of the FFT developed by Frigo and Johnson, called FFTW [12].

The frequency spectrum $[X(0), X(1), \dots, X(N-1)]$ is symmetrically divided into complex conjugate “positive” and “negative” frequencies, the positive ones residing in the interval $[X(0), \dots, X(N/2+1)]$ with $X(0)$ being the signal DC component, which is ignored. In order to obtain the absolute amplitude spectrum, the absolute values of this portion of the spectrum are calculated. Thus, abiding the Nyquist – Shannon sampling theorem, the amplitude frequency spectrum of a signal frame of length N consists of $N/2$ frequency components, each of which is multiple to the frequency resolution given by $\Delta f = F_s/N$, where F_s is the sampling rate.

B. Instantaneous Feature Extraction

In order to acquire the specific signal properties, several features are extracted from the amplitude frequency spectrum [13]. These are referred to as instantaneous features due to the fact that they are extracted from every single spectral frame independently, not relying on previous information. The list of features is signal-specific and is formed during the process of test signal analysis in order to distinguish well separable, desirably weakly correlated features, which best indicate the nature of signal fluctuations corresponding to the concerned events. The six spectral features considered in this paper are extracted from the absolute magnitude spectrum frame $|X_t(k)|$ of length $K = N/2$, $k = 1, \dots, K$.

Root Mean Square (RMS) Energy of the power spectrum conveys the general spectral energy level:

$$X_{RMS} = \sqrt{\frac{1}{K} \sum_{k=1}^K |X_t(k)|^2}. \quad (2)$$

The **band energy** measures the energy of the power spectrum at the i^{th} band and is computed as

$$X_{BE}(i) = \frac{\sum_{l \in S_i} |X_t(l)|^2}{\sum_{k=1}^K |X_t(k)|^2}, \quad (3)$$

where S_i is the set of power spectrum samples belonging to the i^{th} band. The bands are chosen according to the Mel-scale denoted by

$$Mel(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right). \quad (4)$$

The Mel-scale is chosen for its increasing spread towards the higher frequencies, which ultimately means that the bands of lower frequencies where most of the spectral energy resides, are shorter than the bands of low-energy higher frequencies. This allows for better distribution of spectral energy by bands.

The **spectral centroid** represents the first central moment of the magnitude spectrum. It is calculated as the frequency averaged over the absolute magnitude spectrum:

$$X_{SC} = \frac{\sum_{k=1}^K k \cdot |X_t(k)|}{\sum_{k=1}^K |X_t(k)|}. \quad (5)$$

Spectral roll-off measures the frequency below which a certain amount of spectral energy resides. This amount is determined by $TH = [0, 1]$ which is the threshold. For our application we choose it to be equal to $TH = 0.9$ (see Fig. 1).

$$X_{SR} = \arg \max_p \left[\sum_{l=1}^p |X_t(l)|^2 \leq TH \cdot \sum_{k=1}^K |X_t(k)|^2 \right] \quad (6)$$

Spectral slope is a measure of spectral energy decrease in the direction of higher frequencies. It is determined by the gradient and y-intersect parameters of a straight line calculated applying linear regression to the magnitude spectrum frame. Hereby, for a set of data points $(k, |X_t(k)|)$, where $k = 1, \dots, K$, the gradient of the best fitted straight line is denoted as

$$m = \frac{K \sum_{k=1}^K k \cdot |X_t(k)| - \sum_{k=1}^K k \sum_{k=1}^K |X_t(k)|}{K \sum_{k=1}^K k^2 - \left(\sum_{k=1}^K k \right)^2}, \quad (7)$$

and the y-intersect is denoted as

$$c = \frac{\sum_{k=1}^K |X_t(k)| \sum_{k=1}^K k^2 - \sum_{k=1}^K k \sum_{k=1}^K k \cdot |X_t(k)|}{K \sum_{k=1}^K k^2 - \left(\sum_{k=1}^K k \right)^2}. \quad (8)$$

An example of spectral slope for a magnitude spectrum of length $K = 8192$ is presented in Fig. 1. The overall decline of spectral energy towards higher frequencies defines the parameters of the straight line and not the precise energy distribution in bands.

In the proposed algorithm the RMS energy is used independently. The rest of the considered features are concatenated into a feature vector, which is analyzed during the later stages of classification. A combination of features of different nature

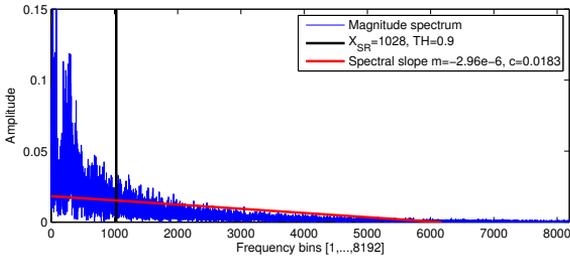


Fig. 1. Spectral roll-off and spectral slope features of an acoustic signal frame.

into a single set may prove harmful in later analysis due to differences of their responsiveness to the incident signal. The method of fuzzy classification, however, overcomes this problem. The issue is further addressed in Section III.

C. Attack Sustain Release Envelope

The process of a vehicle passing the measurement point at a given velocity consists of three stages: approach (spectral energy increases), passing (spectral energy remains stable), retreat (spectral energy decreases). This dynamic pattern is detected by estimating the Attack Sustain Release (ASR) envelope. It is conducted by analyzing the RMS spectral energy (2).

The amount of deviation of RMS energy of the present i^{th} frame $X_{RMS}(i)$ is estimated by the difference between it and the mean value of M previous RMS energy readings. The parameter $\delta \in [0, 1]$ is the lower threshold of energy deviation. RMS energy deviation is coded to three states by the following principle:

$$state_i = \begin{cases} 1, & X_{RMS}(i) > (1 + \delta) \cdot mean_{RMS}(i) \\ -1, & X_{RMS}(i) < (1 - \delta) \cdot mean_{RMS}(i) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where 1 denotes energy increase, 0 denotes stable energy levels and -1 denotes energy decrease. The mean of M previous energy levels is calculated by

$$mean_{RMS}(i) = \frac{1}{M} \sum_{j=i-1}^{i-1} X_{RMS}(j). \quad (10)$$

Therefore the transitions $1 \rightarrow 0 \rightarrow -1$ and $1 \rightarrow -1$ are suspected for a car passing event and the quantities of -1, 0, and 1 coded frames denote the lengths of attack, sustain, and release components, respectively.

III. THE HIERARCHICAL ALGORITHM

The proposed hierarchical algorithm, presented in Fig. 2., consists of two independent stages. The hierarchical decision-making scheme (on the left) firstly differentiates relatively loud sounds from mild background noise, secondly it distinguishes vehicle-produced sounds from heavy background noise and lastly estimates the vehicle type from a set of predefined types. This part of the algorithm operates in a frame-by-frame

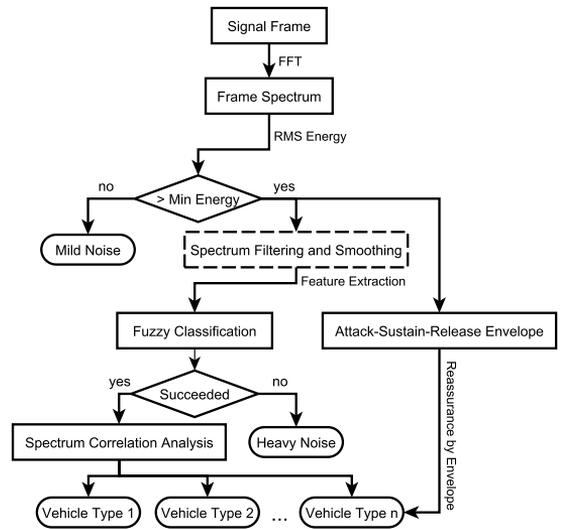


Fig. 2. Block diagram of the proposed hierarchical algorithm for vehicle detection and classification.

manner, computing a single class label per signal frame. The ASR envelope estimation procedure, on the other hand, runs parallel to the decision-making procedure and complements the past frames' classifications with reassurance of positive vehicle-passing event detection. The hierarchy of the algorithm is conditioned by the supremacy of vehicle detection priority over vehicle classification priority, i.e. distinction between vehicle-produced sound and other types of noise is more important than correct vehicle type estimation.

A. Lower Energy Threshold

The first stage of the hierarchical procedure is the estimation of sufficient signal energy. The energy level of a signal frame is calculated and compared to the lower energy threshold, if the threshold is not exceeded, the procedure terminates and the frame is marked as mild noise. The estimation of the lower energy threshold occurs during algorithm parameter estimation by means of test signal analysis. The initial threshold is chosen as the minimal value of RMS energy of all the frames that correspond to vehicle passing instances.

The optional procedure of spectrum filtering and smoothing follows. Digital filtering may improve the Signal to Noise Ratio (SNR) of the spectrum. However, it is effective only in the cases where the spectral band containing the signal is known. In our specific case the vehicle sounds overlap with the background noise and filtering does not improve the classification process. Furthermore, this procedure may corrupt the vehicle acoustic pattern and thus is not applied in our experiments.

B. Fuzzy Classification

The sound pattern of a moving object passing the measuring device is not consistent. Changing signal energy and complex

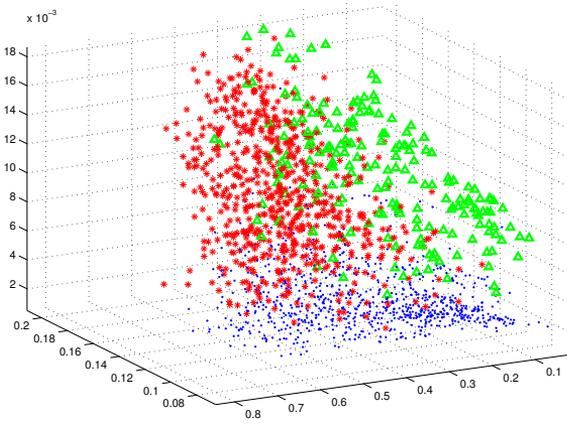


Fig. 3. Three clusters in a three-dimensional feature space. Features represent energy ratios and thus units are not set.

alternations of spectrum shape caused by engine sounds, the Doppler Effect, and also the influence of background noise – all introduce variance to the spectral features of the signal. It is because of this variance why the feature vectors (of length L) corresponding to an event under the same class label form a cluster in an L -dimensional space. Because of the high level of variance among the samples, the clusters can be rather complex shaped (Fig. 3). The separation of samples into a specific class/cluster is carried out by a fuzzy classifier derived by a heuristic training procedure [14]. The classifier operates by applying fuzzy inference to an input feature vector. The algorithm is relatively lightweight if L does not exceed 20-30, which is one of the reasons of applying spectral features instead of the whole spectral vector.

The fuzzy classifier that classifies a feature vector $X = [x_1, \dots, x_L]$ of length L by assigning to it one of T different discrete valued labels, consists of R rules of the following structure:

$$\begin{aligned} &\text{IF } x_1 \text{ is } A_{1r} \text{ AND } x_2 \text{ is } A_{2r} \text{ AND } \dots \text{ AND } x_L \text{ is } A_{Lr} \\ &\quad \text{THEN } y \text{ belongs to class } c_r \end{aligned} \quad (11)$$

where A_{ir} is the linguistic term of the i^{th} input (i.e., feature vector element, $i = 1, \dots, L$) associated with the r^{th} rule and $c_r \in (1, \dots, T)$ is the class label assigned by the r^{th} rule.

The class label is assigned in a winner-takes-all manner by specifying the rule with the highest degree of activation

$$y = c_r, \arg \max_{1 \leq r \leq R} (\tau_r), \quad (12)$$

where τ_r is the activation degree of the r^{th} rule:

$$\tau_r = \prod_{i=1}^L \mu_{ir}(x_i), \quad (13)$$

where μ_{ir} is the MF corresponding to the linguistic term A_{ir} .

Classifier training consists of estimating the parameters of these MFs. For the implementation of the classifier at hand triangle-shaped MFs are used:

$$\mu_{ir}(x_i) = \begin{cases} \frac{x_i - a_{ir}}{b_{ir} - a_{ir}}, & a_{ir} \leq x_i \leq b_{ir} \\ \frac{c_{ir} - x_i}{c_{ir} - b_{ir}}, & b_{ir} < x_i \leq c_{ir} \\ 0, & (x_i < a_{ir}) \vee (x_i > c_{ir}) \end{cases}, \quad (14)$$

where a_{ir} and c_{ir} locate the base of the triangle and b_{ir} locates the peak. The training is performed using a set of reference feature vectors for which a class label is provided manually. This is done during system off-line tuning. The procedure consists of the following steps:

- 1) The set of vectors is partitioned into R subsets S_j , $j = 1, \dots, R$, each consisting of P_j vectors of the same class.
- 2) The parameters of the MFs are calculated as $a_{ir} = \min_{k \in S_j} (x_i(k))$, $c_{ir} = \max_{k \in S_j} (x_i(k))$, $b_{ir} = \frac{1}{P_j} \sum_{k \in S_j} x_i(k)$, $i = 1, \dots, L$.
- 3) The base of each MF is slightly enlarged (by 1% in our case) to give non-zero membership values to the training samples located at the edges of multidimensional space clusters [14].
- 4) The established MFs are added to the classifier rule-base defined by (13).

If the clusters do not separate naturally in the feature space (which is often the case), the extracted rules are bound to have a high degree of overlap (Fig 4, upper subplot). Note that because of (12), the rules compete for the samples and those samples of a class that are at a sufficient distance from the related rule center, will receive a higher activation degree (13) and consequently, the classification decision from a neighbouring rule by what they lose the connection to the rule they were originally assigned to. In such a case, it makes sense to readjust the MF parameters by excluding the lost samples from corresponding S_j and applying the tuning procedure again. Quite often this ignites a minor chain reaction because the updated rules are inclined to lose additional samples to neighbouring rules and we need to readjust them again. In the end, however, what we obtain is a classifier with much more compact rules and MFs (Fig. 4, lower subplot). Moreover, usually this comes at no loss of classification accuracy.

Note, that the classifier that employs triangular MFs cannot operate on samples that fall beyond the rule borders specified by the MF base parameters. This can be fixed, if desired, by replacing the triangular MFs with near-equivalent Gaussian curves defined as

$$\mu_{ir}(x_i) = \begin{cases} \exp \left\{ -\frac{(x_i - b_{ir})^2}{2 \cdot (0.4247 \cdot (b_{ir} - a_{ir})^2)} \right\}, & x_i < b_{ir} \\ \exp \left\{ -\frac{(x_i - b_{ir})^2}{2 \cdot (0.4247 \cdot (c_{ir} - b_{ir})^2)} \right\}, & x_i \geq b_{ir} \end{cases}. \quad (15)$$

While this improves the ability to properly classify the unseen samples, performance of the classifier first and foremost depends on the quality of the training data set. The reference features must be chosen with moderate amounts of background noise. Very noisy reference features will most likely produce large, sparse and heavily overlapping clusters dependent on the stationary properties of this particular noise. On the other

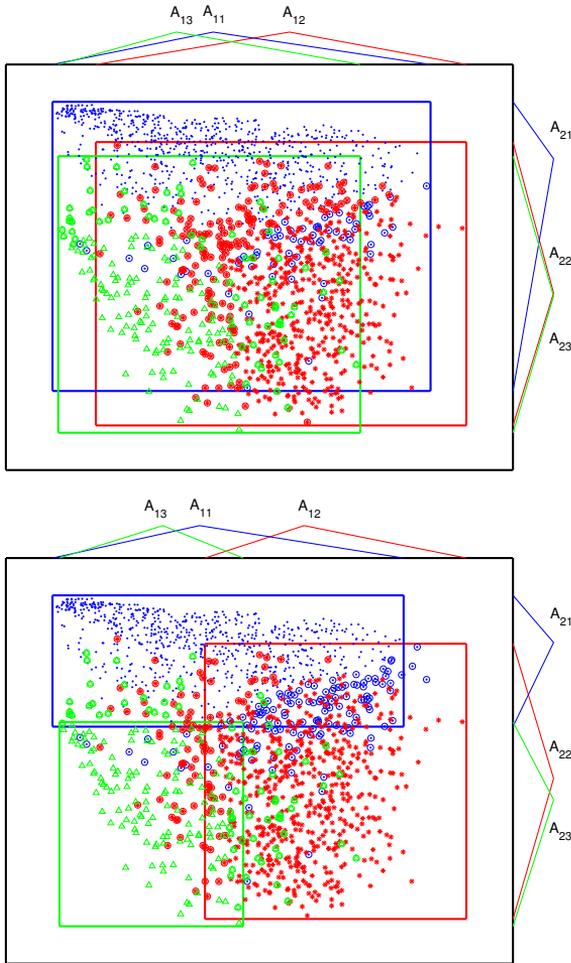


Fig. 4. Initial (upper) and refined (lower) representation of three clusters by triangular MFs.

hand, increasing the size of the reference feature vector will provide more information on the distribution of points in the feature space thus allowing for more efficient MF parameter tuning.

In the proposed hierarchical algorithm, the heuristic fuzzy classification may be applied in two different manners. First, a general cluster corresponding to all vehicle types in question may be estimated and the resulting classifier is used for pure detection purposes, just to distinguish all vehicle produced noise from ambient noise. Final classification is then performed by correlation analysis. Alternatively, a separate cluster for each vehicle class is built and the classifier is applied for specific vehicle type estimation. Here both the fuzzy classifier and the correlation analysis produce separate class estimates, which may reinforce each other. In Section V both methods are used during hierarchical algorithm testing.

C. Correlation Coefficient Analysis

The final stage of vehicle identification is the correlation analysis between the unknown amplitude spectrum vector and the reference spectrum vectors, each corresponding to a single vehicle type class. For a more rigorous classification, several reference vectors per class may also be used. Correlation coefficients are simple and effective metrics for similarity estimation, however, this method is very susceptible to noise. A spectrum of loud background noise may correlate to any of the reference spectra enough to receive incorrect classification. Application of the fuzzy classifier in the previous stage of the algorithm relieves this problem.

During correlation analysis, the correlation coefficients between an unlabeled spectrum vector $x = |X_i(k)|$ and C reference vectors of length K , $r_i = [r_i(1), \dots, r_i(K)]$, $i = 1, \dots, N$, are calculated using the following equation:

$$\rho_i = \left[K \sum_{k=1}^K x(k) \cdot r_i(k) - \sum_{k=1}^K x(k) \sum_{k=1}^K r_i(k) \right] / \left[\sqrt{K \sum_{k=1}^K x(k)^2 - \left(\sum_{k=1}^K x(k) \right)^2} \times \sqrt{K \sum_{k=1}^K r_i(k)^2 - \left(\sum_{k=1}^K r_i(k) \right)^2} \right] \quad (16)$$

The correlation is defined on the interval, $-1 \leq \rho_i \leq 1$, -1 meaning total inverse correlation, 0 specifying uncorrelated patterns, and 1 meaning total direct correlation. The class label corresponding to the reference vector of maximum correlation is declared the winner:

$$y = \arg \max_{1 \leq i \leq C} (\rho_i). \quad (17)$$

D. Reassurance by ASR Envelope

As it was mentioned earlier, the detection of the ASR dynamic of signal energy complements the past identification results. If the ASR pattern is detected, a notification is generated and presented along with the final class estimate. The class labels generated during the period of the detected ASR are inspected for the most frequent one (mode in statistical sense), which is presented in the notification. This reduces inconsistencies in the series of class estimates, e.g. when the vehicle type cannot be clearly classified. Additional restrictions may also be applied to the ASR envelope detection. If the potential velocity of the moving object is known, the lower and upper bounds for the attack, sustain or release components may be specified, so the detection is invalid if these restrictions are not met. For example, if the vehicles are known to stop at the measurement point, the expected values of the sustain component have to be large in order to not confuse this stop with multiple vehicles. On the other hand, for quickly passing vehicles on a highway the ASR components are expected to be short.

IV. ALGORITHM COMPLEXITY MINIMIZATION

The most time consuming operations of the procedure are feature extraction and correlation coefficient calculation

due to a large number of lengthy vector summations. To reduce the number of summations several feature extraction techniques were specifically chosen with similar summands. Analyzing equations (2), (3), (5) – (8), the repeating elements are $\sum_{k=1}^K |X_t(k)|$, $\sum_{k=1}^K |X_t(k)|^2$ and $\sum_{k=1}^K k \cdot |X_t(k)|$, the first two of which are also present in the correlation calculating equation (16). Computing these sums only once and minimizing the number of cycles during feature extraction greatly reduces the number of overall operations.

Equations (7) and (8) may be further simplified if k is taken as an integer vector index of the corresponding frequency component. The closed form for the sum of K first successive integers is equal to

$$\sum_{k=1}^K k = \frac{1}{2}K(K+1), \quad (18)$$

and the sum of squares of K first successive integers is

$$\sum_{k=1}^K k^2 = \frac{1}{6}K(K+1)(2K+1). \quad (19)$$

Even if k is chosen non-integer, the sums of the resulting recurrences may still be evaluated [15], however these closed forms will definitely require more computations than it is needed for (18) and (19).

Calculating the sums of reference vectors and the sums of squared reference vectors only once during the off-line stage of algorithm parameter specification turns (16) to a more lightweight equation with only one specific summation, which must be performed for each correlation coefficient calculation: $\sum_{k=1}^K x(k) \cdot r_i(k)$.

Using a power of two as the signal frame length also reduces computation complexity. FFT operation is optimized for frame lengths multiple to powers of two, also in this case many multiplications and divisions are replaced by simpler and faster bitwise arithmetic shifts.

V. ALGORITHM TESTING RESULTS

The performance of the algorithm is tested on real signals acquired in an open environment. The signals are manually analyzed prior to algorithm accuracy evaluation. This is done in order to estimate the number of classes, that are used in the algorithm and to assign reference class labels to every frame. Each signal is divided into a training and test portion. The partitioning is performed so, that the event of each class occurs at least once in each portion. In our experiments we divide the signals equally, half of the signal is used for training and another half for testing. The training portion is used for fuzzy classifier training and for choosing reference vectors for the correlation analysis procedure. For the fuzzy classifier training, the features are extracted from every frame of the training signal and gathered to the training dataset. For correlation analysis, several spectral vectors corresponding to different classes are manually chosen. The test portion of the signal is used for the estimation of detection and classification accuracy.

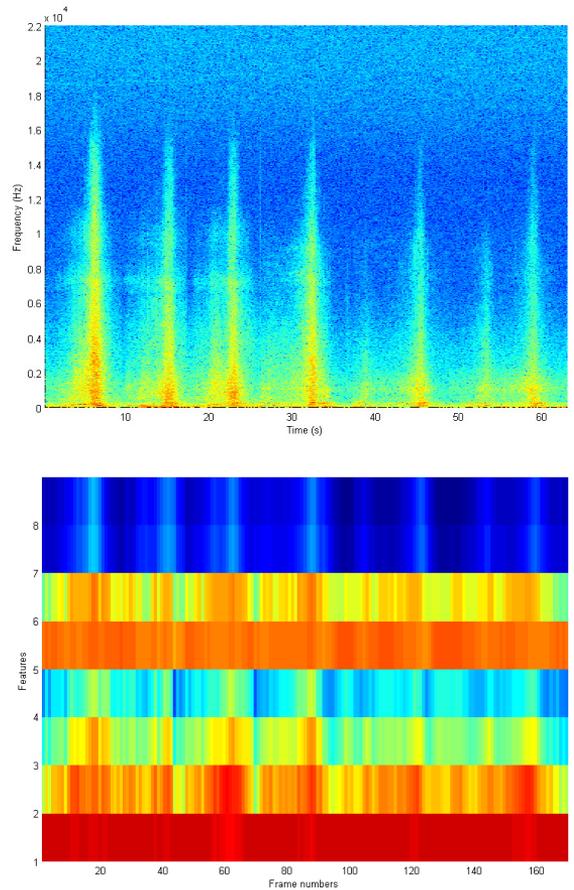


Fig. 5. Spectrogram (upper) and extracted features (lower) of the first experimental signal.

A. First Test Signal

The first test audio signal was measured using a Shure SM58 microphone and a Roland Edirol UA-25EX audio signal processor at 44.1 kHz sampling rate in mono channel mode and saved in 16-bit Waveform Audio File (WAV) format. For the acquisition of the test signal, a microphone was placed at an empty parking lot and two cars (Mercedes S320 and Mazda MX-5) were in turn passing the microphone stand at a speed of 35 – 45 km/h at the passing point, starting to accelerate from a distance of approximately 40 meters. Each car passed the microphone three times: the Mercedes first three times and the Mazda three times afterwards. The sounds were acquired during summer time in mild weather conditions, thus ambient noise levels were relatively low. The signal's spectrogram is presented in Fig. 5. Six instances of passing car sounds are clearly visible.

For testing, the frame length of $2^{14} = 16384$ samples is chosen, which corresponds to 0.3715 seconds at a sampling rate of 44.1 kHz. The signal feature vector comprises of eight features: four band energy features (four bands of 1-824, 824-

2616, 2616-6514, 6514-15000 Hz), spectral centroid, spectral roll-off and spectral slope:

$$X = [X_{BE}(1), \dots, X_{BE}(4), X_{SC}, X_{SR}, m, c]. \quad (20)$$

These features per every signal frame are presented in Fig. 5. Because the SNR is high the features corresponding to car passing events are clearly visually identifiable. For classification a total of 2 class labels are used: 1 for Mercedes and 2 for Mazda. The reference spectral vectors used in correlation analysis are estimated by averaging several spectra of sounds produced by vehicles of the same class, one reference vector per class is applied.

The results of algorithm testing are presented in Fig. 7. The general results are satisfying – every vehicle is detected and successfully classified. As it can be seen in the second and third subplots of Fig. 7, each vehicle passing instance ASR envelope is correctly detected. Though it can be noticed that approximately on the 107th frame the ASR dynamic is falsely detected, the energy of the signal is below the threshold and the fuzzy classifier gives no classification decision, consequently the detection does not occur. Also approximately at the 145th frame the ASR dynamic is present, but is not detected, as for the known vehicle speed corridor the attack and release components of the envelope are set to be no less than 2 frames in duration for the dynamic to be detected. For this signal the fuzzy classifier is trained for detection purposes, i.e. all car types vs. ambient noise. The detection decisions of the fuzzy classifier concur with the ASR envelope.

The heuristic fuzzy algorithm, trained to identify the general vehicle feature space, succeeds in doing so for the majority of signal frames thus allowing the correlation coefficient calculation procedure to analyze only the frames corresponding to vehicle pass time intervals. The fourth subplot of Fig. 7 shows that the correlation coefficient values are unreliable during the periods between vehicle passing instances, during these instances, however, they become more separate, indicating an obvious leader.

B. Second Test Signal

The second test signal was acquired using a condenser microphone Sennheiser KE 4-211-2 and an embedded computing device Gumstix Overo Water. The signal was also sampled at 44.1 kHz mono channel mode and saved to a 16-bit WAV file. Signal acquisition was conducted at a lively two-lane highway during dense traffic in late fall under heavy wind and light rain. Consequently the noise levels in this signal are quite high. The spectrogram, presented in Fig. 6., confirms this. Ambient noise from wind and rain pollutes the whole frequency band, unlike in Fig. 5. Dense traffic results in vehicle passing instances being much less visually separable.

The frame length was chosen the same as for the first test signal. Two vehicle classes were chosen: 1 for passenger cars and 2 for trucks and busses. Feature vectors comprise of eight features, which are the same as for the first signal, except the bands for the band energy features are less spread: 220-818,

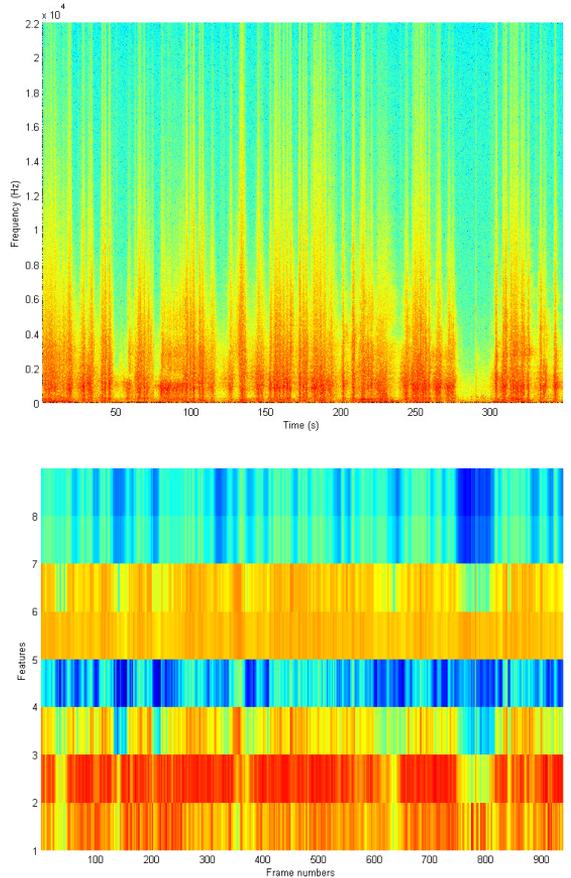


Fig. 6. Spectrogram (upper) and extracted features (lower) of the second experimental signal.

818-2592, 2592-6438, 6438-14780 Hz. Fig. 6. presents these feature vectors. Low SNR makes the features corresponding to vehicle passing instances visually almost unidentifiable. For the derivation of reference spectral vectors, the same technique as for the first signal is used.

The results of signal analysis are presented in Fig. 8. As the time intervals between car passes are very short and often non-existent altogether, reference class labels, which are also used during fuzzy algorithm training, are introduced in the first subplot. The intermediate results are, on the other hand, not presented due to possible problems with readability. The results are as follows: out of 46 instances of class 1 vehicles, 37 were successfully detected and classified, 5 were undetected and 4 were confused with class 2; for 11 instances of class 2 vehicles, 9 were correctly classified, 1 was not detected and 1 confused with class 1. Thus the classification accuracy for class 1 vehicles is 80.43% and for class 2 – 81.82%. Considering the harsh environmental conditions, the overall detection and classification accuracy is admissible.

The main problems causing the lowered classification accuracy are:

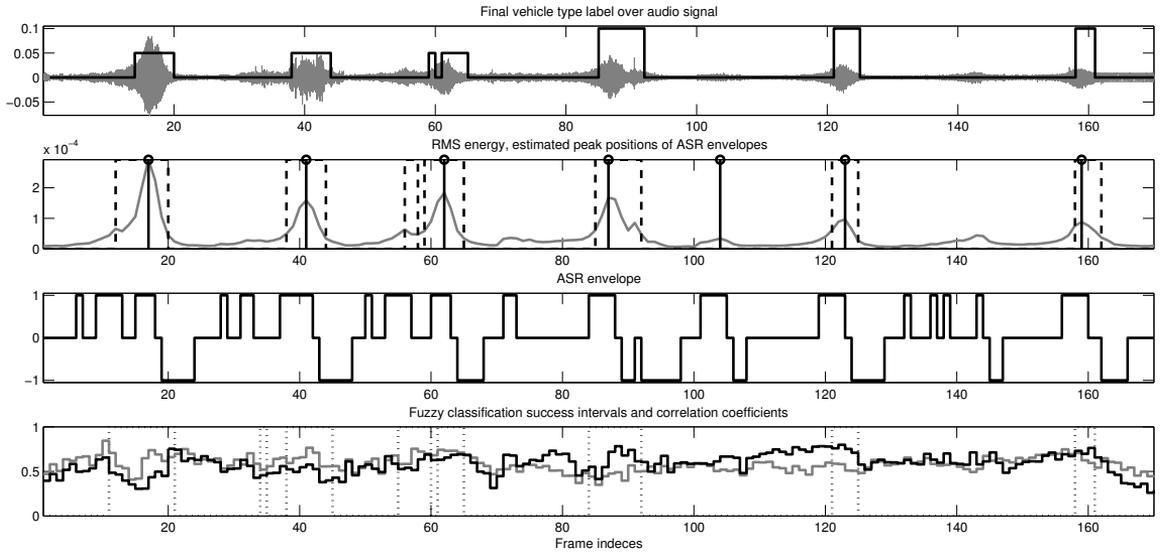


Fig. 7. Algorithm testing results, first test signal. From top to bottom, First Subplot: test signal with 6 instances of passing vehicles (grey); final estimated labels with values 0.05 corresponding to class 1 and 0.1 – to class 2 (black); Second Subplot: RMS energy readings per frame (grey); signal energy threshold (black horizontal line); energy peaks approximated by ASR envelope (black stems); Third Subplot: coded RMS energy dynamic of the ASR envelope; Fourth Subplot: intervals of positive fuzzy membership to vehicle feature subspace (dotted vertical lines); coefficients of correlation to the reference spectral vectors (grey – class 1, black – class 2).

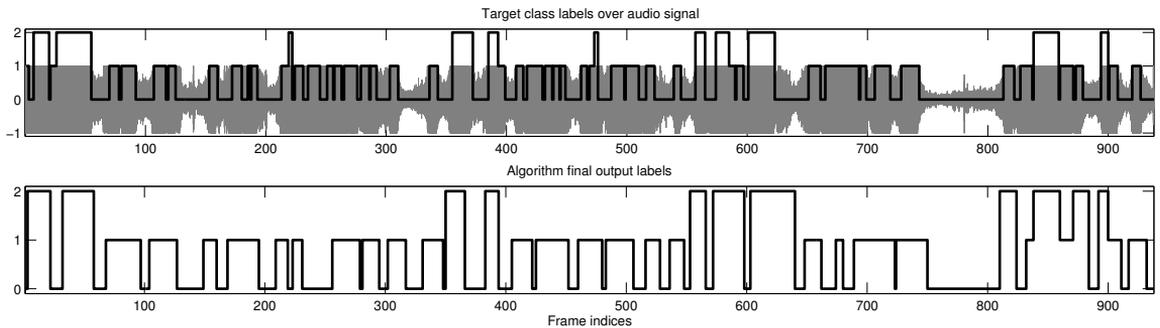


Fig. 8. Algorithm testing results: second test signal. From top to bottom, First Subplot: test signal with instances of passing vehicles (grey), reference labels with values 1 corresponding to class 1 and -1 – to class 2 (black); Second Subplot: final estimated labels with values 1 corresponding to class 1 and -1 – to class 2.

- 1) Severe pollution of the whole signal frequency band with high levels of ambient noise.
- 2) Due to dense traffic the time interval between vehicle passes is very short and often does not allow for distinguishing between successive vehicle passes. Furthermore, sounds of vehicles driving on lanes of opposite direction may overlap and distort one another.
- 3) Sound masking. A heavy truck can emit a noise loud enough to mask the sound of a nearer but lighter car, thus making this car undetectable.
- 4) Intermediate vehicle types (e.g. minibus or pickup truck) make the boundary between passenger and heavy cars more ambiguous. As a result, for some specific types of vehicles precise classification is not possible.

C. General Testing Results

The algorithm operates sufficiently well in both cases of motor vehicles passing with a certain time interval between the passes and heavy traffic. However if the flow of vehicles is consistent and very dense, a decrease of identification quality is witnessed. The influence of background noise, though reduced due to the algorithm's multistage decision-making logic, cannot be eliminated completely. The algorithm is applicable under different weather conditions, which is demonstrated on the examples of both high and low SNR recordings. Possessing a variety of tunable parameters, the sensitivity of the algorithm can be adjusted to the needed extent. This provides the opportunity to apply the algorithm for classification of various types of moving objects not limited to motorized vehicles.

TABLE I
ALGORITHM OPERATION TIMINGS

Processing times (s)	Algorithm sub-procedures				
	1)	2)	3)	5)	Total
Mean	0.0350	0.0022	0.0110	0.0044	0.0526
Maximum	0.0367	0.0027	0.0119	0.0048	0.0564

VI. REAL-TIME OPERATION ON EMBEDDED DEVICE

For the implementation we choose the embedded device Gumstix Overo Water with System-on-Chip OMAP3530 (60 MHz ARM-Cortex-A8), 256MB RAM with a 4GB microSI card. The test signal for the real-time operation experiment is similar to the second test signal considered in Section V-B. The test signal was acquired prior to the experiment. After training the fuzzy rule-base and tuning all the parameters of the algorithm, the signal file is streamed to the device input buffer bypassing the ADC at the rate of the sampling frequency in order to simulate real-time data acquisition and operation [16]. The frame length is chosen, as in the previous sections to be $2^{14} = 16384$, which corresponds to 0.3715 seconds at the sampling rate 44.1 kHz. To operate in real-time, the identification procedure therefore must take less than 0.3715 seconds to compute.

The test signal is 625.27 seconds long, which corresponds to 1683 frames of length 16384. The processing time is measured for the following procedures:

- 1) FFT execution
- 2) RMS energy calculation
- 3) Feature extraction
- 4) Fuzzy Classification
- 5) Calculation of correlation coefficients
- 6) ASR envelope estimation

During the experiment the algorithm is made to run to full extent, not terminating during negative detection, i.e. after 2) or 4), in order to achieve consistent results. The mean and maximal values of processing times are presented in Tab. I. Operations 4) and 6) are excluded from the table, as the times of 4) are either 30-31 or rarely 61 μ s, and for 6) – 30-31 μ s. Thus the mean values do not need to be estimated.

As expected, the most time-consuming operations are FFT execution (consuming more than half of the total processing time) and feature extraction. Process RMS energy calculation takes very little time, so during non-detection the system resources are greatly spared. Correlation computation also consumes much time, growing along with frame length and the number of reference vectors. Thus finding a faster alternative to this method will increase system performance. Altogether, the mean total processing times are significantly smaller than frame durations, thus the algorithm can easily operate in real-time on the given platform.

The distributions of processing times of the computation steps with most variance are presented as histograms in Fig. 9. The variance of processing time is small and thus the predictability of computation time is high. A small amount of values noticeably larger than the mean exist for every sub-procedure and the total process. These abnormalities of long processing durations are most certainly influenced by the

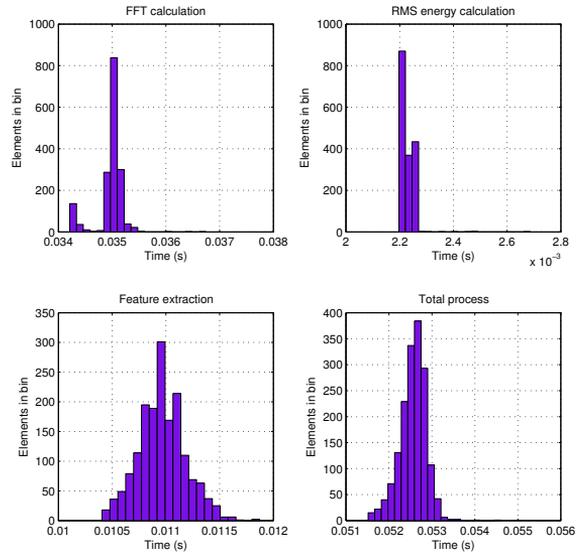


Fig. 9. Histograms of processing times of three sub-procedures with most variance and the whole process.

hardware limitations and problems with memory allocation. Taking this into account and calculating the worst-case total processing time, consisting of maxima of every sub-procedure time, gives 0.0564 seconds for frame length 16384, which is still much less than the signal frame duration.

VII. CONCLUSION

In this paper we have introduced a hierarchical algorithm for mobile vehicle identification by means of acoustic noise analysis. The algorithm is developed specifically for real-time applications and is therefore computationally inexpensive. The hierarchical structure of the algorithm embeds several signal analysis and classification techniques that are applied in a systemized manner to support complex decision-making. The testing results indicate that the algorithm has a potential to detect and classify motor vehicles under varying weather conditions. A possible implementation of the algorithm on an embedded device is presented and its real-time operation capability is experimentally proven.

For future developments the algorithm's robustness may be increased by applying soft discretization to the transitions of the algorithm decision-making path [17] thus transforming its appearance to a fuzzy tree. The final class label therefore may be decided based on degrees of membership. Final class label derivation logic may also be reconsidered to support hierarchy in class label assignment, that will allow for increasing the number of classes by using sub-classes and thus covering the intermediate vehicle types. Another direction of development lies in the integration of other types of sensors in order to enable enhanced environment perception by means of data fusion [3]. On the other hand, application of several microphones in an array configuration would permit to localize the vehicles' positions in the monitored area [18].

REFERENCES

- [1] P. E. William and M. W. Hoffman, "Classification of military ground vehicles using time domain harmonics' amplitudes," vol. 60, no. 11, pp. 3720–3731, 2011.
- [2] E.-H. Ng, S.-L. Tan, and J. G. Guzman, "Road traffic monitoring using a wireless vehicle sensor network," in *Proc. Int. Symp. Intelligent Signal Processing and Communications Systems ISPACS 2008*, 2009, pp. 1–4.
- [3] A. Klausner, A. Teng, and B. Rinner, "Vehicle classification on multi-sensor smart cameras using feature- and decision-fusion," in *Proc. First ACM/IEEE Int. Conf. Distributed Smart Cameras ICDCS '07*, 2007, pp. 67–74.
- [4] T. Takechi, K. Sugimoto, T. Mandono, and H. Sawada, "Automobile identification based on the measurement of car sounds," in *Proc. 30th Annual Conf. of IEEE Industrial Electronics Society IECON 2004*, vol. 2, 2004, pp. 1784–1789.
- [5] A. Starzacher and B. Rinner, "Single sensor acoustic feature extraction for embedded realtime vehicle classification," in *Proc. Int Parallel and Distributed Computing, Applications and Technologies Conf*, 2009, pp. 378–383.
- [6] N. A. Rahim, M. P. Paulraj, A. H. Adom, and S. Sundararaj, "Moving vehicle noise classification using backpropagation algorithm," in *Proc. 6th Int Signal Processing and Its Applications (CSPA) Colloquium*, 2010, pp. 1–6.
- [7] S. Maithani and R. Tyagi, "Noise characterization and classification for background estimation," in *Proc. Int. Conf. Signal Processing, Communications and Networking ICSCN '08*, 2008, pp. 208–213.
- [8] S. S. Yang, Y. G. Kim, and H. Choi, "Vehicle identification using wireless sensor networks," in *Proc. IEEE SoutheastCon*, 2007, pp. 41–46.
- [9] G. Grötsch, N. Donath, B. Kohn, and M. Litzenberger, "Night-time vehicle classification with an embedded, vision system," in *Proc. 12th Int. IEEE Conf. Intelligent Transportation Systems ITSC '09*, 2009, pp. 1–6.
- [10] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave patterns," *Trans. Sig. Proc.*, vol. 57, no. 1, pp. 30–47, Jan. 2009.
- [11] M. Zivanovic, A. Röbel, and X. Rodet, "Adaptive threshold determination for spectral peak classification," *Comput. Music J.*, vol. 32, no. 2, pp. 57–67, Jun. 2008.
- [12] M. Frigo and S. G. Johnson, "FFTW: an adaptive software architecture for the FFT," in *Proc. IEEE Int Acoustics, Speech and Signal Processing Conf*, vol. 3, 1998, pp. 1381–1384.
- [13] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," CUIDADO I.S.T. Project Report, Tech. Rep., 2004.
- [14] A. Riid and E. Rustern, "An integrated approach for the identification of compact, interpretable and accurate fuzzy rule-based classifiers from data," in *Proc. 15th IEEE Int Intelligent Engineering Systems (INES) Conf*, 2011, pp. 101–107.
- [15] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics: a foundation for computer science*. Addison-Wesley Reading, MA, 1994, vol. 2.
- [16] S. Astapov, J. S. Preden, and E. Suurjaak, "A method of real-time mobile vehicle identification by means of acoustic noise analysis implemented on an embedded device," in *Proc. 13th Biennial Baltic Electronics Conf. (BEC)*, 2012, pp. 283–286.
- [17] Y. Peng and P. Flach, "Soft discretization to enhance the continuous decision tree induction," in *Integrating Aspects of Data Mining, Decision Support and Meta-Learning*, C. Giraud-Carrier, N. Lavrac, and S. Moyle, Eds. ECML/PKDD'01 workshop notes, September 2001, pp. 109–118.
- [18] V. C. Ravindra, Y. Bar-Shalom, and T. Damarla, "Feature-aided localization of ground vehicles using passive acoustic sensor arrays," in *Proc. 12th Int. Conf. Information Fusion FUSION '09*, 2009, pp. 70–77.

Publication 2

Reference

S. Astapov, J. Berdnikova, and J.-S. Preden, “Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.

Abstract

Acoustic localization by means of sensor arrays has a variety of applications, from conference telephony to environment monitoring. Many of these tasks are appealing for implementation on embedded systems, however large dataflows and computational complexity of multi-channel signal processing impede the development of such systems. This paper proposes a method of acoustic localization targeted for distributed systems, such as Wireless Sensor Networks (WSN). The method builds on an optimized localization algorithm of Steered Response Power with Phase Transform (SRP-PHAT) and simplifies it further by reducing the initial search region, in which the sound source is contained. The sensor array is partitioned into sub-blocks, which may be implemented as independent nodes of WSN. For the region reduction two approaches are handled. One is based on Direction of Arrival estimation and the other — on multilateration. Both approaches are tested on real signals for speaker localization and industrial machinery monitoring applications. Experiment results indicate the method’s potency in both these tasks.

Optimized Acoustic Localization with SRP-PHAT for Monitoring in Distributed Sensor Networks

Sergei Astapov, Julia Berdnikova, and Jürjo-Sören Preden

Abstract—Acoustic localization by means of sensor arrays has a variety of applications, from conference telephony to environment monitoring. Many of these tasks are appealing for implementation on embedded systems, however large dataflows and computational complexity of multi-channel signal processing impede the development of such systems. This paper proposes a method of acoustic localization targeted for distributed systems, such as Wireless Sensor Networks (WSN). The method builds on an optimized localization algorithm of Steered Response Power with Phase Transform (SRP-PHAT) and simplifies it further by reducing the initial search region, in which the sound source is contained. The sensor array is partitioned into sub-blocks, which may be implemented as independent nodes of WSN. For the region reduction two approaches are handled. One is based on Direction of Arrival estimation and the other – on multilateration. Both approaches are tested on real signals for speaker localization and industrial machinery monitoring applications. Experiment results indicate the method's potency in both these tasks.

Keywords—Acoustic localization, wireless sensor networks, direction of arrival, multilateration, SRP-PHAT

I. INTRODUCTION

IN recent years acoustic signal analysis has grown in popularity in environment monitoring applications. Acoustic signal analysis has a wide area of application because one-dimensional audio signals are relatively easy to process, they are highly comprehensive without additional manipulations, and because acoustic signal acquisition does not require either full direct sight of view of the monitored object, or sufficient highlighting. On the other hand, acoustic signals are prone to noise pollution, especially in unconfined environments, where ambient noise variance and the nature of different background noises are undetermined. For single-sensor solutions, noise poses a great problem because these solutions are unable to efficiently filter unknown noise types [1]. However, the situation changes radically if the acoustic sensors are used in array configurations. Multi-sensor solutions enable concentrating on a specific region of the monitored area and consequently filtering the sound incoming from that region

This research was supported by the Innovative Manufacturing Engineering Systems Competence Centre IMECC, by the European Union Regional Development Fund (project EU30006), co-financed by the Estonian ICT doctoral school and the Nations Support Program for ICT in Higher Education "Tiger University".

S. Astapov and J.-S. Preden are with the Laboratory for Proactive Technologies, Department of Computer Control, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mails: sergei.astapov@dcc.ttu.ee; jurgo.preden@dcc.ttu.ee).

J. Berdnikova is with the Department of Radio and Communication Engineering, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mail: juliad@lr.ttu.ee).

alone. Another application of sensor array solutions lies in sound source localization and tracking.

There exists a variety of methods for acoustic localization, e.g. [2]–[5]. These methods are all based on simple principles of acoustic wave propagation. Having several sensors set in a specific configuration, the direction and distance to the sound source can be estimated by the time delays of wave arrival to these sensors, also called Time Difference of Arrival (TDOA). The Direction of Arrival (DOA) can be estimated from the TDOA or by other methods, as for example in the MUSIC algorithm [4]. For our application we employ the localization method of Steered Response Power with Phase Transform (SRP-PHAT). The method is established to be robust and tolerant to both noise and acoustic reverberation.

One of the main problems related to acoustic localization methods is high computational complexity. Multi-channel signal processing requires large amounts of computational resources for real-time operation. The significantly reduced resources of embedded hardware of Wireless Sensor Networks (WSN) aggravate the situation. Furthermore, for WSN the amounts of data exchanged between nodes must also be maximally reduced. For these reasons the main focus of research in the area lies in the simplification of localization algorithms. Yet, WSN applications with small embedded hardware solutions allow to widen the ordinary localization techniques with more complex multi-node sound source detection and recognition solutions, e.g. [6]–[9].

In this paper we propose a method of Initial Search Region Reduction (ISRR) for the SRP-PHAT, that significantly reduces computational load. For the implementation we use several linear microphone arrays, that together constitute a large-aperture array with a wide area of observation. The ISRR is performed by estimating the DOA for every sub-array and finding the region of common direction. Alternatively we also use multilateration for the region estimation. For final localization we apply the optimized version of SRP-PHAT, which uses Stochastic Region Contraction (SRC) for global energy maximum search. The proposed method is tested on real signals for moving speaker localization and industrial machinery monitoring [10] applications. Based on the results, we consider the advantages and shortcomings of the DOA and multilateration approaches to ISRR.

II. ACOUSTIC LOCALIZATION WITH SRP-PHAT

Acoustic localization may be performed either in a three or two-dimensional space. For our grounded applications we focus on the horizontal plane, thus acoustic source coordinates

(x, y) are estimated. In the two-dimensional space the use of linear arrays is sufficient and computationally less complicated. Linear arrays consist of several microphones with equal distances between each other. The TDOA from one microphone to another then specifies the DOA of the source. The calculation of DOA relies firstly on the speed of sound (in air in our case), the dependence of which on the ambient temperature is expressed by the following equation:

$$c = 331.45\sqrt{1 + \theta/273}, \quad (1)$$

where c is the speed of sound and θ is the temperature in Celsius. Secondly, DOA depends on the assumption of near or far field signal source location. For our implementation we assume the far field disposition of the sound source. The near and far field assumptions specify the trigonometry to be used for DOA computation. Sound waves propagate spherically, and while in the near field this curvature of the wave front is accounted for in DOA calculation, in the far field the fronts are well spread and considered linear. We combine several linear array blocks to achieve a large-aperture array with a Field of View (FOV) of up to 25 m². A FOV is an area where the sound source is localizable, it directly depends on the array's configuration. Large FOV require much time and resources for the source to be localized.

A. Conventional SRP-PHAT

SRP-PHAT is a technique of estimating the DOA of sound signals in a reverberant environment. The SRP $P(\vec{a})$ is a real-valued functional of a spatial vector \vec{a} , which is defined by the FOV of a specific array. The high maxima in $P(\vec{a})$ indicate the estimates of the sound source location. $P(\vec{a})$ is computed for each direction as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) value across all pairs of microphones at the theoretical time-delays associated with the chosen direction. Consider a pair of signals $x_k(t)$, $x_l(t)$ of an array consisting of M microphones. The time instances of sound arrival from a point $a \in \vec{a}$ for the two microphones are $\tau(a, k)$ and $\tau(a, l)$ respectively. Hence the time delay between the signals is $\tau_{kl}(a) = \tau(a, k) - \tau(a, l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega(\tau(a,k) - \tau(a,l))} d\omega, \quad (2)$$

where $X(\omega)$ is the spectrum (the Fourier transform) of signal x and $X^*(\omega)$ is the conjugate of the spectrum [11]. Ψ_{kl} is the PHAT weight of the inverse of the spectral magnitudes:

$$\Psi_{kl} = \frac{1}{|X_k(\omega) X_l^*(\omega)|}. \quad (3)$$

The PHAT is an effective weighting of a GCC for finding TDOA from signals in a highly-reverberant environment.

Computing the SRP for every point in the area \vec{a} results in a SRP image of the whole observable FOV. These images are highly suitable for manual analysis as they portray signal energy distributions and reverberation effects very clearly. For example, consider a result of speaker localization in a room

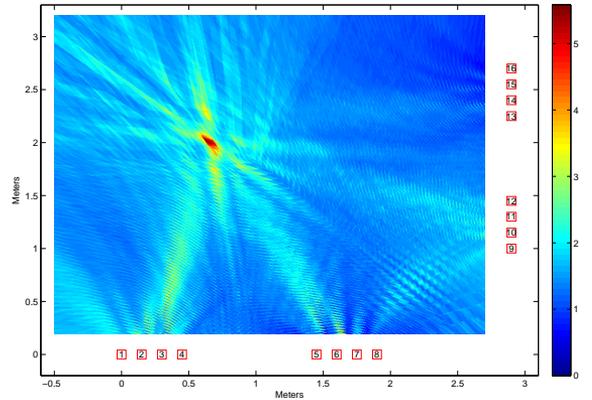


Fig. 1. SRP image of speaker localization using the conventional SRP-PHAT.

performed by conventional SRP-PHAT, presented in Fig. 1. However, for automated processing (i.e. global maximum estimation) these images contain an overwhelming amount of information. Consequently, the processing, as the image generation itself, is highly time and resource consuming. Several propositions have been made for SRP optimization [11]–[13]. For our work we choose the method of locating high maxima of SRP energy by applying Stochastic Region Contraction (SRC).

B. SRP-PHAT with SRC

The conventional SRP-PHAT performs as many functional evaluations (2), or FE, as there are points in \vec{a} , the number of which is defined by the dimensionality of the FOV and the accuracy measure, that partitions the area into small discrete regions. This analysis is highly resource demanding, particularly when applied for large areas of observation. The number of computations is significantly reduced by applying Stochastic Region Contraction, which iteratively reduces the search volume for the global maximum. SRC starts with the initial search volume (i.e. the whole FOV), stochastically explores the functional of that volume by randomly picking a specific number of points, and then contracts the volume into the sub-volume containing the desired global optimum and proceeds iteratively until the global maximum can be located with a finite precision [11]. The procedure may be described in pseudo code as:

- 1) Initialize iteration $i = 0$.
- 2) Set initial parameters: $V_0 = V_{\text{room}}$ – initial volume; J_0 – the number of random points that need to be evaluated to ensure, that one or more is likely to reside in the sub-volume of higher values, surrounding the global maximum; N_0 – number of points used to define the new sub-volume V_{i+1} .
- 3) Calculate $P(\vec{a})$ for J_i points.
- 4) Sort out the best (highest) $N_i \ll J_i$ points.
- 5) Contract the search volume to the smaller volume V_{i+1} , defined by a rectangular boundary vector $B_{i+1} = [x_{\min}(i+1), x_{\max}(i+1), y_{\min}(i+1), y_{\max}(i+1), z_{\min}(i+1), z_{\max}(i+1)]$, that contains these N_i points.

- 6) **IF** $V_{i+1} < V_u$ (a sufficiently small sub-volume, in which the global optimum is contained) **AND** $FE_i < \Phi$ (the total number of FE-s for iteration i is less than the maximum number of allowed FE-s), **THEN** determine the global maximum, **STOP**.
- 7) **ELSE IF** $FE_i \geq \Phi$, **STOP**, discard results.
- 8) **ELSE** Among the N_i points keep a subset G_i of points, which have values greater than the mean μ_i of the N_i points.
- 9) Evaluate J_{i+1} new random points in V_{i+1} .
- 10) Form the set of N_{i+1} as the union of G_i and the best $N_{i+1} - G_i$ points from the J_{i+1} just evaluated. This gives N_{i+1} high points for iteration $i + 1$.
- 11) $i = i + 1$, **GO TO** Step 5.

There are several proposed ways of selecting N_i and J_i depending on the specific FOV and on the condition of monotonic or non-monotonic increase of the mean μ_i . The one, emphasized in [11], consists of fixing N_i and adjusting J_i incrementally in the following fashion: N_i is chosen as $N_i \equiv N = 100$; J_i is the number of FE-s to find $N - G_i$ points greater than μ_i . For our system we propose a different method, which is presented in Section III-C.

III. INITIAL SEARCH REGION REDUCTION

However, optimized localization algorithms still require a significant amount of resources while starting the evaluation on the initial search area. Furthermore, the convergence on a sharp maximum may be guaranteed only if it exists in the FOV. For many applications and monitored objects this is not always true. Large objects, like vehicles or other machinery, do not have a single point of sound emission, rather they appear as distributed regions of heightened acoustic energy with several maxima. On the other hand, if no sound source is present, the localization algorithms will search for maxima in ambient noise, which produces useless results while consuming resources. The reduction of the initial search area, firstly, allows to estimate the presence of the sound source in the FOV, and secondly, greatly reduces the computational load of localization.

We focus on an array setup targeted for use in WSN. The sub-array blocks are placed in different positions in the environment, their orientation may be at all random. The position of the sub-array is specified by the coordinates of its center (which may be found using [14] or [15]) and the angle α , by which the array is steered from the global zero angle, as it is shown in Fig. 2. Knowing the coordinates of a block center (x_0, y_0) , i th sensor before rotation (x_i, y_i) and the angle α , the steering is performed as

$$\begin{bmatrix} x_i^{(rot)} \\ y_i^{(rot)} \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x_i - x_0 \\ y_i - y_0 \end{bmatrix}. \quad (4)$$

Such a configuration is convenient for WSN, where each sub-block may be implemented on a separate network node. Sub-arrays with common FOV form large-aperture arrays and cooperate on localization. Such a configuration enables ad-hoc array composition and increases robustness due to high decentralization. Also a large number of sub-blocks simplifies

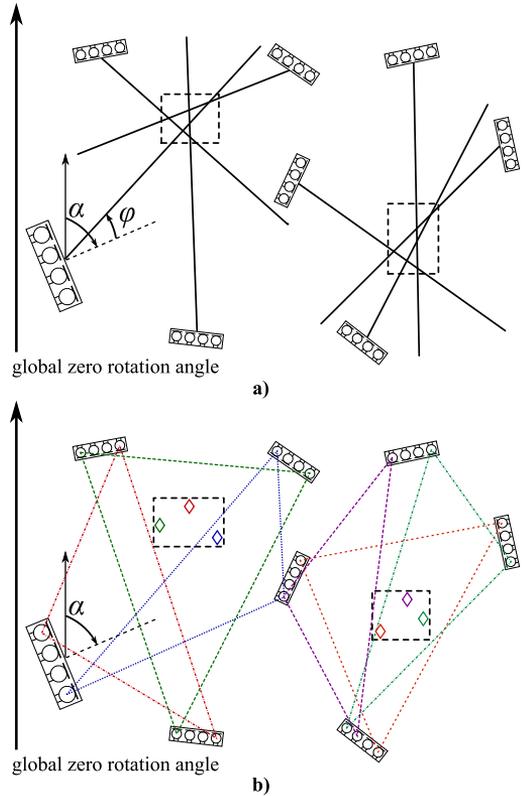


Fig. 2. Initial search region estimation by (a) DOA calculation and (b) multilateration in random configuration of sensor array blocks.

multiple source localization, as the monitored area is divided into smaller local regions.

The ISRR is performed by estimating the DOA for every sub-array and finding the region of common DOA (i.e. the intersection of vectors pointed by the DOA) as is shown in Fig. 2a. We also consider an alternative approach of choosing sensor triplets and performing multilateration to retrieve the source coordinate estimates. The aggregate of these estimates then denotes the sought-for region (Fig. 2b).

A. The DOA Approach to ISRR

Having K microphone arrays, each consisting of M microphones, observing a common FOV, the ISRR is performed in the following steps:

- 1) Estimate the DOA for each of K arrays.
- 2) Generate vectors spanning from the arrays' centers to the bounds of the FOV in the directions of DOA.
- 3) Find points of intersections of these vectors.
- 4) Find groups of points no farther than D_{max} distance units (meters) from their centroid and enclose the areas, in which these groups coincide, in rectangles.
- 5) Perform control of false detection, discard areas not meeting specific criteria (optional).

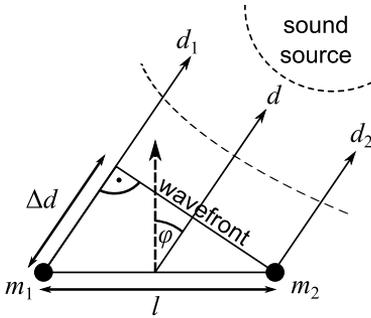


Fig. 3. DOA estimation for a pair of microphones.

The DOA are estimated for the array front, i.e. from -90° to 90° . As we operate in the horizontal plane, it is sufficient to acquire the azimuth (angle of arrival) of the incoming signal to estimate the DOA [16]. The estimation is performed for all $\binom{M}{2}$ combinations of M microphone pairs. Considering Fig. 3, the sound wave originating from a source in the far field is acquired by the microphones m_1 and m_2 with a time delay $\tau \in [-\tau_{\max}, \tau_{\max}]$, where τ_{\max} is the delay of sound traveling directly from one microphone to the other (i.e. at $\pm 90^\circ$). To estimate τ we apply cross-correlation to the two signals:

$$R(\tau) = \sum_{k=0}^n m_1(k) \cdot m_2(k - \tau), \quad (5)$$

where n is the length of the signals in samples. The maximum of the cross-correlation defines the time delay, and the azimuth is obtained by

$$\varphi = \arcsin \frac{\tau \cdot c}{l} = \arcsin \frac{\Delta k / f_s \cdot c}{l}, \quad (6)$$

where c is the speed of sound, l is the distance between two microphones and τ is represented in terms of delay in samples Δk and the sampling frequency f_s . Depending on the chosen pair of microphones in the array, l will vary from l to $l(M-1)$. At this point data validation may be performed. If the maximum of correlation is less than some threshold, the azimuth φ may be discarded. This way, in absence of a sound source or in case of high noise, invalid estimates are avoided early on. We use the deviation from the mean for this metric:

$$\max(R(\tau)) > (1 + TH) \cdot \overline{R(\tau)}, \quad (7)$$

where TH is the threshold of deviation, which depends on the Signal to Noise Ratio (SNR). We use 0.2-0.3 in our experiments.

Having $C_i \leq \binom{M}{2}$ azimuth estimates for every array (varying slightly due to varying inter-microphone distance and accounting for the far field error), the final DOA for each i th array, $i \in (1, \dots, K)$, is derived according to the following special cases:

- 1) DOA spread uniformly (leftmost pairs point left, rightmost – right, and center – straight); no common DOA, $\phi_i = \emptyset$.

- 2) DOA are consensual with slight variance: common DOA is the mean of pair-wise ones

$$\phi_i = \frac{1}{C_i} \sum_{j=1}^{C_i} \varphi_{i,j}. \quad (8)$$

- 3) Same as Case 2), but with some DOA outside variance of consensual group: exclude these DOA from mean.
- 4) Several distinct groups of consensual DOA: choose one with more members and less variance (several may be considered for heavily multi-source applications), calculate mean.

Having estimated $K_1 \leq K$ of the existing array DOA ϕ_{i^*} , $i^* \in (1, \dots, K_1)$ and added the nodes' rotation angles α_i to them, vectors $\overrightarrow{AB}_{i^*}$ are computed with the starting point $A_{i^*} = (x_{1,i^*}, y_{1,i^*})$ being the coordinate of i^* th array's center and the ending point $B_{i^*} = (x_{2,i^*}, y_{2,i^*})$ being the point at a bound of the FOV steered by ϕ_{i^*} from the array's center. Intersection points of all pairs $\overrightarrow{AB}_h, \overrightarrow{AB}_k$ are calculated by

$$I_{hk} = (I_x, I_y) = \left(\frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(x_{1,k} - x_{2,k}) - (x_{1,h} - x_{2,h})(x_{1,k}y_{2,k} - y_{1,k}x_{2,k})}{(x_{1,h} - x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k} - x_{2,k})}, \frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k}y_{2,k} - y_{1,k}x_{2,k})}{(x_{1,h} - x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k} - x_{2,k})} \right). \quad (9)$$

As a result we have a set of $I_{i^{**}}$ intersections, $i^{**} \in (1, \dots, K_2)$, $K_2 \leq \binom{M}{2}$. To get the initial search areas, these intersection points are partitioned by their relative distance. For the maximum distance D_{\max} the partitioning is performed in the following manner:

- 1) **IF** no points $I = \emptyset$ **THEN** no partitions $P = \emptyset$, **STOP**
- 2) **ELSE IF** only 1 point I_1 **THEN** $P_1 = I_1$ **STOP**
- 3) **ELSE** number of partitions $j = 0$
- 4) **WHILE** $|I| > 0$, where $|I|$ is the cardinality of the set I , calculate centroid of free points $\text{Cent} = 1/|I| \cdot \sum I$.
- 5) Calculate Euclidean distance of all free points to centroid $D_k = \sqrt{\sum_{s=1,2} (I_{k,s} - \text{Cent}_s)^2}$, choose point with minimal distance, $j = j + 1$, insert point to P_j , remove point from set of free points I .
- 6) Calculate partition centroid $\text{Cent}(P_j) = 1/|P_j| \cdot \sum P_j$, get Euclidean distance for all free points $D_k = \sqrt{\sum_{s=1,2} (I_{k,s} - \text{Cent}(P_j)_s)^2}$.
- 7) **IF** $\min(D) \leq D_{\max}$ **THEN** insert point corresponding to $\min(D)$ into P_j , delete point from set of free points I , **GO TO** Step 6.
- 8) **ELSE IF** $|I| > 1$ **THEN GO TO** Step 4.
- 9) **ELSE** $j = j + 1$, put last remaining point to P_j .

After obtaining the partitions, their areas are enclosed by rectangles with the edges denoted by the partitions' minimal and maximal values of x and y , added a constant in order to ensure minimal area (in the experimental part we choose 0.1 m). As a result several initial regions may occur in the same FOV. Also while a vector of DOA from one array may cross with several other vectors, redundant "echoing" regions may arise. These may be removed by additional control metrics or by analyzing previously estimated positions (i.e. tracking).

B. Multilateration Approach to ISRR

Multilateration is a technique of estimating the signal source coordinates based on TDOA from the source to the receiving sensors. The distance between the sensor with coordinates x_i, y_i, z_i and the acoustic object could be defined as the length of vector \vec{d} :

$$\|\vec{d}\| = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}, \quad (10)$$

where x, y, z are the acoustic source coordinates. For the multi-sensor WSN ground applications we simplify the solution with constant z dimension. Thus having a TDOA τ_{ij} between two nodes i and j , the acoustic source location coordinates are calculated directly by

$$d_{ij} = c \cdot \tau_{ij} = c (\tau_i - \tau_j) = \frac{\sqrt{(x_i - x)^2 + (y_i - y)^2} - \sqrt{(x_j - x)^2 + (y_j - y)^2}}{(11)}$$

where d_{ij} is the distance difference estimate between sensors i and j , and (x_i, y_i) and (x_j, y_j) are the sensors' respective coordinates [16]. If τ_{ij} is represented in terms of delay in samples Δk_{ij} with sampling frequency f_s , then the difference, similar to (6), is computed as $d_{ij} = \Delta k_{ij} / f_s \cdot c$. The delay τ_{ij} is calculated using cross-correlation, as in (5), also applying the control metric (7). For any three separate sensors (1, 2, 3) acoustic source is localizable by the following system of equations:

$$\begin{cases} d_{12} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} - \sqrt{(x_2 - x)^2 + (y_2 - y)^2} \\ d_{13} = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} - \sqrt{(x_3 - x)^2 + (y_3 - y)^2} \end{cases} \quad (12)$$

To estimate the solution to this system of nonlinear equations we apply a numerical method called Trust-Region Dogleg [17].

We use multiple sensor triplets in order to establish several triangles for multilateration. Every triplet gives a separate position estimate and then all the estimates are partitioned by minimal distance, as in Section III-A, in order to get the reduced regions. The general direct multilateration solution in real-time WSN applications is solved with larger number of nodes [18], where the incorrectly placed regions or multiple sound sources are eliminated by feedback from the object tracking stage. We, however, do not expand beyond three sensor batches in order to simplify and accelerate the solution estimation procedure.

C. Application of SRP-PHAT with SRC to ISRR Estimates

Our approach initializes the SRP-PHAT on already contracted areas and often more than once for a single signal frame. The typical approach to SRC suggests choosing fixed values for $N_i \equiv N = 100$ and $J_0 = 3000$ for a FOV of approximately 20 m^2 , however this is not suitable for constantly varying initial search areas. In our approach the parameters are rather estimated by linear functions. Building on the test results in [11], considering peak estimation quality, and performing our own testing, we derive the two functions

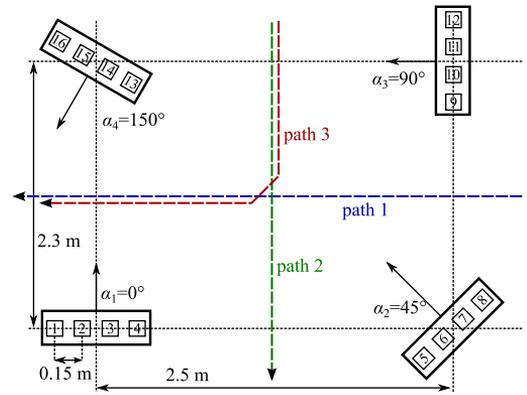


Fig. 4. Layout of the experiment with one speaker and four array blocks.

for the task:

$$J_0(s) = \begin{cases} [297.6 \cdot s + 24], & S < 10 \\ 3000, & S \geq 10 \end{cases}, \quad (13)$$

$$N(s) = \begin{cases} [9.9 \cdot s + 1], & S < 10 \\ 100, & S \geq 10 \end{cases},$$

where s is the area of the FOV in m^2 and $[\cdot]$ denotes the operation of rounding to the nearest integer. The application of these functions optimizes the SRC process by greatly reducing the number of SRP evaluations for reduced regions of acoustic source search [19].

IV. EXPERIMENTAL RESULTS

For the experimental installation we use Vansonic PVM-6052 condenser microphones. The microphones are mounted with a spacing of 15 cm between each other. We use 4 sub-arrays with 4 microphones in each sub-array (width of a single sub-array is thus equal to 45 cm), which results in a large aperture 16-microphone array. For signal acquisition an Agilent U2354A data acquisition device (DAQ) is used with the sampling rate set to 8 kS/s per channel. The data is acquired to and processed in the Matlab environment using the Data Acquisition Toolbox. Processing is performed in frames with a step of 0.2 seconds by conventional SRP-PHAT and by ISRR followed by SRP-PHAT SRC on estimated regions.

A. Human Speaker Localization

For the human speaker localization experiment the microphones are placed in a room as it is portrayed in Fig. 4. The FOV is set to be 1 meter wider in every direction than the corner points of the array (approximately 18 m^2). Sub-arrays SA1, SA2 and SA3 form a triangular array, while sub-array SA4 is diverted from the common direction of view, simulating the belonging to a different group. The speaker takes 3 paths while walking with an average pace (approximately 1-1.5 m/s) and reciting the rainbow passage (designed to contain all the English phonemes and used in speech evaluations). For the DOA approach to ISRR all 16 microphones are used, for

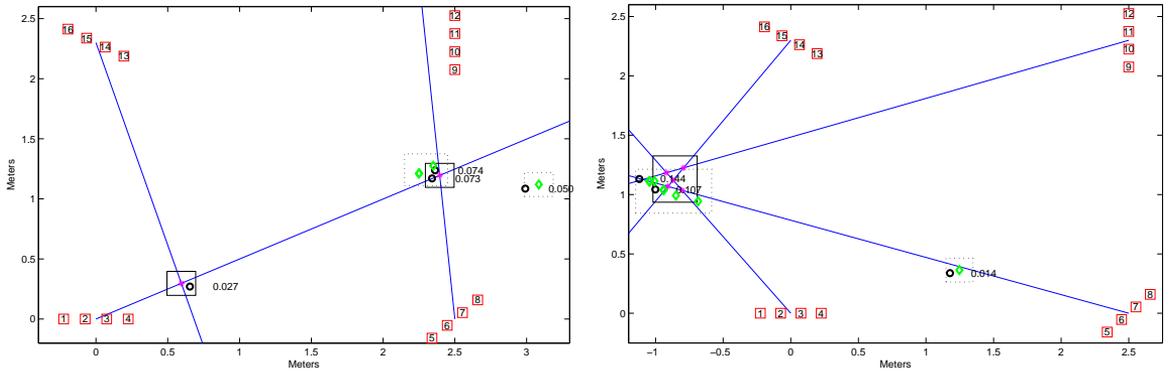


Fig. 5. Two instances of ISRR and localization for the speaker experiment. Blue lines denote sub-array DOA, pink stars – the intersections of DOA rays, black boxes – the estimated regions and black circles - the energy maximum of a region followed to the right by its value. For triangulation the coordinate estimates are denoted by green diamonds, the regions – by dotted boxes and energy maxima – by black circles.

TABLE I
RESULTS OF REGION REDUCTION FOR SPEAKER LOCALIZATION

FOV $\approx 18 \text{ m}^2$	DOA estimation	Multilateration
Mean area (m ²)	0.1374	0.0621
RMSE x (m)	0.1143	0.1227
RMSE y (m)	0.1107	0.1230

multilateration the triplets are chosen in the following order: 1 4 12; 1 4 16; 5 8 12; 5 8 16; 1 8 12; 1 8 16; 4 5 9; 4 5 13. Several resulting triangles with small areas may perform better on closer distances to the source and those with larger areas – on greater distances.

In this experiment the ISRR with DOA estimation and multilateration operate with approximately equal accuracy. Problems arise for both approaches in the region behind and between SA2 and SA3 (path 1), where neither SA2 or SA3 have a sufficient view of the source and SA1 and SA4 are overly steered away (Fig. 5 left). For SA4 the DOA totally exceeds its limits. A slight advantage of multilateration is, however, evident due to its non-directional approach. The latter part of path 1 and both paths 2 and 3 are well traceable by both approaches. In the leftmost region of the FOV, where SA4 is also active, the ISRR achieves the best results (Fig. 5 right).

The impact of the ISRR is substantial, the mean area is reduced from 18 m² of the whole FOV range to a fraction of a square meter (see Tab. I). To estimate the divergence from the global maximum estimated over the FOV, we find the difference between the result of conventional SRP-PHAT and the result of our method. Error variation over time is presented in Fig. 6, and the Root Mean Square Errors (RMSE) are presented in Tab. I. As it can be seen from the x -axis values in Fig. 6, the DOA approach discards less frames than multilateration (due to non-detection operation) and is therefore more sensitive to the sound source. Also the RMSE is slightly lower for the DOA approach. The overall errors are sufficiently low for speaker localization. Rare bursts of error do occur, however they are instantaneous and appear only during moments of speaker acceleration.

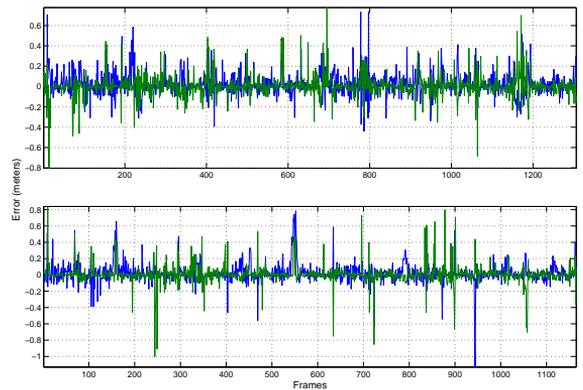


Fig. 6. Difference in localization between SRP-PHAT over the whole FOV and using ISRR: DOA approach (upper) and multilateration (lower). Blue line denotes the x and the green line – the y coordinates.

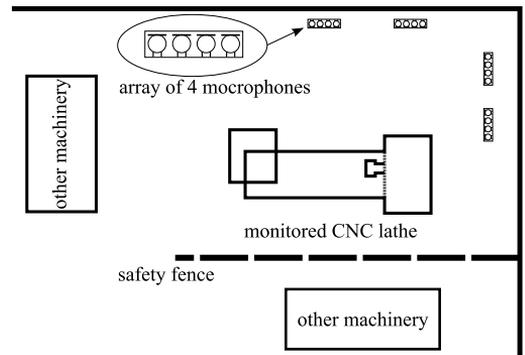


Fig. 7. Array placement at an industrial facility for CNC lathe monitoring.

B. Industrial Machinery Monitoring

For the industrial machinery monitoring experiment the same hardware implementation as for speaker localization is used. The microphones and their triplets are chosen in the

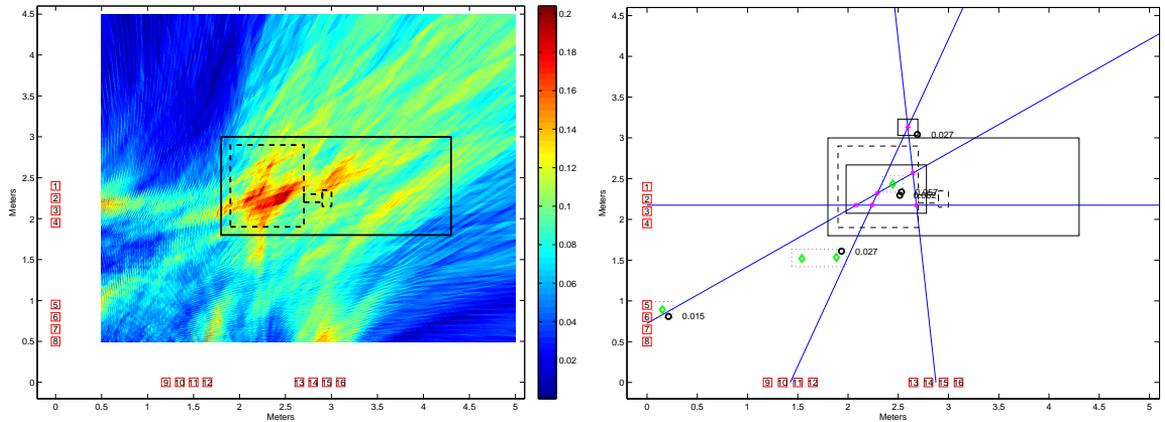


Fig. 8. CNC lathe noise localization with conventional SRP-PHAT (left) and using ISRR followed by SRP-PHAT with SRC (right).

same manner. The placement scheme and the room layout is presented in Fig. 7. The two sub-arrays are placed near a wall at a right angle to the other two sub-arrays, i.e. $\alpha = \{0, 0, 90, 90\}$. The monitored object in the experiment is a large Computer Numerical Control (CNC) lathe. The main noise sources of the lathe are the motor with the gear box and the spindle. The lathe is put through a short working cycle: the motor is activated, the spindle rotates and the carriage with the cutting tool moves beside the cutting area, after which the spindle and then the motor are shut down.

In this experiment the DOA approach to ISRR performs significantly better than multilateration. It seems that multilateration, considering this specific configuration of the array and the large sound emitting area of the lathe, cannot estimate the region confined by the specified maximal distance $D_{\max} = 0.5$ (see Section III-A). A frame corresponding to the active motor and spindle noise is presented in Fig. 8. The region of the motor noise is correctly located by the DOA approach and corresponds to the result of the conventional SRP-PHAT. Multilateration estimates only one small correct region. The triangular array configuration used is evidently not appropriate for localization of large sound sources by multilateration.

V. DISCUSSION AND FUTURE WORK

The ISRR method has proven to perform well for both experimental tasks. The DOA and multilateration approaches perform differently in various situations. The non-directional nature of multilateration enables it to locate the sources out of view for DOA. On the other hand, the directional approach eliminates the possibility of regions duplicating on the opposite side of the array due to reverberation. For a more complicated task of localizing a large noise region, the supremacy of the DOA approach is more evident. The DOA method performs better in a triangular configuration and worse in a square-like configuration. The situation with multilateration is absolutely opposite. Thus, ISRR type may be chosen based on the configuration of the array and the specific application. Both approaches may be used in conjunction for mutual reassurance.

For future work we intend to develop a fully embedded system with array blocks implemented on individual devices. The dataflows between the devices must be thoroughly researched in order to achieve smooth real-time operation. As SRP-PHAT demands information from different sub-arrays, a cooperation scheme with resource allocation must be developed. The operation may proceed in an ad-hoc manner, where the operations are equally distributed between nodes, or a separate node may be allocated for sophisticated computations.

A. Aspects of WSN Organization

The underlying computation and communication system realizing the localization method described in the paper must be able to cope with the real-time requirements set by the employed algorithms. The data delivered to the fusion algorithm from the individual microphone arrays must be temporally valid, i.e. the age of the data delivered to the fusion algorithm must not exceed a set maximum and the data from distinct sources processed by an algorithm must be coherent in time and space. In the current application the spatial aspect is of special importance as due to uncertainties inherent in a distributed architecture – the locations of the individual arrays are not known beforehand and the configuration of the system may change over time. This means that the spatial aspects must be explicitly considered in communication and computation. As the communication and processing delays are dynamic, the system must be also able to cope with these variances. In order to manage with these uncertainties we suggest the use of a proactive middleware [20] as an active mediator of data to and from the individual computing nodes. The middleware enables deterministic data exchange between autonomous (sensing) systems according to the constraints set by individual fusion algorithms and devices. This is achieved by performing constraint propagation from the computing nodes and online data validation based on the propagated constraints. In addition to the data propagation tasks, the middleware can be also used as a tool to synchronize the spatial properties of devices, such as location and orientation, making it for example possible to perform data alignment

for the angles and coordinates among the WSN nodes at the level of the middleware. The proactive middleware can also distribute the individual tasks (e.g. to compute the SRP per FOV) among the WSN nodes using a prescribed scenario (partition) as proposed in [21].

VI. CONCLUSION

The paper proposes a method of search region reduction for the purpose of optimizing acoustic localization. The method targets the sensor array configurations implementable on separate nodes of WSN. Two approaches to the method are proposed and tested on two real experimental signals. The results are positive with the method succeeding with substantially reducing the search region and localizing with small amounts of error. The differences in localization quality for the two approaches under different circumstances do not show definite supremacy of either approach. The results suggest the application of both approaches to region reduction in the final implementation.

REFERENCES

- [1] S. Astapov and A. Riid, "A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device," *International Journal of Electronics and Telecommunications*, vol. 59, no. 2, pp. 151–160, 2013.
- [2] D. Blatt and A. O. Hero, "APOCS: a rapidly convergent source localization algorithm for sensor networks," in *Proc. IEEE/SP 13th Workshop Statistical Signal Processing*, 2005, pp. 1214–1219.
- [3] F. Masson, D. Puschini, P. Julian, P. Croce, L. Arlenghi, P. Mandolesi, and A. G. Andreou, "Hybrid sensor network and fusion algorithm for sound source localization," in *Proc. IEEE Int. Symp. Circuits and Systems ISCAS 2005*, 2005, pp. 2763–2766.
- [4] C. T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2009*, 2009, pp. 2027–2032.
- [5] H. Hongsen, L. Jing, and G. Yang, "Acoustic direction of arrival estimation based on spatial circular prediction," in *Information Technology and Applications, 2009. IFITA '09. International Forum on*, vol. 3, 2009, pp. 177–180.
- [6] A. Dhawan, R. Balasubramanian, and V. Vokkarane, "A framework for real-time monitoring of acoustic events using a wireless sensor network," in *Proc. IEEE Int Technologies for Homeland Security (HST) Conf*, 2011, pp. 254–261.
- [7] T. Liu, Y. Liu, X. Cui, G. Xu, and D. Qian, "MOLTS: Mobile object localization and tracking system based on wireless sensor networks," in *Proc. IEEE 7th Int Networking, Architecture and Storage (NAS) Conf*, 2012, pp. 245–251.
- [8] Z. Merhi, M. Elgamel, and M. Bayoumi, "A lightweight collaborative fault tolerant target localization system for wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 12, pp. 1690–1704, 2009.
- [9] G. Vakulya and G. Simon, "Fast adaptive acoustic localization for sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1820–1829, 2011.
- [10] S. Astapov, J. S. Preden, T. Aruvali, and B. Gordon, "Production machinery utilization monitoring based on acoustic and vibration signal analysis," in *Proc. 8th Int DAAAM Baltic Industrial Engineering Conf*, 2012, pp. 268–273.
- [11] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 2007*, vol. 1, 2007.
- [12] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 71–74, 2011.
- [13] X. Zhao, J. Tang, L. Zhou, and Z. Wu, "Accelerated steered response power method for sound source localization via clustering search," *Science China Physics, Mechanics and Astronomy*, vol. 56, no. 7, pp. 1329–1338, 2013.
- [14] E. Mangas and A. Bilas, "FLASH: Fine-grained localization in wireless sensor networks using acoustic sound transmissions and high precision clock synchronization," in *Proc. 29th IEEE Int. Conf. Distributed Computing Systems ICDCS '09*, 2009, pp. 289–298.
- [15] Q. Wang, R. Zheng, A. Tirumala, X. Liu, and L. Sha, "Lightning: A hard real-time, fast, and lightweight low-end wireless sensor election protocol for acoustic event localization," *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 570–584, 2008.
- [16] Y. Liu and Z. Yang, *Location, Localization, and Localizability: Location-awareness Technology for Wireless Networks*. Springer, 2010.
- [17] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer verlag, 1999.
- [18] S. Zhang, G. Li, W. Wei, and B. Yang, "A novel iterative multilateral localization algorithm for wireless sensor networks," *Journal of Networks*, vol. 5, no. 1, pp. 112–119, 2010.
- [19] S. Astapov, J.-S. Preden, and J. Berdnikova, "Simplified acoustic localization by linear arrays for wireless sensor networks," in *Digital Signal Processing (DSP), 2013 18th International Conference on*, 2013, pp. 1–6.
- [20] L. Motus, M. Meriste, and J. Preden, "Towards middleware based situation awareness," in *Military Communications Conference, 2009. MILCOM 2009. IEEE*. IEEE, 2009, pp. 1–7.
- [21] J. Preden and J. Helander, "Situation aware computing in distributed computing systems," in *Proceedings of the 10th Symposium on Programming Languages and Software Tools*, 2007, pp. 280–292.

Publication 3

Reference

S. Astapov, J. Ehala, and J.-S. Preden, "Performing acoustic localization in a network of embedded smart sensors," *International Journal of Microelectronics and Computer Science (IJMCS)*, vol. 6, no. 3, pp. 86–95, 2015.

Abstract

Situation awareness is an important aspect of ubiquitous computer systems, as these systems of systems are highly integrated with the physical world and for successful operation they must maintain high awareness of the environment. Acoustic information is one of the most popular modalities, by which the environment states are estimated. Multi-sensor approaches also provide the possibility for acoustic source localization. This paper considers an acoustic localization system of dual channel smart sensors interconnected through a Wireless Sensor Network (WSN). The low computational power of smart sensor devices requires distribution of localization tasks among WSN nodes. The Initial Search Region Reduction (ISRR) method is used in the WSN to meet this requirement. ISRR, as opposed to conventional localization methods, performs significantly less complex computations and does not require exchange of raw signal between nodes. The system is implemented on smart dust motes utilizing Atmel ATmega128RFA1 processors with integrated 2.4GHz IEEE 802.15.4 compliant radio transceivers. The paper discusses complications introduced by low power hardware and ad-hoc networking, and also reviews conditions of real-time operation.

Performing Acoustic Localization in a Network of Embedded Smart Sensors

Sergei Astapov, Johannes Ehala, and Jürjo-Sören Preden

Abstract—Situation awareness is an important aspect of ubiquitous computer systems, as these systems of systems are highly integrated with the physical world and for successful operation they must maintain high awareness of the environment. Acoustic information is one of the most popular modalities, by which the environment states are estimated. Multi-sensor approaches also provide the possibility for acoustic source localization. This paper considers an acoustic localization system of dual channel smart sensors interconnected through a Wireless Sensor Network (WSN). The low computational power of smart sensor devices requires distribution of localization tasks among WSN nodes. The Initial Search Region Reduction (ISRR) method is used in the WSN to meet this requirement. ISRR, as opposed to conventional localization methods, performs significantly less complex computations and does not require exchange of raw signal between nodes. The system is implemented on smart dust motes utilizing Atmel ATmega128RFA1 processors with integrated 2.4GHz IEEE 802.15.4 compliant radio transceivers. The paper discusses complications introduced by low power hardware and ad-hoc networking, and also reviews conditions of real-time operation.

Index Terms—Acoustic localization, Wireless Sensor Networks, Direction of Arrival, smart dust, distributed computing.

I. INTRODUCTION

THE continuous process of computer systems integration into all aspects of everyday life paves the way for cyber-physical systems with diverse abilities for interfacing with human operators and the environment, in which these systems exist. Future Internet of Things applications are also envisioned to be ubiquitous systems, which must maintain good situation awareness in order to be able to provide the expected services proactively. Situation awareness is achieved by constant analysis of environment states by sensing different modalities (e.g. acoustic, video, vibration, magnetic, etc.) and sophisticated decision-making through data fusion and system component cooperation. One of the most popular modalities for the majority of environments and human-machine interaction is acoustic signals. Acoustic information is widespread and may be acquired during various physical processes accompanied by sound emission and during human speech analysis.

Acoustic signal analysis has been applied for a great variety of tasks concerning both environment monitoring and human-machine interfaces (HMI). Applications for open environments

span from traffic monitoring [1] and military reconnaissance [2] to monitoring woodland and aquatic wildlife [3]. For confined environments the main applications are person and process monitoring, ranging from home automation [4] and security systems [5] to industrial process control [6]. The majority of monitoring tasks assume pattern matching and classification and use single-sensor solutions. Single channel information is sufficient for low-noise environments with a well defined list of expected events and signal patterns (e.g. HMI with a finite list of voice commands). For the majority of open environments, however, multi-sensor systems are beneficial from several standpoints. Firstly, the monitored area is observed from multiple points of view, which provides more information than a single-sensor system. Secondly, if joint analysis is applied, position of the acoustic source may be localized in observed space. Thirdly, if localization is possible, the sound emitted by a specific source may be filtered from sounds incoming from other directions via beamforming.

The advantages of multi-sensor acoustic systems are even more evident if implemented in Wireless Sensor Networks (WSN). Compact sensor network solutions allow to widen the ordinary localization techniques with more complex multi-node source detection and recognition solutions, e.g. [7]–[10]. A WSN consists of several smart sensor nodes distributed in the observed environment and communicating with each other. The aggregate of local measurements from single nodes can be used to generate a global assessment of the situation. The downside of WSN application is low computational power of sensor nodes. In order to ensure the small size of smart sensors and the longevity of their power supplies, the hardware used in even modern smart sensors is quite limited in terms of computational power. However, these limitations may be overcome to a certain extent through node cooperation and distributed computations.

In this paper we consider a localization approach of Initial Search Region Reduction (ISRR), previously developed by our research group [11], [12], and its implementation in WSN. In our work we use smart sensors of particularly small size and low computational power, known as smart dust motes. Low power imposes many restrictions on signal acquisition and processing, e.g. low sampling rates, limited memory. The paper addresses these restrictions and proposes a system architecture for workload distribution, as well as discusses inter-node communication problems and system real-time operation capabilities. The decrease in localization quality resulting from the sampling rate decrease is demonstrated on a practical example of single speaker localization performed on an eight sensor system.

This research was partly supported by the Estonian Ministry of Defence, the European Defense Agency project IN4STARS, cofinanced by the Estonian ICT doctoral school and the Estonian IT Foundation for Education: IT Academy program.

All authors are with the Research Laboratory for Proactive Technologies, Department of Computer Control, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia

e-mail: sergei.astapov@ttu.ee, johannes.ehala@ttu.ee, jurjo.preden@ttu.ee

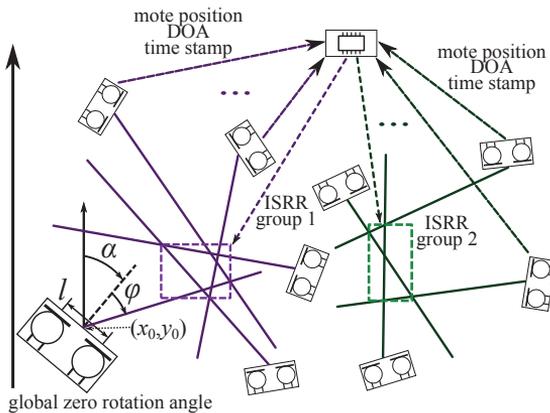


Fig. 1. Schematic diagram of proposed WSN architecture.

II. DISTRIBUTED LOCALIZATION IN WSN

The WSN system is designed for localizing grounded acoustic sources. The motes are placed in the monitored environment in the horizontal plane and localization is performed by estimating the coordinates (x, y) of sound emitting objects. Each mote is equipped with two acoustic sensors spaced by a specific distance l from one another. Localization is based on estimating the time delays of acoustic wave arrival to the sensors, also called Time Difference of Arrival (TDOA). The Direction of Arrival (DOA) of sound from a specific acoustic source is calculated using TDOA. The whole localization system consists of a large number of motes, each one of which plays its specific role in the localization process. This section presents the proposed WSN architecture along with the distributed approach to localization.

A. WSN Architecture

The proposed WSN architecture is designed for applications in both open (urban, woodland, etc.) and confined (home, office, industrial facility, etc.) environments. The network consists of two types of motes: smart sensors and fusion nodes. Dual channel smart sensors acquire acoustic information and perform DOA estimation. Fusion nodes gather DOA estimates and perform further steps of localization, which are discussed in Section II-C. The schematic diagram of the architecture is presented in Fig. 1.

The sensor motes are dispersed in the monitored environment either in an orderly or random fashion. In confined environments an orderly placement is more likely, because sensors are usually mounted on room walls or ceilings. In open environments, however, it is rarely the case — the sensors may be attached to buildings, light posts, trash bins, etc. in urban and to trees, rocks, etc. in natural environments. Thus a general case is assumed, where the sensor's location is defined by the coordinates of its point of reference (x_r, y_r) and the angle α , by which the sensor is steered from the global angle reference. For example, sensor location may be estimated via the Global Positioning System (GPS), in which case the point

of reference is the GPS unit. For environments, where GPS signals are unavailable, other location algorithms based on Radio Frequency (RF) [13] or sound [14] may be adopted. The global angle reference may be defined by Earth's magnetic field and the angle α estimated using a magnetometer. The central point of the microphone pair (x_0, y_0) , for which the DOA is actually estimated, is defined by the reference point (x_r, y_r) and may coincide with it. The coordinates of the i -th microphone (x_i, y_i) are shifted by $\pm l/2$ from (x_0, y_0) and then the steering by α is performed as

$$\begin{pmatrix} x_i^{(rot)} \\ y_i^{(rot)} \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \end{pmatrix}. \quad (1)$$

Sensor motes are partitioned into groups, where a single mote can belong to any number of groups. Each group must have a common Field of View (FOV), i.e. all motes observe the same area. The whole network may consist of several groups or each group can constitute a separate sub-network. Group partitioning is in essence a clustering task, for which two aspects are taken into consideration. Firstly, motes must have a common field of view as the considered localization procedure uses a directional approach. In this regard, the observed area is not necessarily enclosed by motes, as shown in Fig. 1, but may be observed from one or several sides. Secondly, a group must have certain homogeneity. Motes located too far from the group's centroid may be useless to the localization effort in low Signal to Noise Ratio (SNR) environments or when the sound emitted by the source of interest is too weak. Furthermore, non-homogeneous groups present additional challenges for wireless communication.

Fusion nodes of the WSN perform mote grouping during network initialization and later participate in localization. For an orderly configuration of motes a single fusion node may be assigned to coordinate the activity of the whole WSN. In a random configuration each sensor mote may be a part of several groups and each group may be governed by several fusion nodes. In order to ensure coverage of all groups, fusion nodes reach an agreement concerning which node will govern which mote group. In this process communication signal strength is taken into account, meaning that a fusion node will adopt a group, to which it has the strongest connection. However, if there exists an ungoverned group, a redundant (i.e. covering an already covered group) fusion node closest to it will switch to that group. Mote communication is further discussed in Section III-C.

B. Acoustic Source Localization

Acoustic localization consists of estimating the DOA and distance to the sound source. DOA, in turn, consists of estimating the Angle of Arrival (AOA) and elevation of the acoustic wave front. As we operate only in the horizontal plane, we assume zero elevation, and thus for DOA estimation only the AOA is needed to be computed. The AOA, as it was mentioned earlier, is calculated based on TDOA. In choosing the trigonometric approach to AOA calculation an assumption of near or far field source location must be made. As sound waves propagate spherically, wave front curvature must be

accounted for in the calculations. The near field disposition assumes spherical fronts, whereas waves originating in the far field are spread enough by the time they reach the sensor to be considered linear. The far field assumption is met for a linear microphone array if the inequality

$$|r| > \frac{2(Md)^2}{\lambda_{\min}} \quad (2)$$

holds, where M is the number of microphones, d is the inter-microphone distance, λ_{\min} is the minimal wave length of the wide-band acoustic signal and r is the radial distance from the array center to the source. For our implementation we assume the far field disposition.

There exists a variety of methods for acoustic localization, most of which also employ TDOA as a basic principal. The methods utilize sensor array structures, in which a large number of microphones is arranged in some specific manner (e.g. linear, tetrahedron, spherical, etc.). The TDOA and consequently DOA is generally estimated using some measure of correlation between different sensor signals. For example a popular method of Steered Response Power with Phase Transform (SRP-PHAT) computes cross-correlation across all pairs of microphones at the theoretical time delays associated with all possible DOA to estimate the cumulative signal energy for each discrete point of the FOV [15]. Multiple Signal Classification (MUSIC) applies eigenspace analysis to the signal correlation matrix in order to get the largest eigenvalues corresponding to the most probable DOA [16]. Multilateration methods estimate distances from every sensor to the source and calculate the position of that source using geometry of triangles and circles (spheres for 3D cases). Distance estimation is usually based on TDOA [17].

Notice that typical acoustic localization methods utilize information from every sensor. This fact does not pose a problem for wired systems with a single powerful computational hub. In WSN, however, collecting raw signals from nodes is a real challenge, especially if the number of nodes is large and signal frames are long. Recent developments in distributed localization combine individual sensor estimates for source positioning by applying, for example, maximum likelihood iterative search [18], or fuzzy clustering [19]. We try to overcome problems associated with communicating signal frames by applying a simplified localization approach of Initial Search Region Reduction (ISRR), recently developed by our research group.

C. Initial Search Region Reduction in WSN

The main idea behind ISRR lies in maximally confining the region of acoustic source disposition as a preliminary procedure to SRP-PHAT or other localization method [12]. Having already established that SRP-PHAT requires raw information from all sensors in the network, we do not apply it for this specific implementation. For object or person monitoring applications, where localization to a single point is not obligatory, ISRR confined regions serve as a sufficient estimate of object location. This section presents ISRR for the specific implementation of dual sensor mote WSN.

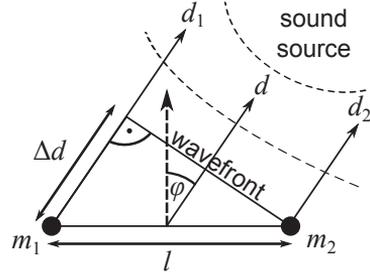


Fig. 2. DOA estimation for a pair of microphones.

Having a group of K dual sensor motes, the ISRR is performed in the following steps:

- 1) Estimate the DOA for each of K motes.
- 2) Generate vectors spanning from the mote sensor pair centers to the bounds of the FOV in the directions of DOA.
- 3) Find points of intersections of these vectors.
- 4) Find groups of points no farther than D_{\max} distance units (meters) from their centroid and enclose the areas, in which these groups reside, in rectangles.
- 5) Perform control of false detection, discard areas not meeting specific criteria (optional).

Step 1 is performed on each sensor mote, steps 2–5 are performed on the group's fusion node.

The DOA are estimated for the front view of the sensor pair, i.e. from -90° to 90° . Considering Fig. 2, the sound wave emitted by a source in the far field is acquired by the microphones m_1 and m_2 with a time delay $\tau = \Delta d/c$, where c is the speed of sound in m/s. The delay takes the values $\tau \in [-\tau_{\max}, \tau_{\max}]$, where τ_{\max} is the delay of sound traveling directly from one microphone to the other (i.e. at $\pm 90^\circ$). To estimate τ we apply cross-correlation to the two signals:

$$R(\tau) = \sum_{k=0}^{n-1} x_{m_1}(k) \cdot x_{m_2}(k - \tau), \quad (3)$$

where n is the length of the signals in samples. The maximum of the cross-correlation defines the time delay, and the AOA is obtained by

$$\varphi = \sin^{-1} \frac{\tau \cdot c}{l} = \sin^{-1} \frac{\Delta k / f_s \cdot c}{l}, \quad (4)$$

where l is the distance between the microphones and τ is represented in terms of delay in samples Δk and the sampling frequency f_s . The speed of sound in air is dependent on the ambient temperature and is equal to

$$c = 331.45 \sqrt{1 + \theta/273}, \quad (5)$$

where θ is the air temperature in Celsius.

At this point AOA validation is performed. If the correlation peak is not sharp and outstanding enough, the AOA φ is discarded. This way, in absence of a sound source or in case of high ambient noise, invalid estimates are avoided early on. We use the deviation from the mean for this metric:

$$\max(R(\tau)) > (1 + TH) \cdot \overline{R(\tau)}, \quad (6)$$

where TH is the threshold of deviation, which depends on the SNR in the environment. We use $TH = 0.2$ in our experiments. The angles φ from every mote are sent as DOA estimates to the fusion node.

The fusion node receives $K_1 \leq K$ DOA estimates $\phi_{i^*}, i^* \in (1, \dots, K_1)$, and adds the mote's rotation angles α_i to them. Vectors $\overrightarrow{AB}_{i^*}$ are computed with the starting point $A_{i^*} = (x_{1,i^*}, y_{1,i^*})$ being the coordinate of i^* -th sensor pair's center and the ending point $B_{i^*} = (x_{2,i^*}, y_{2,i^*})$ being the point at a bound of the FOV steered by ϕ_{i^*} from the pair's center. Intersection points of all pairs $\overrightarrow{AB}_h, \overrightarrow{AB}_k$ are calculated by

$$\mathbf{I}_{hk} = (I_x, I_y) = \left(\frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(x_{1,k} - x_{2,k}) - (x_{1,h} - x_{2,h})(x_{1,k}y_{2,k} - y_{1,k}x_{2,k})}{(x_{1,h} - x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k} - x_{2,k})}, \frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k}y_{2,k} - y_{1,k}x_{2,k})}{(x_{1,h} - x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k} - x_{2,k})} \right). \quad (7)$$

As a result we have a set of \mathbf{I}_{i^*} intersections, $i^* \in (1, \dots, K_2)$, $K_2 \leq \binom{K_1}{2}$. To get the initial search areas, these intersection points are partitioned by their relative distance. For the maximum distance D_{\max} the partitioning is performed in the following manner:

- 1) **IF** no points $\mathbf{I} = \emptyset$ **THEN** no partitions, $\mathbf{P} = \emptyset$ **STOP**
- 2) **ELSE IF** only 1 point \mathbf{I}_1 **THEN** $\mathbf{P}_1 = \mathbf{I}_1$ **STOP**
- 3) **ELSE** number of partitions $j = 0$
- 4) **WHILE** $|\mathbf{I}| > 0$, where $|\mathbf{I}|$ is the cardinality of the set \mathbf{I} , calculate centroid of free points $C_{\mathbf{I}} = 1/|\mathbf{I}| \cdot \sum \mathbf{I}$.
- 5) Calculate Euclidean distance of all free points to centroid $D_k = \sqrt{\sum_{s=1,2} (I_{k,s} - C_{\mathbf{I},s})^2}$, choose point with minimal distance, $j = j + 1$, insert point to \mathbf{P}_j , remove point from set of free points \mathbf{I} .
- 6) Calculate partition centroid $C_{\mathbf{P}_j} = 1/|\mathbf{P}_j| \cdot \sum \mathbf{P}_j$, get Euclidean distance for all free points $D_k = \sqrt{\sum_{s=1,2} (I_{k,s} - C_{\mathbf{P}_j,s})^2}$.
- 7) **IF** $\min(D) \leq D_{\max}$ **THEN** insert point corresponding to $\min(D)$ into \mathbf{P}_j , delete point from set of free points \mathbf{I} , **GO TO** Step 6.
- 8) **ELSE IF** $|\mathbf{I}| > 1$ **THEN GO TO** Step 4.
- 9) **ELSE** $j = j + 1$, put last remaining point to \mathbf{P}_j .

After obtaining the partitions \mathbf{P} , their areas are enclosed by rectangles with the edges denoted by the partitions lowest leftmost and highest rightmost points. As a result several regions may occur in the same FOV. Also while a vector from one array may cross with several other vectors, redundant 'echoing' regions may arise. These can be removed by applying SRP-PHAT to every region and comparing SRP values, or by using tracking filters. This work, however, does not focus on redundant region removal.

The procedure is applicable to multiple target localization. If more than two sensors are used in the array, several AOA may be estimated [12]. Each dual channel mote, however, points to a single direction of the strongest acoustic source. As sound pressure decreases exponentially with propagation, each mote group identifies a source closest to it. If a group is well spread, several targets may be identified within the FOV based on the same principle.

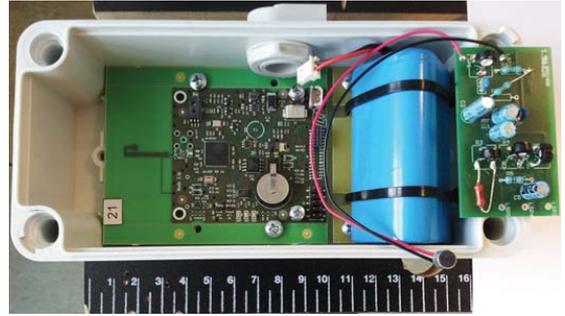
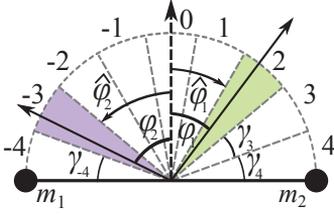


Fig. 3. Packaged WSN mote with a sensor amplification circuit (scale in cm).

III. WSN IMPLEMENTATION

The proposed distributed localization method with ISRR is implemented on a small WSN comprising several smart dust motes. The motes are equipped with Atmel ATmega128RFA1 microprocessors, which conveniently provide an on-chip AD converter for signal acquisition and a radio transceiver for WSN communication. The microprocessor has a clock speed of 16 MHz and provides 16kB of SRAM memory for operation with an additional 128kB of flash memory for program code. The on-chip Analog to Digital Converter (ADC) has a resolution of 10 bits and is able to sample with rates up to 330 kHz. However, actual experiments were carried out with a sampling rate of 2 kHz for each microphone channel, since higher sampling rates provided inconsistent and erroneous results during data acquisition. We were able to determine that erroneous results were caused by signal leakage from the previous ADC channel to the succeeding channel when switching between channels, but the cause of the leakage could not be determined. Mote-to-mote communication was realized with the IEEE standard 802.15.4 compliant radio transceiver with an effective indoor communication range of approximately 30 meters. The IEEE 802.15.4 standard supports transfer rates up to 250 kbit/s and packet sizes not larger than 127 bytes. Vansonic PVM-6052 electret condenser microphones were used for acoustic signal acquisition with additional circuitry performing signal amplification and the normalization needed for the microprocessor ADC input. For every mote a pair of microphones was mounted facing the same direction on a plastic board, which was then attached to the mote's plastic chassis.

The smart sensor mote chosen for the experiments is presented in Fig. 3. Microphone amplification circuitry is situated on the right and the microphone itself — in the bottom right corner. The mote is powered by a 3.7 V, 6600 mAh battery block (left from the sensor circuit). The motes are packaged in protective frames 16 cm in length. The poor computational characteristics listed above are typical for smart sensor motes. The reason for this is that these motes must work ubiquitously and autonomously with the battery they are provided for as long as possible. For example, the battery used in our configuration can sustain the motes for 1–1.5 years in a low duty cycle mode and approximately a month in constant

Fig. 4. Discretization of the AOA scope, defined by Δk_{\max} .TABLE I
INTER-SENSOR DISTANCE FOR DIFFERENT SAMPLING RATES

Δk_{\max}	n_{AOA}	l (cm)		
		$f_s = 44.1$ kHz	$f_s = 2$ kHz	$f_s = 500$ Hz
1	3	0.8	17.2	68.7
2	5	1.6	34.4	137.4
3	7	2.4	51.6	206.1
10	21	7.8	171.7	686.8

operation mode. The goal here is to show that if localization and ISRR can be carried out on a smart sensor mote network, it is reasonable to assume it can also be implemented on larger networks with computationally more powerful motes.

A. Implications of Using Low Sampling Rates

The essential operation for AOA estimation is the signal cross-correlation (3). As our time delay τ is bounded by τ_{\max} and τ is expressed in delay in samples Δk , then Δk is also bounded by a maximal sample shift $\Delta k \in [-\Delta k_{\max}, \Delta k_{\max}]$, where Δk_{\max} is calculated as

$$\Delta k_{\max} = \left\lfloor l \cdot \frac{f_s}{c} \right\rfloor, \quad (8)$$

where $\lfloor \cdot \rfloor$ denotes rounding to the largest previous integer (floor function). Consequently the view scope of the sensor pair is reduced to the number of possible AOA values $n_{AOA} = 2 \cdot \Delta k_{\max} + 1$. Fig. 4 depicts a view scope divided into 9 sectors. For any actual AOA φ , only its discrete margin $\hat{\varphi} \in [\gamma_{-\Delta k_{\max}}, \gamma_{\Delta k_{\max}}]$, corresponding to the correlation maximum, can be estimated. For devices capable of only low sampling rates this poses a problem in terms of compromise between the values of l and Δk_{\max} . Consider Table I. The standard CD sampling rate of 44.1 kHz is used for reference and l is calculated using (8): $l = \Delta k_{\max} \cdot c / f_s$. The table shows that to provide even the smallest n_{AOA} the inter-sensor distances must be quite considerable at low f_s . It is clear that if mote dimensions do not exceed 15–20 cm, it would not be reasonable to make $l = 1.7$ m to provide the sensor scope with 21 possible AOA.

We chose $l = 0.7$ m for our motes, which gives $n_{AOA} = 9$ possible AOA values with an average step of 19.7° at the used rate of $f_s = 2$ kHz. These are calculated using (4) and presented in Fig. 5. The substantial difference with high sampling rates is also evident from the figure — for the same sensor distance at a rate of $f_s = 44.1$ kHz, the AOA number is equal to 179 with an average step of 0.92° . A small n_{AOA} introduces additional error into the localization process as the

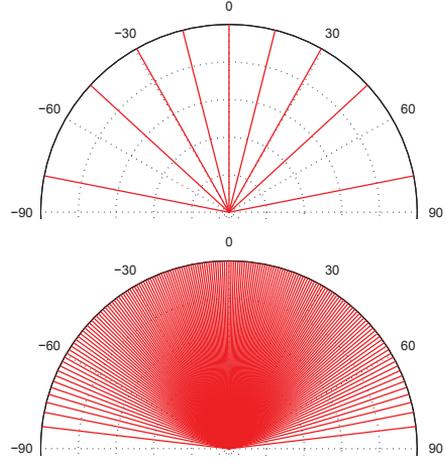
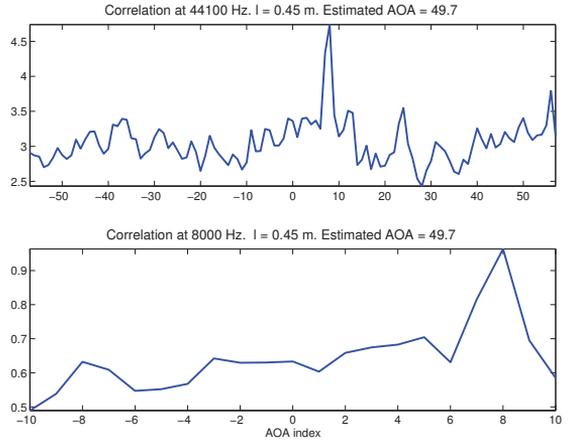
Fig. 5. Possible values of AOA for a sensor pair with $l = 0.7$ m and sampling rate $f_s = 2$ kHz (top); $f_s = 44.1$ kHz (bottom).

Fig. 6. Results of signal cross-correlation at different sampling rates.

ISRR estimated regions may become larger and get shifted from the true area occupied by the sound source. For example an angle step of 19.7° can give an error of 1.8 m if the sound source is situated only 5 m away from the sensors. To manage the situation a large number of motes must be used, preferably steered by different angles α (i.e. not facing in exactly the same direction). Random mote placements allow AOA uncertainty regions to superimpose on one another thus reducing the discrete gaps.

Low sampling rates also influence cross-correlation in two ways. Firstly, if the signal contains many strong components in the higher frequencies, they are not acquired at low sampling rates. As a result aliasing may occur, which in turn reduces the correlation reliability. A precise peak corresponding to a single Δk loses its steepness and spreads to several values. This makes AOA estimation and the control metric (6) less

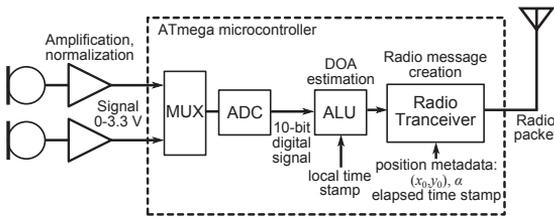


Fig. 7. Sensor mote architecture schematic.

reliable. Secondly, the cross-correlation yields exactly n_{AOA} coefficients, and if this number is low, the correlation peak cannot stand out from the average correlation level as much as in the case of high sampling rates. At low SNR the peak becomes almost uniform with the average level and control metric (6) declares the result invalid for the majority of signal frames. Both effects are evident from Fig. 6. The upper subplot shows the result of cross-correlation of two signals sampled at 44.1 kHz and the lower — at 8 kHz. For both cases the inter-microphone distance was equal to $l = 0.45$ m and the AOA from the acoustic source, as well as signal power, were the same. As n_{AOA} is more than five times larger in the case of $f_s = 44.1$ kHz and more signal energy information is contained in a single frame, the correlation peak is much more steep and evident than in the $f_s = 8$ kHz case. Generally at a fixed f_s correlation results are improved by increasing the signal frame length, thus providing more signal energy information. Here a compromise between correlation result reliability and the device refresh rate, as well as the amounts of required memory must be reached.

B. Resource Management and Scheduling

The smart sensor mote must divide its computational resource between two main tasks: sampling the ADC and performing cross-correlation (3) on the sensor signals. With the current hardware setup and computational power of the motes, sampling the ADC and doing correlation calculations at the same time is not possible, therefore currently these tasks are performed separately. A simplified schematic of sensor mote architecture and computational steps is presented in Fig. 7. First the ADC samples both channels and obtains a 0.2 second frame (totaling 400 samples at 2 kHz sampling rate) from each channel. Since we have a single ADC, both channels cannot be sampled simultaneously. Therefore there is a channel switching delay of about 150 microseconds. The phase shifts between channels due to these switching delays can be accounted for and they do not affect cross-correlation calculations significantly.

After the frames have been acquired, the ADC is stopped and the cross-correlation of frames is calculated. If a sound source is detected in the FOV of the sensor mote and the DOA of sound waves is found, then this information along with the spatial and temporal metadata is broadcast over the sensor network. Calculating the cross-correlation of frames and composing and transmitting the DOA message does not require much time (50 – 100 ms). Nevertheless, the smart

sensor mote keeps track of this elapsed time and right before broadcasting the DOA message, appends the elapsed time to the message. Elapsed time is the time difference between the moment when the acoustic data reaches the processing unit and the moment of message composition. It indicates the 'age' (in milliseconds) of the calculated DOA value.

The fusion node has also two key tasks: listening to messages broadcast by sensor motes and performing sound source region estimation. As mentioned in Section II-A, sensor motes are partitioned into groups. Group membership is established by position metadata (coordinates and steering angle) found in the messages. Every fusion node maintains a lookup table consisting of coordinates of motes, which the node includes in its group. The table is updated every time a mote with new coordinates appears. A fusion node is only interested in messages arriving from motes in its own group and discards others. Sound source region estimation is performed when enough DOA messages with fresh data have been received. When the last data received becomes too old to be useful and no new data is received then the fusion node switches to idle mode until new messages arrive. Currently the expiration time for DOA information is 3 seconds, after which the fusion node discards the data.

C. Communication Strategy and Real-Time Operation

The benefit from using smart sensors is that preliminary sensor signal analysis is done on spot. With large networks it is not conceivable that raw sensor data is forwarded to some fusion unit. With the sampling parameters proposed for the WSN experiment (2 kHz sampling rate on both channels and frame lengths of 400 samples) it would take one sensor node in ideal conditions at least 0.15 seconds to communicate its entire measurement buffer to a fusion node. It is clear that with only a small number of motes the communication channel will be congested and system operation will be paralyzed. Hence, it is necessary to perform signal processing on the sensor motes and only communicate forward the results. This is what is proposed in our approach.

In the WSN experiment the communication scheme is built upon indirect messaging, i.e. motes broadcast the messages, which are to be received by fusion nodes. The indirect approach offers a quick and robust method for validating the data processing algorithm. The approach makes the network scalable to a certain extent and easily integrable into a larger system of systems [20].

In the algorithm validation experiment the sensor nodes broadcast messages about their latest DOA estimates and include their location and orientation metadata as well as temporal metadata with the broadcast message. A fusion node receiving these messages collects the DOA and metadata information and regularly initiates the ISRR algorithm to locate possible sound sources. Note that communication is performed in one direction — from sensor motes to fusion nodes. Therefore sensor motes do not possess any information concerning group partitioning.

The proposed WSN communication strategy is asynchronous, i.e. sensor motes do not have a global clock, which

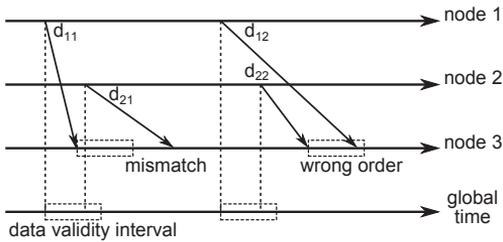


Fig. 8. Main problems situated with asynchronous data interchange.

would enable coordination of signal acquisition and message broadcasting. Rather every mote transmits a DOA estimate after every signal acquisition and processing loop (in our case it lasts 250–300 ms). The receiving fusion node can then estimate the time of DOA calculation in its own local time using the elapsed timestamp value and the common understanding of the millisecond time unit. For real-time operation two parameters must be strictly defined: the maximum duration of the DOA estimation loop and the maximum communication and processing delay for incoming messages. This must be done to enable estimating the validity of DOA estimates for position estimation. The asynchronous decentralized approach described above is robust and simple, suitable for algorithm validation, however in operational systems better control over data paths is desirable, which can be achieved by applying proactive middleware, as described in the next section.

D. Proactive Middleware and Data Validation

Performing computations in dynamic ad-hoc wireless sensor networks presents many challenges in terms of guaranteeing data correctness. The data used in the fusion process must satisfy certain temporal and spatial constraints (i.e. its age must not be greater than a pre-specified value or come from a certain location). It is easy to achieve such guarantees in a system with a fixed configuration, however in a dynamic setting the systems must evaluate these data properties at runtime.

For effective acoustic localization the DOA calculations ideally must be performed simultaneously. In real conditions a time interval must be specified in which the estimates are considered temporally coherent. Due to undefined transmission delays the data may arrive with considerable delays and therefore not satisfy the coherence requirement. Coherent data is vital for all signal processing tasks, like tracking and trajectory estimation. Fig. 8 graphically explains the validity interval mismatch and wrong order of message arrival. The validity interval specifies the time period, during which the data is considered coherent. Due to undetermined transmission delays the coherent messages may not fit in the interval. On the other hand, they may arrive in time, but in a wrong order, which may later cause errors in tracking procedures.

In [20], [21] we have presented the concept of proactive middleware, called ProWare, which is a lightweight distributed middleware component running on every element of the WSN system (see Fig. 9). ProWare implements a subscription based

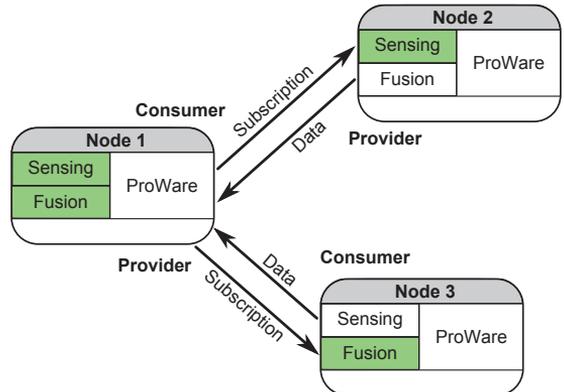


Fig. 9. Principle diagram of proactive middleware mediation.

information exchange scheme, where the data consumer (fusion node) can subscribe for data from the providers (sensor nodes computing DOA estimates). ProWare also handles data validity checking ensuring that the data received at the fusion node satisfies the constraints for a given fusion operation (i.e., that the data is temporally coherent). ProWare manages the process of finding appropriate data providers (in our case sensor motes with overlapping fields of view) and setting up the data exchange paths with the consumers (fusion nodes). Both the validity checking and provider-consumer agreements are performed on-line. Among other tasks the middleware component keeps track of the different clock offsets of the motes and regularly checks and updates the change (caused by clock drift, jitter etc.) in these offsets. This temporal information is then used to estimate the time of measurement of the data in local time of the data consumer.

IV. EXPERIMENTAL RESULTS

We demonstrate the applicability of our proposed method of acoustic localization and the implications situated with using low sampling rates by performing an experiment of single speaker localization. For the initial experimental validation we performed data acquisition using an Agilent U2354A data acquisition device (DAQ) and performed localization offline in the MATLAB environment. Data acquisition is performed at two sampling rates: $f_s = 8$ kHz and $f_s = 2$ kHz per sensor (as the motes are able to sample the signal at 2 kHz). For the experiment we use four microphone pairs arranged in an angular configuration — two microphone pairs are placed perpendicular to the other two. For comparison with our proposed approach we apply the SRP-PHAT method, which was mentioned in Section II-B. The fact that SRP-PHAT is a highly resource demanding procedure is another reason for choosing MATLAB for computations. The ISRR procedure implemented in MATLAB is identical to the one running on the motes, therefore there is no difference in localization. Mote communication and asynchronous data validation is not considered in this experiment. Additionally we apply simplified SRP-PHAT with Stochastic Region Contraction (SRC) to the

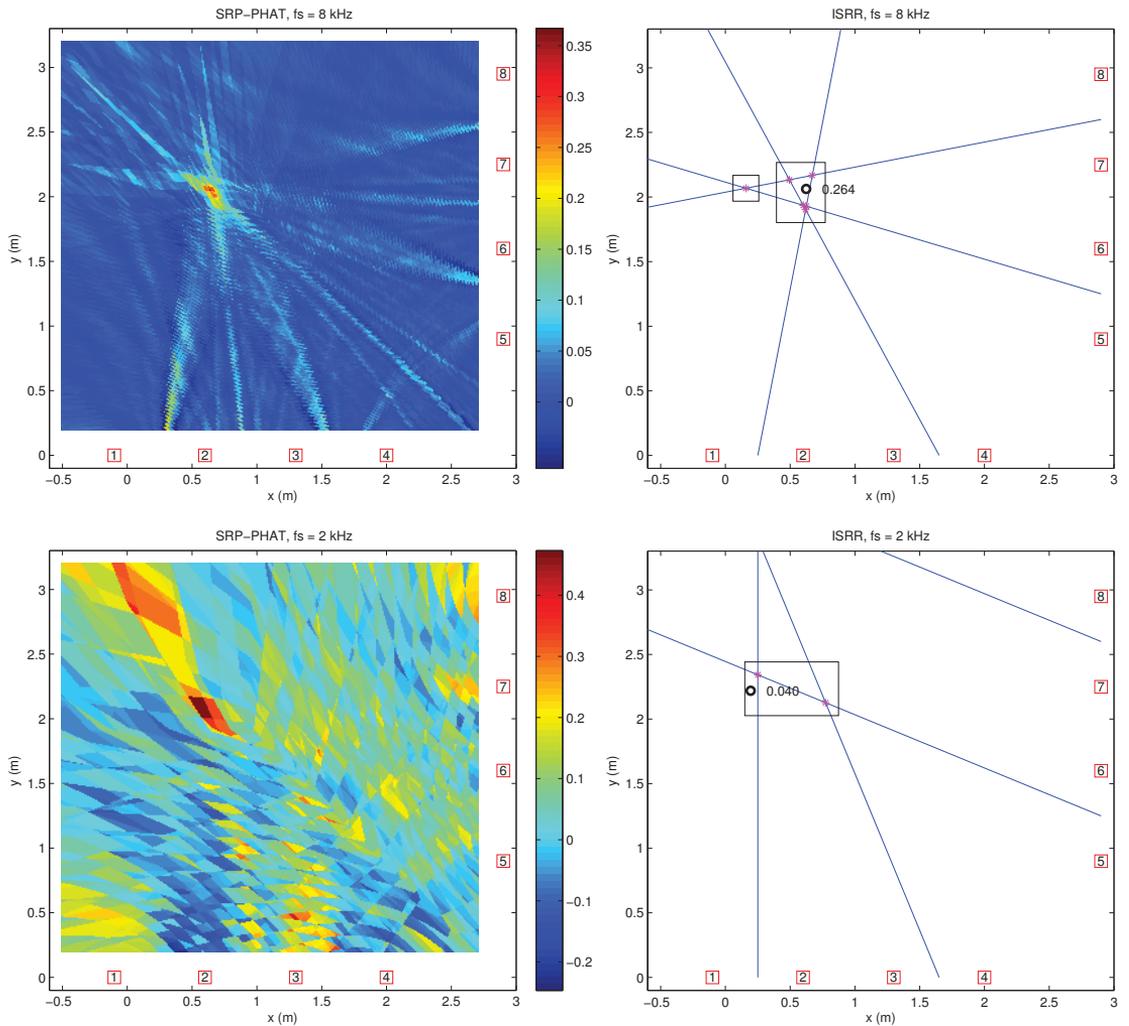


Fig. 10. Results of acoustic source localization using four pairs of microphones and two approaches: SRP-PHAT and the proposed ISRR.

source areas estimated by ISRR in order to determine the cumulative acoustic energy levels in these areas. The principles of SRP-PHAT and SRC are briefly introduced in the Appendix. For better comprehension of the result analysis it is advised to get familiarized with these localization methods.

The results of speaker localization are presented in Fig. 10. The speaker is placed at position (0.7, 2) meters and a short speech recording is made. We analyze the signal frame by frame, as it is done in our WSN implementation, using a frame length of 200 ms. For SRP-PHAT the area discretization value is set to 1 cm². Fig. 10 portrays localization results for a single signal frame related to the same time instance. In the figure microphones are represented by red squares with numbers inside them. The four microphone pairs are then 1, 2; 3, 4; 5, 6; 7, 8. SRP-PHAT results of cumulative energy values

are plotted using a red-green-blue color scale. DOA vectors of ISRR are denoted by blue lines, their intersections — by magenta stars, the estimated regions — by black rectangles and the maximum of SRP-PHAT with SRC — by black circles.

It can be seen from the two upper plots of Fig. 10, that at $f_s = 8$ kHz both SRP-PHAT and ISRR localize the sound source efficiently. The SRP-PHAT region of particularly high cumulative energy (i.e. orange to red on the scale) is reduced to approximately 0.01 m². The region estimated by ISRR is significantly larger, but proportionate to the SRP-PHAT region of medium cumulative energy (i.e. green on the scale). However, the sound source is fully confined by the region, as confirmed by the SRP-PHAT with SRC estimate on the region.

On the other hand, both methods suffer from the problems situated with the low sampling rate of $f_s = 2$ kHz, as it can

be clearly seen in the lower plots of Fig. 10. SRP-PHAT high cumulative energy region enlarges to approximately 0.25×0.3 meters with an ‘echoing’ region situated in the top left corner of the FOV. The decrease of n_{AOA} and the number of signal samples in a frame, described in Section III-A, affects SRP-PHAT, producing a more rough and less comprehensive image. Nevertheless, SRP-PHAT localizes the source properly. ISRR performs worse, missing the source slightly up the y -axis. The reason lies in the fourth microphone pair failing to estimate the DOA correctly. Although the confined region is close to the source position, it does not confine it fully. This example clearly shows the need for a larger number of microphone pairs (motes) to be used for successful localization.

Generally the performance of ISRR is comparable to the performance of SRP-PHAT in terms of localization accuracy, with the deviation of ISRR estimates from the localization results over the whole FOV being less than 0.13 m [12]. Considering the dimensions of usual acoustic sources under localization (no less than 15–20 cm), this deviation is quite permissible. In the case of low sampling rates, however, both SRP-PHAT and ISRR become less reliable. Therefore it cannot be explicitly stated, that our proposed method suffers from the limitations of embedded hardware more than the other. On the contrary, ISRR significantly reduces the number of computations required for localization up to an area of a fraction of a square meter [11].

V. DISCUSSION AND FUTURE WORK

The paper mainly considers the limitations of data processing hardware and touches upon the implications of asynchronous data interchange slightly. The proactive middleware component, which has been tested on abstract data, was not fully implemented for the purposes of acoustic localization on a large number of smart dust motes. Thus further testing on a large network of motes is required in order to estimate the feasibility of the component for our specific purpose.

Considering our recent developments, we have achieved a reliable 4 kHz sampling rate per channel on the Atmel ATmega128RFA1. As a consequence, the localization quality has noticeably improved. The problem of signal leakage when changing channels is still not solved, but the higher sampling rate was reached by changing the ADC clock speed, which consequently affected the settling time (i.e., the time automatically inserted by hardware to clear and prepare the ADC registers for channel change) in-between changing ADC channels and alleviated the signal leakage problem. We do not plan on improving this hardware platform any further, instead we are considering more powerful embedded devices, such as the Gumstix Overo series, for our localization approach. The implementation on Atmel ATmega displays promising results, however, it requires a significant number of motes and their significant dispersion in the FOV to sustain the localization quality on low sampling rates, and not every application and environment will allow these things. For the applications, where only a few motes are permitted, the motes must estimate the DOA more accurately, and that requires more resources. With the increase of computational power it

will be also possible to increase the number of sensors on each mote, which will increase the reliability of DOA estimates.

Increasing the number of sensors per mote will also allow for 3D acoustic localization. As the elevation AOA cannot be accurately estimated by a pair of horizontally placed microphones, 3D localization will require additional microphones to be utilized to estimate the AOA in the vertical plane. For this direction of future research the proposed ISRR method is to be expanded in order to be able to confine volumetric regions, as opposed to planar areas, discussed in this paper.

VI. CONCLUSION

The paper considers an acoustic source localization system and its implementation in a WSN consisting of dual channel low power smart sensors. A decentralized ad-hoc WSN structure for distributed computation is proposed, which reduces the number of computations per network node and introduces redundancy to the system, making it more reliable. The applied localization approach is presented and different problems situated with system implementation on specific hardware are handled. Computationally weak smart sensor hardware imposes limitations on the signal sampling rate, processing time and communication bandwidth. A compromise between a reliable sampling rate, suitable sensor pair geometry and localization accuracy is established. The applied asynchronous communication strategy reduces message interchange and does not overwhelm the network’s fusion nodes. A practical experiment is held to test the proposed localization method and compare it to a popular and effective, but resource demanding approach. Experimental results show, that both methods suffer from the limitations induced by low power embedded hardware. However, the proposed method is capable of localization with permissible accuracy.

APPENDIX

OVERVIEW OF SRP-PHAT AND SRC

Steered Response Power with Phase Transform (SRP-PHAT) is a technique of estimating the DOA of sound signals. The SRP $P(\vec{a})$ is a real-valued functional of a spatial vector \vec{a} , defined by the FOV of a specific microphone array. The high maxima in $P(\vec{a})$ indicate the estimates of sound source location. $P(\vec{a})$ is computed for each direction as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) across all pairs of microphones at the theoretical time delays associated with the chosen direction. Consider a pair of signals $x_k(t)$, $x_l(t)$ of an array consisting of M microphones. The times of sound arrival from point a to the two microphones are $\tau(a, k)$ and $\tau(a, l)$ respectively. Hence the time delay between the two signals is $\tau_{kl}(a) = \tau(a, k) - \tau(a, l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega\tau_{kl}(a)} d\omega, \quad (9)$$

where $X_i(\omega)$ is the spectrum (the Fourier transform) of signal x_i and $X_i^*(\omega)$ is the conjugate of that spectrum. Ψ_{kl} is the PHAT weight of the inverse of spectral magnitudes:

$$\Psi_{kl} = \frac{1}{|X_k(\omega) X_l^*(\omega)|}. \quad (10)$$

Conventional SRP-PHAT performs as many evaluations (9), as there are points in \bar{a} , the number of which is defined by the dimensionality of the FOV and the accuracy measure, that partitions the area (or volume) into small discrete regions. The method is highly resource demanding, particularly when applied to large areas of observation. The number of evaluations (9) is significantly reduced by applying Stochastic Region Contraction (SRC), which iteratively narrows down the search volume for the global maximum [15]. SRC starts with the initial search volume (i.e. the whole FOV), stochastically explores the functional of that volume by randomly picking a specific number of points, then contracts the search volume into a sub-volume containing the desired global optimum and proceeds iteratively until the SRP maximum can be located with a finite precision.

REFERENCES

- [1] S. Astapov and A. Riid, "A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device," *International Journal of Electronics and Telecommunications (JET)*, vol. 59, no. 2, pp. 151–160, 2013.
- [2] S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, "Object detection for military surveillance using distributed multimodal smart sensors," in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP 2014)*, 20–23 Aug. 2014, pp. 366–371.
- [3] H. Lohrasbipeydeh, A. Zielinski, and T. Gulliver, "A new acoustic method for passive sperm whale depth tracking," in *Proc. IEEE Region 10 Conference TENCON 2012*, Nov 2012, pp. 1–5.
- [4] J.-C. Wang, C.-H. Lin, E. Siahaan, B.-W. Chen, and H.-L. Chuang, "Mixed sound event verification on wireless sensor network for home automation," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 803–812, Feb 2014.
- [5] Y. Lee, K. Kim, D. Han, and H. Ko, "Acoustic and visual signal based violence detection system for indoor security application," in *Proc. 2012 IEEE Int. Conf. on Consumer Electronics (ICCE)*, 2012, pp. 737–738.
- [6] S. Astapov, J. S. Preden, T. Aruvalli, and B. Gordon, "Production machinery utilization monitoring based on acoustic and vibration signal analysis," in *Proc. 8th Int. Conf. DAAAM Baltic Industrial Engineering*, 2012, pp. 268–273.
- [7] A. Dhawan, R. Balasubramanian, and V. Vokkarane, "A framework for real-time monitoring of acoustic events using a wireless sensor network," in *Proc. IEEE Int. Conf. Technologies for Homeland Security (HST)*, 2011, pp. 254–261.
- [8] T. Liu, Y. Liu, X. Cui, G. Xu, and D. Qian, "MOLTS: Mobile object localization and tracking system based on wireless sensor networks," in *Proc. IEEE 7th Int. Conf. Networking, Architecture and Storage (NAS)*, 2012, pp. 245–251.
- [9] Z. Merhi, M. Elgamel, and M. Bayoumi, "A lightweight collaborative fault tolerant target localization system for wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1690–1704, 2009.
- [10] G. Vakulya and G. Simon, "Fast adaptive acoustic localization for sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1820–1829, 2011.
- [11] S. Astapov, J.-S. Preden, and J. Berdnikova, "Simplified acoustic localization by linear arrays for wireless sensor networks," in *Proc. 18th Int. Conf. on Digital Signal Processing (DSP)*, 2013, pp. 1–6.
- [12] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [13] Q. Wang, R. Zheng, A. Tirumala, X. Liu, and L. Sha, "Lightning: A hard real-time, fast, and lightweight low-end wireless sensor election protocol for acoustic event localization," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 570–584, 2008.
- [14] E. Mangas and A. Bilas, "FLASH: Fine-grained localization in wireless sensor networks using acoustic sound transmissions and high precision clock synchronization," in *Proc. 29th IEEE Int. Conf. Distributed Computing Systems ICDCS*, 2009, pp. 289–298.
- [15] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP*, vol. 1, 2007, pp. 121–124.
- [16] C. T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2009*, 2009, pp. 2027–2032.
- [17] Y. Liu and Z. Yang, *Location, Localization, and Localizability: Location-awareness Technology for Wireless Networks*. Springer, 2010.
- [18] D. Blatt and A. O. Hero, "Aposc: a rapidly convergent source localization algorithm for sensor networks," in *Proc. IEEE/SP 13th Workshop Statistical Signal Processing*, 2005, pp. 1214–1219.
- [19] Z. Merhi, M. Elgamel, and M. Bayoumi, "Acoustic target localization in sensor networks with FUZZYART," in *Proc. 50th Midwest Symp. Circuits and Systems MWSCAS 2007*, 2007, pp. 1536–1539.
- [20] J. S. Preden, J. Liinas, G. Rogova, R. Pahtma, and L. Motus, "On-line data validation in distributed data fusion," in *SPIE Defense, Security and Sensing, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV*, vol. 8742, 2013, pp. 1–12.
- [21] J. S. Preden, L. Motus, R. Pahtma, and M. Meriste, "Data exchange for shared situation awareness," in *2012 IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, March 2012, pp. 198–201.



Sergei Astapov received his M.Sc. degree in the field of Computer System Engineering at the Tallinn University of Technology in 2011. He continues his education as a PhD student at the Department of Computer Control at the Tallinn University of Technology and is a member of the Department's Research Laboratory for Proactive Technologies. His research interests include object tracking using wide-band signal analysis, classification tasks and distributed computing in embedded multi-agent systems. His recent research concerns object localization and identification in open environments and acoustic signal based diagnostics of industrial machinery.



ProWare and developing

Johannes Ehala received his M.Sc. degree in the field of Computer System Engineering at the Tallinn University of Technology in 2012. Currently he is a PhD student at the Department of Computer Control at the Tallinn University of Technology and is a member of the Research Laboratory for Proactive Technologies. His doctoral studies and research interests include self-organization and emergent behavior in cyber-physical systems and computational models of distributed computer systems. Currently he is involved in analyzing the temporal aspects of a computational model to describe ProWare.



processing, computation

Jürjo-Sören Preden received his PhD degree in Computer Science from the Tallinn University of Technology, the topic being "Enhancing Situation-Awareness, Cognition and Reasoning of Ad-Hoc Network Agents". He is a senior researcher and the head of the Research Laboratory for Proactive Technologies at the Tallinn University of Technology. Jürjo's research interests are focused on distributed computing systems, more specifically on cognition and situation awareness of such systems. His spectrum of activities involves sensing technologies, data and communication in ad-hoc sensing systems.

Publication 4

Reference

S. Astapov, A. Riid, J.-S. Preden, and T. Aruvali, “Industrial process monitoring by multi-channel acoustic signal analysis,” in *Proc. 14th Biennial Baltic Electronics Conference (BEC)*, Tallinn, Estonia, Oct 2014, pp. 209–212.

Abstract

Machinery monitoring at the shop floor bears relevance in preventive maintenance applications and for manufacturing process optimization. As the installation of monitoring hardware directly on the machinery may be hazardous and expensive due to installation costs, the use of contactless sensors is preferable. In this paper we propose a solution for machinery monitoring based on multi-channel acoustic information analysis. We apply large aperture microphone arrays, perform machine noise source localization using the SRP-PHAT method and classify machine acoustical patterns by means of fuzzy rule-based classification. The results of experiments, performed in an industrial setting, indicate the feasibility of our solution in real conditions.

Industrial Process Monitoring by Multi-Channel Acoustic Signal Analysis

Sergei Astapov[#], Andri Riid[#], Jürjo-Sören Preden[#], Tanel Aruväli^{*}

[#]Department of Computer Control, ^{*}Department of Machinery

Tallinn University of Technology

Ehitajate tee 5, 19086, Tallinn, Estonia

sergei.astapov@ttu.ee, andri.riid@ttu.ee, jurgo.preden@ttu.ee, tanel.aruvali@student.ttu.ee

Abstract—Machinery monitoring at the shop floor bears relevance in preventive maintenance applications and for manufacturing process optimization. As the installation of monitoring hardware directly on the machinery may be hazardous and expensive due to installation costs, the use of contactless sensors is preferable. In this paper we propose a solution for machinery monitoring based on multi-channel acoustic information analysis. We apply large aperture microphone arrays, perform machine noise source localization using the SRP-PHAT method and classify machine acoustical patterns by means of fuzzy rule-based classification. The results of experiments, performed in an industrial setting, indicate the feasibility of our solution in real conditions.

Index Terms—Sensor arrays, industrial monitoring, acoustic source localization, pattern classification.

I. INTRODUCTION

Accurate monitoring of industrial machinery is essential in almost any production process. Receiving on-line feedback on equipment utilization ensures uninterrupted flow of the production process. Analysis of machinery utilization also allows for process optimization and thus increases efficiency of the production facility [1]. Monitoring performance and condition of the machinery makes it possible to eliminate faults and malfunctions in equipment prior to breakdown, which decreases the number of accidental failures.

Most high-end industrial devices are equipped with integrated monitoring solutions that enable interconnection and data exchange through a network. Very often, however, a production process is manually operated or equipped with outdated machinery, that lacks integrated sensors and control hardware. Applying a monitoring system to this machinery presents a set of challenges, including integration of hardware. Installation of sensors may prove hazardous and also expensive due to interruptions in the production process. Furthermore, if the machine requires a safe zone or is concealed in a safety chamber, sensor installation may be simply prohibited by safety regulations. One of the solutions lies in the application of contactless, remote sensors. High levels of acoustic noise emitted by industrial machinery suggest the choice of acoustic sensors among others [2], [3].

Our previous work regarding acoustic information analysis applied in industrial machinery monitoring [2] was concerned with single-sensor solutions. High levels of noise originating

from different machines in close proximity in a highly reverberant environment of the shop floor may pollute the signal to a degree, where the classification of the machine's working state becomes unreliable. Using several sensors in an array configuration provides the possibilities of concentrating on a specific noise source and attenuating unwanted noise digitally. In this paper we propose a multi-sensor solution for an acoustic monitoring system.

Microphone arrays in industrial applications are mainly used for acoustic holography, in which machinery or products are examined for areas of high or abnormal noise [4]. These techniques have been applied in a wide range of applications, e.g. diagnostics of vehicle internal and external noise sources [4], estimation of fan and turbine noise [5], detection of compressed air leakage [6], etc. Multi-channel acoustic signal analysis also benefits the optimization of processes, for example laser welding [7]. Research in our particular field of interest, i.e. machinery noise localization at the shop floor and monitoring, has been described in [8] and [9].

II. SYSTEM COMPOSITION AND METHODS

The proposed solution consists of several small sub-arrays that form a large-aperture array with a Field of View (FOV) of 15–25 m² (FOV is a region where the sound source is susceptible to be found by a specific array). The position of a sub-array is specified by the coordinates of its center and the angle α , by which the array is steered from the global zero angle. With each such array, the system monitors a separate machine or part of the conveyer-chain. Several tasks are performed, namely localization of the machine's noise source and signal de-noising, processing, and classification of machine status. We use acoustic localization in the horizontal plane and identify the signal from the monitored machine and/or its specific components. To obtain machine state estimates, we apply beamforming to attenuate unwanted noise, perform feature extraction on the de-noised signal and do classification of the signal features.

A. SRP-PHAT Acoustic Localization

Steered Response Power with Phase Transform (SRP-PHAT) is a popular technique of estimating the DOA of sound signals. The SRP $P(\vec{a})$ is a real-valued functional of a spatial vector \vec{a} , defined by the FOV of an array. High maxima in

$P(\vec{a})$ indicate the source location. $P(\vec{a})$ is computed for each direction as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) value across all pairs of microphones at the theoretical time delays associated with the chosen direction:

$$P(a) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega\tau_{kl}(a)} d\omega, \quad (1)$$

where $X_i(\omega)$ is the spectrum (the Fourier transform) of signal x_i , $X_i^*(\omega)$ is the conjugate of the spectrum and $\tau_{kl}(a) = \tau(a, k) - \tau(a, l)$ is the time delay between signals $x_k(t)$ and $x_l(t)$ of a wave originating from a point $a \in \vec{a}$ and reaching sensors k and l . Ψ_{kl} is the PHAT weight of the inverse of the spectral magnitudes:

$$\Psi_{kl} = \frac{1}{|X_k(\omega) X_l^*(\omega)|}. \quad (2)$$

We compute the SRP for every discrete point of the observed area in order to obtain SRP images of the whole FOV. These images are very informative for manual analysis, however for real-time autonomous processing we propose to use a different approach.

B. Simplified Approach to Localization

The conventional SRP-PHAT performs as many evaluations (1), as there are points in \vec{a} , the number of which is defined by the dimensionality and discretization of the FOV. This analysis is highly resource demanding, particularly when applied to large FOVs. The number of computations is significantly reduced by applying Stochastic Region Contraction (SRC), which iteratively reduces the search region for the global maximum [10]. SRC starts with the initial search region (i.e. the whole FOV), stochastically explores the functional by randomly picking a specific number of points, contracts into a sub-region containing the desired global optimum and proceeds iteratively until the SRP peak is located. To reduce the computational load further we propose to first estimate the initial search region instead of using the whole FOV [11].

Having K microphone arrays, each consisting of M microphones, observing a common FOV, the initial search region reduction is performed in several steps:

- 1) Estimate the DOA for each of K arrays.
- 2) Generate vectors spanning from the array centers to the bounds of the FOV in the directions of DOA.
- 3) Find points of intersections of these vectors.
- 4) Find groups of points no farther than D_{\max} distance units (meters) from their centroid and enclose the areas, in which these groups reside, in rectangles.
- 5) Perform verification and control of false detection.

As we work in the horizontal plane, for DOA it is sufficient to know the azimuth (i.e. angle of arrival) of the incoming signal. The DOA estimation is performed for all $\binom{M}{2}$ combinations of M microphone pairs. Consider a pair m_1 and m_2 . The sound wave originating from a source in the far field (i.e. assuming a planar wave front) is acquired by the pair with a time delay $\tau \in [-\tau_{\max}, \tau_{\max}]$, where τ_{\max} is the delay of sound traveling

directly from one microphone to the other (i.e. at $\pm 90^\circ$). To estimate τ we apply cross-correlation to the two signals:

$$R(\tau) = \sum_{k=0}^{N-1} x_{m_1}(k) \cdot x_{m_2}(k - \tau), \quad (3)$$

where N is the signal frame length in samples. The maximum of the cross-correlation defines the time delay, and the azimuth is obtained by

$$\varphi = \arcsin \frac{\tau \cdot c}{l} = \arcsin \frac{\Delta k / f_s \cdot c}{l}, \quad (4)$$

where c is the speed of sound, l is the distance between two microphones and τ is represented in terms of delay in samples Δk and the sampling frequency f_s . Depending on the chosen pair of microphones in the array, l will vary from l to $l(M-1)$.

Having $C_i \leq \binom{M}{2}$ azimuth estimates for every array, the final DOA ϕ_i for each i -th array, $i \in (1, \dots, K)$, is computed by finding the common direction of azimuth. If no common direction is found, ϕ_i is not assigned, $\phi_i = 0$.

Having estimated $K_1 \leq K$ of the existing array DOA ϕ_{i^*} , $i^* \in (1, \dots, K_1)$ and added the nodes' rotation angles α_i to them, vectors \vec{AB}_{i^*} are computed with the starting point $A_{i^*} = (x_{1,i^*}, y_{1,i^*})$ being the coordinate of i^* -th array's center and the ending point $B_{i^*} = (x_{2,i^*}, y_{2,i^*})$ being the point at a bound of the FOV steered by ϕ_{i^*} from the array's center. Intersection points of all pairs of these vectors are then calculated. As a result we have a set of i^{**} intersections, $i^{**} \in (1, \dots, K_2)$, $K_2 \leq \binom{K_1}{2}$. To get the initial search areas, these intersection points are partitioned by their relative distance. The resulting partitions are strictly speaking clusters \mathbf{P}_j with each point not further than the maximum distance D_{\max} from the cluster's centroid $Cent(\mathbf{P}_j) = 1/|\mathbf{P}_j| \cdot \sum \mathbf{P}_j$. After obtaining the partitions, their areas are enclosed by rectangles with the edges denoted by the partitions minimal and maximal values on both the x - and y -axis.

C. Feature Extraction

Prior to feature extraction the signal is de-noised using Delay and Sum beamforming (DSB). Beamforming is a technique of steering the main lobe of the array's directivity pattern to the desired direction. Given the direction or coordinates of the sound source, the DSB first temporally shifts signals by $\tau = [\tau_1, \dots, \tau_M]$, where $\tau_i = \Delta k_i / f_s$ is the time of sound arrival from the specified source to the microphone i , and then weigh-sums the signals into a single array output. We use equal weights for all microphones, so the DSB is performed as

$$x_{DSB}(k) = \frac{1}{M} \sum_{i=1}^M x_i(k - \Delta k_i). \quad (5)$$

To prepare the de-noised signal for classification, we extract specific signal properties that represent the acoustic pattern of the machine state in a compact way. Below we present only some of the utilized features, a more detailed review may be found in our previous research [12] and in a larger glossary of audio signal features [13].

For every signal frame's absolute magnitude spectrum $|X(k)| = |\mathcal{F}\mathcal{F}\mathcal{T}\{x_{DSB}(k)\}|/N$ of length N , the features are extracted as follows:

Band energy, which measures the energy of the power spectrum in the i -th band and is computed as

$$F_{BE}(i) = \sum_{l \in S_i} |X(l)|^2 / \sum_{k=0}^{N-1} |X(k)|^2, \quad (6)$$

where S_i is the set of power spectrum samples belonging to the i -th band, chosen according to the Mel-scale.

Spectral centroid, calculated as the frequency averaged over the absolute magnitude spectrum:

$$F_{SC} = \sum_{k=0}^{N-1} k \cdot |X(k)| / \sum_{k=0}^{N-1} |X(k)|. \quad (7)$$

Spectral roll-off that determines the frequency, below which a certain amount of spectral energy resides. The amount is determined by a threshold $TH = [0, 1]$:

$$F_{SR} = \arg \max_p \left[\sum_{l=0}^p |X(l)|^2 \leq TH \cdot \sum_{k=0}^{N-1} |X(k)|^2 \right]. \quad (8)$$

Spectral slope, which is a measure of spectral energy decrease in the direction of the higher frequencies. It is specified by the gradient and the y-intersect of a straight line fitted to the magnitude spectrum by linear regression.

The extracted features are concatenated into a feature vector $\mathbf{F} = [F_1, \dots, F_L]$, where L is the number of features. This vector is later used during the classification stage.

D. Fuzzy Rule-Based Classification

Acoustic signals emitted by some industrial machinery in steady working states can be considered stationary and periodic, which makes them classifiable using simple correlation methods [2]. We propose to use a more robust fuzzy rule-based method in order to account for variance in signal features. This brief section describes only the rule base structure and the inference method. The full process of rule base derivation is described in Section III.

A feature vector $\mathbf{F} = [F_1, \dots, F_L]$ is classified by assigning one of R different discrete valued labels to it, using the classifier that consists of R rules:

$$\text{IF } F_1 \text{ is } A_{1r} \text{ AND } \dots \text{ AND } F_L \text{ is } A_{Lr} \\ \text{THEN } \mathbf{F} \text{ belongs to class } c_r, (r = 1, \dots, R), \quad (9)$$

where A_{ir} is the linguistic term of the i -th input (i.e. feature vector element) associated with the r -th rule and c_r is the class label assigned to the r -th rule. This form is the minimal rule classifier (MRC) that contains only one rule for each class (see [12] for details). The class label is assigned in a winner-takes-all manner by specifying the rule with the highest degree of activation

$$c_r = \arg \max_{1 \leq r \leq R} (\tau_r), \quad (10)$$

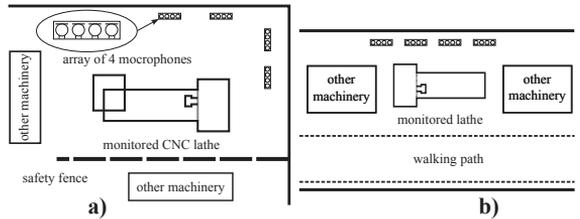


Fig. 1. Microphone sub-array placement at the shop floor.

where τ_r is the activation degree of the r -th rule:

$$\tau_r = \bigcap_{i=1}^L \mu_{ir}(F_i), \quad (11)$$

with μ_{ir} being the Membership Function (MF) corresponding to the linguistic term A_{ir} .

III. EXPERIMENTAL RESULTS

For our experiments we use Vansonic PVM-6052 condenser microphones mounted with a spacing of 15 cm between each other. In total we use 4 sub-arrays with 4 microphones in each sub-array, and place them in a factory workshop in different configurations. For data acquisition Agilent U2354A data acquisition device (DAQ) is used with the sampling rate set to 8 kS/s per channel. Two array configurations are used: linear, where the sub-arrays are placed in a straight line and angular, where two sub-arrays are placed perpendicular to the remaining two. For monitored machinery we choose a Computer Numerical Control (CNC) lathe, and a manually operated lathe. The array installations for the CNC lathe and manual lathe are presented in Fig. 1a and 1b, respectively.

Acoustic localization proves to be successful for every tested machine. Lathe parts that emit most noise are the DC-motor, transmission, spindle and the area of manufactured object. In the motor no-load state (i.e. the transmission is off the spindle) the motor itself is well identifiable, as it is shown in Fig. 2 left; spindle rotation causes a much more diffused, but nevertheless detectable region (see Fig. 2 center). Angular array configuration performs better than the linear one, which is not able to confine the broad noise emitting regions of the monitored machine. Results of the proposed approach to localization coincide with regular SRP-PHAT, as it can be seen in Fig. 2 right compared to Fig. 2 left.

Classification experiments were performed on the CNC and manual lathe. For both machines we specified 3 classes of working state: 1 — Idle, 2 — Motor no-load, 3 — Spindle rotation. The signals were de-noised using DSB and 8 features were extracted (4 band energies, spectral centroid, spectral roll-off and 2 features of spectral slope). For the DSB we used the sound source coordinates, obtained during localization. Two datasets were generated for both machines: a training set for fuzzy rule base derivation and a test set for verification of classification quality.

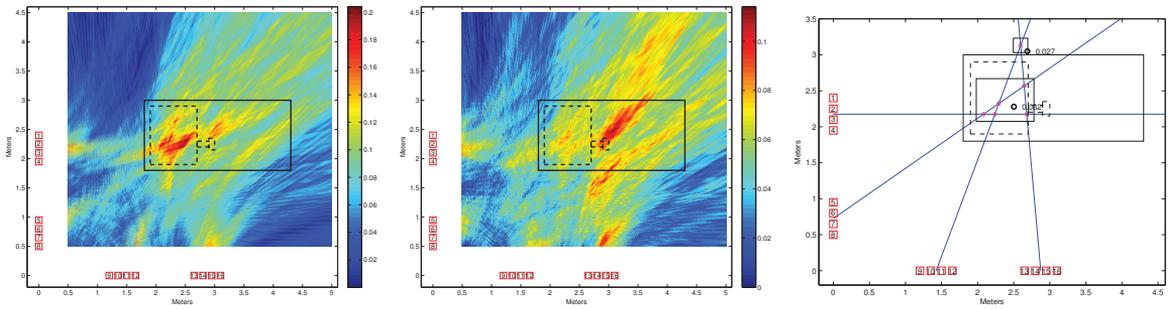


Fig. 2. SRP images of CNC lathe, motor no-load (left) and spindle rotation (center). Proposed localization (right) for the signal instance covered in left plot.

The training and testing datasets for the manual lathe signal contain 165 and 140 samples respectively. Using the training data, we construct a MRC, which is usually not 100% accurate (see Table I). To reduce the error, the number of classification rules is then iteratively increased until the error on training data disappears. At each iteration of this procedure a subset of data corresponding to the rule with the highest number of erroneous samples is split into two further subsets using the Hierarchical Complete Linkage Clustering Algorithm [14]. Upon those two subsets, two new rules are modeled and replace the original rule. As Table I shows, this classifier is 100% accurate on training data, but very inaccurate on test data. In the next step, the number of rules is decreased by rule consolidation in which samples are transferred from minor rules (those governing few samples) to major rules (those governing a lot of samples) as long as classification accuracy does not suffer. The consolidated classifier contains 7 rules, including two singleton rules (governing just one sample), two minor rules and three major rules. We discard the singleton and minor rules (assuming that they represent outliers or irregular data) and the remaining three rules undergo rule compression (that discards less relevant features from individual rules, see [14]). In the last step triangular MFs of the classifier are converted into Gaussian ones to improve generalization ability [15].

The resulting simple classifiers contain less than 8 rules. For example, for the distinction between classes 1 and 2 of the manual lathe, it is sufficient to consider only the y-intersect of spectral slope and class 3 is separated from the rest by the 3rd band energy. The final classification quality for the manual lathe is 94.5% and for the CNC lathe it reaches 90.5%, which is more than satisfying, considering the high surrounding noise and reverberation levels.

REFERENCES

- [1] T. Aruvali, R. Serg, and T. Otto, "Machinery utilization monitoring and pause identification prototype model design," in *Proc. 8th Int. Conf. DAAAM Baltic Industrial Engineering*, 2012, pp. 256–261.
- [2] S. Astapov, J. S. Preden, T. Aruvali, and B. Gordon, "Production machinery utilization monitoring based on acoustic and vibration signal analysis," in *Proc. 8th Int. Conf. DAAAM Baltic Industrial Engineering*, 2012, pp. 268–273.
- [3] M. A. A. Elmaleeh, N. Saad, and M. Awan, "Incipient fault detection of industrial pilot plant machinery via acoustic emission," in *Proc. 8th Int. Symp. Mechatronics and its Applications (ISMA)*, 2012, pp. 1–6.

TABLE I
RESULTS OF FUZZY CLASSIFIER TRAINING STEPS FOR MANUAL LATHE

Procedure	No. of rules	Erroneous samples		Accuracy
		Training	Testing	
Initialization	3	5	22	91.1%
Rule base expansion	13	0	45	85.3%
Consolidation	7	0	23	92.5%
Minor rule removal	3	11	23	88.9%
Compression	3	11	13	92.1%
Conversion	3	6	8	95.4%

- [4] J. Lanslots, F. Deblauwe, and K. Janssens, "Selecting sound source localization techniques for industrial applications," *Sound and Vibration*, vol. 44, no. 6, pp. 6–10, 2010.
- [5] O. Minck, N. Binder, O. Cherrier, L. Lamotte, and V. Pommier-Budinger, "Fan noise analysis using a microphone array," in *Proc. Int. Conf. on Fan Noise, Technology, and Numerical Methods*, Aug. 2012.
- [6] P. Eret and C. Meskell, "Compressed air leak detection using microphone array techniques," in *Leveraging Technology for a Sustainable World*. Springer, 2012, pp. 347–352.
- [7] A. Sansan, L. Zhen, Z. Nan, B. Xianzheng, and L. Yang, "Microphone arrays for acoustic detection during laser welding of cold-rolled steel strip," in *Proc. Int. Conf. on Intelligent System Design and Engineering Application*, vol. 1, 2010, pp. 935–939.
- [8] M. Smith, K. Kim, and D. Thompson, "Noise source identification using microphone arrays," in *Proc. Institute of Acoustics*, vol. 29, Oxford, UK, Oct. 2007.
- [9] B. Atmaja and D. Arifianto, "Machinery fault diagnosis using independent component analysis (ICA) and instantaneous frequency (IF)," in *Proc. Int. Conf. on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, Nov 2009, pp. 1–5.
- [10] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP*, vol. 1, 2007, pp. 121–124.
- [11] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [12] S. Astapov and A. Riid, "A hierarchical algorithm for moving vehicle identification based on acoustic noise analysis," in *Proc. 19th Int. Conf. Mixed Design of Integrated Circuits and Systems*, 2012, pp. 467–472.
- [13] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," CUIDADO I.S.T. Project Report, Tech. Rep., 2004.
- [14] A. Riid and N. Saadallah, "Unsupervised learning of well drilling operations: Fuzzy rule-based approach," in *Proc. 16th IEEE Int. Conf. on Intelligent Engineering Systems (INES)*, June 2012, pp. 375–380.
- [15] A. Riid and E. Rustern, "An integrated approach for the identification of compact, interpretable and accurate fuzzy rule-based classifiers from data," in *Proc. 15th IEEE Int. Conf. Intelligent Engineering Systems (INES)*, 2011, pp. 101–107.

Publication 5

Reference

S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, “Object detection for military surveillance using distributed multimodal smart sensors,” in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP)*, Hong Kong, China, Aug 2014, pp. 366–371.

Abstract

This paper considers an autonomous ground Intelligence, Surveillance and Reconnaissance (ISR) system comprising of multiple distributed, wirelessly communicating smart sensors. The ISR system, in turn, is a part of a larger System of Systems (SoS) consisting of aerial, manned, etc. surveillance systems and information collection centers. The smart sensors of the ISR system perform environment monitoring using different modalities and exchange object detection and identification results to assess the situation and provide other SoS components with this information. In the paper we discuss using acoustic, magnetic and Passive Infrared (PIR) sensor information for target detection and identification. We also propose an approach of distributed acoustic source localization and a method of velocity estimation using PIR data. For sensor communication an asynchronous ad-hoc WSN configuration is proposed. The system is implemented on low power smart sensors utilizing Atmel ATmega128RFA1 processors with integrated 2.4GHz IEEE 802.15.4 compliant radio transceivers.

Object Detection for Military Surveillance Using Distributed Multimodal Smart Sensors

Sergei Astapov, Jürgo-Sören Preden, Johannes Ehala, Andri Riid

Laboratory for Proactive Technologies

Tallinn University of Technology

Ehitajate tee 5, 19086, Tallinn, Estonia

sergei.astapov@ttu.ee, jurgo.preden@ttu.ee, johannes.ehala@ttu.ee, andri.riid@ttu.ee

Abstract—This paper considers an autonomous ground Intelligence, Surveillance and Reconnaissance (ISR) system comprising of multiple distributed, wirelessly communicating smart sensors. The ISR system, in turn, is a part of a larger System of Systems (SoS) consisting of aerial, manned, etc. surveillance systems and information collection centers. The smart sensors of the ISR system perform environment monitoring using different modalities and exchange object detection and identification results to assess the situation and provide other SoS components with this information. In the paper we discuss using acoustic, magnetic and Passive Infrared (PIR) sensor information for target detection and identification. We also propose an approach of distributed acoustic source localization and a method of velocity estimation using PIR data. For sensor communication an asynchronous ad-hoc WSN configuration is proposed. The system is implemented on low power smart sensors utilizing Atmel ATmega128RFA1 processors with integrated 2.4GHz IEEE 802.15.4 compliant radio transceivers.

Index Terms—Multimodal information, acoustic source localization, object detection, distributed computing, smart sensors.

I. INTRODUCTION

Military Intelligence, Surveillance and Reconnaissance (ISR) is steadily heading in the direction of unmanned operation. For ISR large areas of (often ragged) terrain must be constantly and systematically monitored. With increasing availability of smart sensors and devices it is only natural that computer systems become more involved in ISR. Very often human operators in such systems handle raw data incoming from surveillance cameras, reconnaissance drones, motion detectors, etc., and perform situation evaluation themselves. In state of the art systems with higher levels of computer system integration, raw data is processed by the autonomous system and human operators handle the provided situation assessments, which adds agility and greater area coverage.

Furthermore, modern ISR systems are increasingly being assembled from standalone systems, so the resulting ISR systems are Systems of Systems (SoS). Situation evaluation in such systems is performed on-site by every sub-system, rather than in the information hub and system-to-system interaction greatly exceeds machine-to-human interaction. In order to

The work presented in this paper was partially supported by the Estonian Ministry of Defence, the European Defense Agency project IN4STARS and the Nations Support Program for ICT in Higher Education "Tiger University".

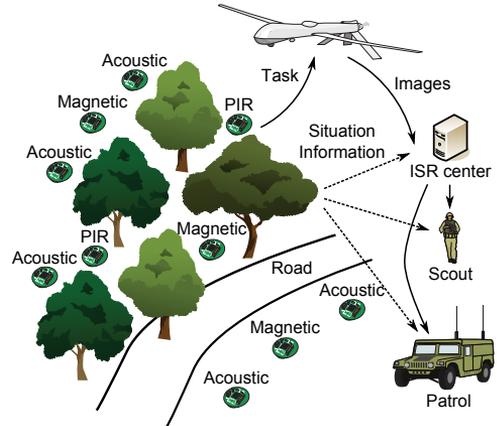


Fig. 1. Concept diagram of the ground ISR system in the SoS.

take full advantage of the capabilities of the ISR SoS, the architecture and the design of these SoS should be different from the approaches currently used in design of ISR systems.

In this work we consider an ISR SoS for ground reconnaissance comprising of multiple distributed smart sensors interconnected in a Wireless Sensor Network (WSN). The smart sensors perform environment monitoring using different modalities and exchange object detection and identification results to achieve situation awareness. The larger SoS comprises of various ground and aerial ISR systems, information collection and fusion databases. Autonomous situation evaluation results are available to human operators in the operation center, as well as to local scout patrols (see Fig. 1). All these systems are considered external information consumers as the focus lies on the ground ISR SoS alone.

This paper presents the ISR system structure, smart sensor operation and communication principals. Smart sensors employ several modalities, which include acoustic, magnetic and Passive Infrared (PIR) sensor information. Acoustic signals are used for both acoustic source localization and identification. For localization a recently developed method of Initial Search Region Reduction (ISRR) is used [1] to manage localization

on low power hardware of smart sensors. Magnetic and PIR sensor readings complement acoustic information in detection of vehicles and foot soldiers. PIR sensor also estimates object movement speed. Local situation assessments are collected for global situation evaluation and sent to external consumers. For efficient and timely data interchange, a WSN communication strategy is proposed. Experimental results of system component tests are presented along the course of the paper.

II. ISR SYSTEM ARCHITECTURE AND OPERATION

The ground ISR system is intended for operation in open (urban, woodland, desert, etc.) environments. The system consists of a large number of smart sensors, known in military applications as Unattended Ground Sensors (UGS). UGS positioning depends on the landscape and may be at all random (see Fig. 2). The position of each UGS is defined by its coordinates (x_0, y_0) and the angle α , by which the sensor is steered from the global angle reference (which may be defined, for example, by the Earth's magnetic field). The coordinates may be estimated by GPS or other location algorithms, e.g. based on Radio Frequency (RF) [2]. Each UGS employs one or several modalities for environment state assessment in its local region. The local metadata is interchanged amongst UGS in order to get a global estimate of the situation in the entire Field of View (FOV) of the ISR system. In our case we use low power devices which cannot simultaneously process acquired data and collect the metadata from other UGS, so in our topology we divide the UGS for local assessments and central UGS for information collection and decision making.

A. Chosen Modalities and WSN Topology

The variety of possible signal types applicable for ground monitoring is quite large. This short paper handles only a few in a specific system configuration. This being stated, we choose acoustic, magnetic and PIR signals for the following reasons. Firstly, the acquisition of these types of signals does not require direct contact with either the monitored object, or the surface, with which that object interacts (as is not the case with e.g. seismic sensors). Secondly, acoustic and infrared waves and magnetic field do not suffer from minor obstacles in the line of view to the monitored object (e.g. thin tree branches) as seriously as, for example, video information. Thirdly, these types of sensors operate equally well during any time of day because they are independent of lighting conditions. The proposed signal processing methods for every sensor type are reviewed in Chapter III.

The system operates in an ad-hoc WSN with an undetermined number of nodes. This way UGS that have been for some reason disabled or restored at certain times do not seriously affect the overall system operation. Therefore, providing a number of redundant UGS introduces additional robustness and reliability to the system. UGS do not exchange raw signal data in order not to overwhelm the network, rather send metadata on local and global situation assessments. Data exchange procedures are discussed further in Chapter IV.



Fig. 2. UGS distribution in real environment (upper left), prototype of acoustic UGS with the microphone amplification circuit (bottom left) and a PIR UGS in a camouflage case designed by Defendec Inc. (right).

B. Implementation Hardware

UGS prototypes are implemented on embedded hardware running TinyOS, equipped with Atmel ATmega128RFA1 microprocessors, which provide an on-chip 10-bit ADC for signal acquisition and a IEEE 802.15.4 compliant radio transceiver for WSN communication. Each UGS is powered by a 3.7 V, 6600 mAh battery block which can sustain the sensors for 1–1.5 years in sleep mode and for approximately a month in operation mode. The sensors are packaged in protective boxes 16 cm in length (see Fig. 2).

Acoustic UGS are equipped with a pair of condenser microphones (Vansonic PVM-6052), sampled at 2 kS/s per channel (due to specific hardware limitations) and processed in frames of 400 samples (frame duration is thus 0.2 s). Magnetic UGS uses a single 3-axis sensor which is sampled at 80 S/s and processed in a sliding window of 10 samples. PIR UGS acquires data by a pyroelectric crystal sensor with detects changes in thermal radiation in zones, defined by a specific Fresnel lens. The lens used in our experiments produces zones for 180° observation angle from the level of the lens and upwards. This design is aimed at reducing false alarms produced by small animals. The slowly oscillating PIR signal is acquired at 100 S/s and is processed in frames of 5 s duration (500 samples) with a window step of half a frame.

III. MULTIMODAL SIGNAL PROCESSING IN WSN

This chapter reviews signal processing procedures for all considered types of UGS. Testing results for every signal type are presented separately to avoid ambiguity.

A. Acoustic Source Localization

For grounded acoustic localization the UGS are placed in the monitored environment in the horizontal plane and localization is performed by estimating the coordinates (x, y) of sound emitting objects. Each UGS is equipped with two acoustic sensors spaced by a specific distance l from one

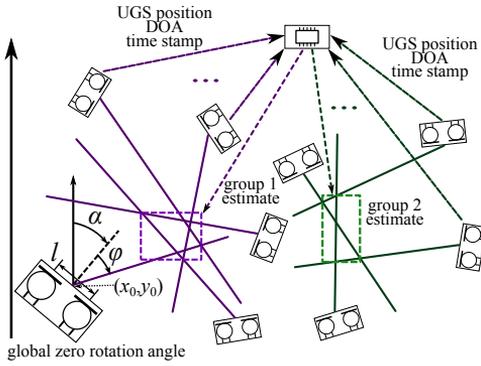


Fig. 3. Acoustic source localization in a network of distributed UGS.

another. Localization is based on estimating the time delays of acoustic wave arrival to the sensors, also called Time Difference of Arrival (TDOA). The Direction of Arrival (DOA) of sound from a specific acoustic source is calculated using TDOA. The network consists of two types of UGS: smart sensors and central nodes. Dual channel smart sensors acquire acoustic information and perform DOA estimation. Nodes that have assumed a central role gather DOA estimates and perform source region estimation. The schematic diagram of system architecture is presented in Fig. 3.

Sensors are partitioned into groups. Each group must have a common Field of View (FOV), i.e. all UGS observe the same area. Group partitioning is performed by clustering, taking two aspects into consideration. Firstly, UGS must be facing in the common direction as the considered localization procedure uses a directional approach. In this regard, the observed area is not necessarily enclosed by UGS, as shown in Fig. 3, but may be observed from one or several sides. Secondly, a group must have certain homogeneity. UGS located too far from the group's centroid may be useless to the localization effort in low Signal to Noise Ratio (SNR) environments or when the sound emitted by the source of interest is too weak. Furthermore, non-homogeneous groups present additional challenges for wireless communication.

Central nodes of the WSN perform UGS grouping during network initialization and later participate in localization. Central nodes must reach an agreement concerning which node governs which UGS group. In this process communication signal strength is taken into account, meaning that a central node will adopt a group, to which it has the strongest connection.

Typical acoustic localization methods, like SRP-PHAT [3], MUSIC [4] or multilateration [5] utilize information from every available sensor. This fact does not pose a problem for wired systems with a single powerful computational hub. In WSN, however, collecting raw signals is a real challenge, especially if the number of nodes is large and signal frames are long. To overcome problems associated with communicating signal frames we apply a simplified localization approach of Initial Search Region Reduction (ISRR), recently developed

by our research team [1]. The main idea behind ISRR lies in maximally confining the region of acoustic source disposition as a preliminary procedure to SRP-PHAT or other localization method [6]. Having already established that SRP-PHAT requires raw information from all sensors in the network, we do not apply it for this specific ISR system configuration. For detection of vehicles or people, localization to a single point is not obligatory and ISRR confined regions serve as a sufficient estimate of object location.

Having a group of K dual sensor UGS, the ISRR is performed in the following steps:

- 1) Estimate the DOA for each of K UGS.
- 2) Generate vectors spanning from sensor pair centers (x_0, y_0) to the FOV bounds in the directions of DOA.
- 3) Find points of intersections of these vectors.
- 4) Find groups of points not farther than D_{max} distance units (meters) from their centroid and enclose the areas, in which these groups reside, in rectangles.
- 5) Perform control of false detection, discard areas not meeting specific criteria (optional).

Step 1 is performed on each UGS, steps 2–5 are performed on the group's central node.

The DOA are estimated for the front view of the sensor pair, i.e. from -90° to 90° . A pair of microphones m_1, m_2 receive a sound wave emitted by a source in the far field (i.e. wave fronts are considered linear) with a time delay τ , which takes the values $\tau \in [-\tau_{max}, \tau_{max}]$, where τ_{max} is the delay of sound traveling directly from one microphone to the other (i.e. at $\pm 90^\circ$). To estimate τ we apply cross-correlation to the two signals:

$$R(\tau) = \sum_{k=0}^n x_{m_1}(k) \cdot x_{m_2}(k - \tau), \quad (1)$$

where n is the signal length in samples. The maximum of the cross-correlation defines the time delay, and the DOA is obtained by

$$\varphi = \arcsin \frac{\tau \cdot c}{l} = \arcsin \frac{\Delta k / f_s \cdot c}{l}, \quad (2)$$

where l is the distance between the microphones, c is the speed of sound in air and τ is represented in terms of delay in samples Δk and the sampling frequency f_s .

At this point DOA validation is performed. If the correlation maximum is less than some threshold, the DOA φ is discarded. This way, in absence of a sound source or in case of high ambient noise, invalid estimates are avoided early on. We use the deviation from the mean for this metric:

$$\max(R(\tau)) > (1 + TH) \cdot \overline{R(\tau)}, \quad (3)$$

where TH is the threshold of deviation, which depends on the SNR in the environment. We use $TH = 0.2$ in our experiments. DOA estimates, which pass control, are sent to central nodes.

A central node receives $K_1 \leq K$ DOA estimates $\phi_{i^*}, i^* \in (1, \dots, K_1)$ and adds UGS rotation angles α_i to them. Vectors AB_{i^*} are computed with the starting point $A_{i^*} = (x_{1,i^*}, y_{1,i^*})$

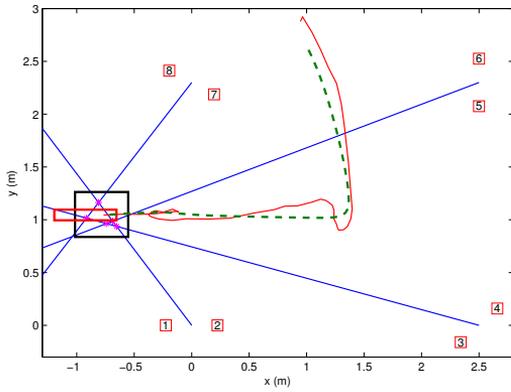


Fig. 4. Localization and tracking of a moving acoustic source using four smart sensors. Blue lines denote DOA vectors of every sensor; magenta stars — vector intersection points; black box — localized region; red line — trajectory estimated by Kalman filter; green dotted line — smoothed trajectory; red box — predicted region of acoustic source.

being the coordinate of i^* -th sensor pair's center and the ending point $B_{i^*} = (x_{2,i^*}, y_{2,i^*})$ being the point at a bound of the FOV steered by ϕ_{i^*} from the pair's center. Intersection points of all pairs of established vectors are then calculated.

As a result we have a set of $I_{i^{**}}$ intersections, $i^{**} \in (1, \dots, K_2)$, $K_2 \leq \binom{K_1}{2}$. To get the initial search areas, these intersection points are partitioned by their relative distance. The resulting partitions are strictly speaking clusters \mathbf{P}_j with each point not further than the maximum distance D_{\max} from the cluster's centroid $Cent(\mathbf{P}_j) = 1/|\mathbf{P}_j| \cdot \sum \mathbf{P}_j$.

After obtaining the partitions, their areas are enclosed by rectangles defined by the lower left and upper right points. As a result several regions may occur in the same FOV. Also while a vector from one array may cross with several other vectors, redundant “echoing” regions may form. These can be removed by applying tracking filters. Localization results for four UGS and a single moving acoustic source are presented in Fig. 4. Here the source position is estimated as a roughly 0.5×0.4 m region, which is well sufficient for ISR application.

The procedure is also applicable to multiple target localization. If more than two sensors are used in the array, several DOA may be estimated. Each dual channel UGS, however, points to a single direction of the strongest acoustic source. As sound pressure decreases exponentially with propagation, each UGS group identifies a source closest to it (if not masked by louder sources). If a group is well spread, several targets may be identified within the FOV based on the same principle.

B. Detection and Speed Estimation Using PIR Sensors

We propose to use PIR data for both movement detection and speed estimation, aimed mainly at foot soldier activity monitoring. The principal block-diagram of the proposed processing procedure is presented in Fig. 5.

The first stage of signal processing consists of raw signal pre-processing (bias removal, scaling, etc.), Low Pass filtering

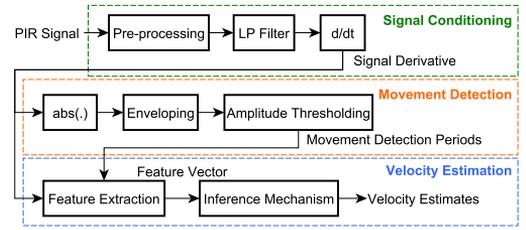


Fig. 5. Diagram of PIR signal processing steps.

(we use 5th order Butterworth IIR filter) and taking the signal derivative. The derivative is calculated as the first order finite difference $\Delta x[k] = x[k] - x[k-1]$, where $x[k]$ and $x[k-1]$ are the k -th and previous signal samples respectively. Further signal analysis is performed on the signal derivative due to the fact that signal settling time after movement is quite long — not less than 1–2 seconds for our specific sensor, and these signal perturbations during settling time are reduced in the derivative. The effect is evident in Fig. 6, where the perturbations on approximately 11th–13th and 20th–22nd seconds are due to movements stopping.

The movement patterns in the PIR signal can easily be contaminated by ambient noise or by insignificant movements. For this reason we perform target movement detection prior to object velocity estimation. Movement detection consists of computing the envelope of the absolute valued signal, and thresholding the envelope (see Fig. 6 middle subplot). Low signal dynamics (causing near zero derivative) suggest to use a sliding average value of the envelope amplitude

$$\bar{x}[k] = \frac{1}{N} \sum_{k-N < i \leq k} x[i], \quad (4)$$

where N is the sliding window length.

For pattern matching we extract a set of temporal features and apply an inference mechanism, which maps the feature vectors onto speed estimates. PIR is a weakly oscillating signal with lengthy perturbations. For this reason frequency analysis and correlation based pattern matching is inapplicable. The features are extracted during periods of movement detection:

- 1) Zero crossing interval (ZCI) — time between two successive signal zero crossings $x_{ZCI} = k_{ZC(i-1)} - k_{ZC(i)}$, which can be considered a rough estimate of half-period. All other features are computed once for every ZCI.
- 2) Amplitude (AMP) — maximum of half-period absolute values $x_{AMP} = \max(|x[k_{ZC(i-1)}]|, \dots, |x[k_{ZC(i)}]|)$.
- 3) Gradient (GR) — value of half-period maximal increase rate $x_{GR} = \max(|\Delta x[k_{ZC(i-1)}]|, \dots, |\Delta x[k_{ZC(i)}]|)$.
- 4) Mean gradient (MGR) — mean value of half-period increase rate $x_{MGR} = \frac{1}{x_{ZCI}} \sum_{j=k_{ZC(i-1)}}^{k_{ZC(i)}} |\Delta x[j]|$.
- 5) Mean square energy (MSE) — mean value of signal half-period squared amplitudes $x_{MSE} = \frac{1}{x_{ZCI}} \sum_{j=k_{ZC(i-1)}}^{k_{ZC(i)}} x^2[j]$.

In our approach we apply fuzzy inference, which provides efficient non-linear mapping for a set of features. For this solution

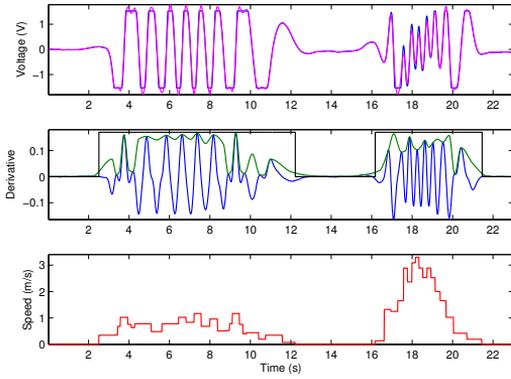


Fig. 6. Walking speed estimation by PIR sensor. Upper: raw PIR signal (blue) and filtered signal (magenta). Middle: signal derivative (blue), its envelope (green), movement detection intervals (black). Lower: speed estimate.

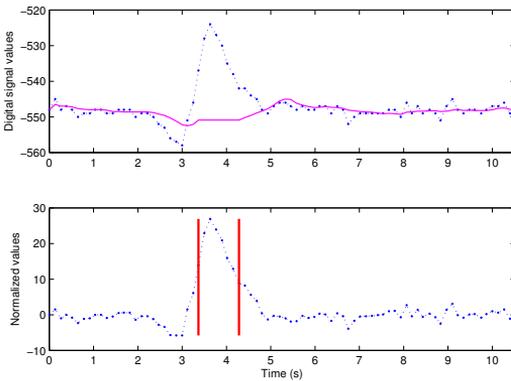


Fig. 7. Vehicle detection by magnetic sensor. Upper: digital signal (dotted blue) and signal average level (magenta). Lower: normalized signal (dotted blue) and interval of detection (between red lines).

we use a Sugeno-type Fuzzy Inference System (FIS), which is composed as an Adaptive Network-based Fuzzy Inference System (ANFIS) and trained using pre-acquired feature data. An example of walking speed estimation is presented in Fig. 6. In the experiment a person went past the sensor with two speeds. It can be seen that the approach distinguishes between moderate walking (app. 1 m/s) and jogging (app. 2.5 m/s) speeds well.

C. Object Detection Using Magnetic Sensors

Applied magnetic sensors operate on the electromagnetic induction principle, i.e. passing of ferromagnetic material in front of the sensor produces voltage fluctuations in the signal. Thus, magnetic sensors are susceptible to vehicles and other personnel carrying things made from ferromagnetic materials. In the proposed system we apply a fairly simple approach to detection based on observing signal deviation.

For every signal sample $x[k]$, the mean amplitude $\bar{x}[k]$ of $N = 10$ previous samples is computed using (4) and

subtracted from the signal to remove the bias. If the absolute value $|x[k] - \bar{x}[k]|$ is above a pre-defined threshold Δ_{\min} , the moment t_0 is fixed and the detection system status goes from state 0 to state 1 (during which the hypothesis whether the deviation is caused by a passing object is verified). When the value $|x[k] - \bar{x}[k]|$ drops below Δ_{\min} , the time t_f is fixed and the system drops to state 0. Note that in state 1 the value of $\bar{x}[k]$ is not updated.

To eliminate false detection due to noise, the lower threshold for detection period t_{\min} is defined. If the considered period $t_f - t_0$ is greater than the threshold $t_f - t_0 > t_{\min}$, detection is considered successful and the values t_0 , t_f are sent to control nodes along with the absolute deviation amplitude $|\max(|x[i]|) - \bar{x}[k]|$, $i \in [t_0, t_f]$.

The choice of Δ_{\min} and t_{\min} determines the sensitivity of the algorithm. Using these parameters the sensor may be tuned to detect certain objects with known properties. The specific values of t_0 and t_f , along with the perturbation amplitude give a rough estimate of object type, e.g. vehicle or armed soldier. We use a 3-axis sensor and each axis is processed separately. Fig. 7 presents a result of vehicle detection with $\Delta_{\min} = 10$ and $t_{\min} = 0.1$ for the x-axis. In the experiment a passenger car drove by 1.5 m from the sensor at 20 km/h. A large Δ_{\min} results in a short but definite detection interval.

D. Object Identification, Tracking and Global Assessment

In the current ISR system configuration object identification is possible through acoustic data analysis. A previously developed multistage algorithm for vehicle identification [7] is applied in acoustic UGS. In [8] we show the ability of the method to operate in real-time on embedded hardware with various frame lengths. For this work we, however, use separate UGS for acoustic identification in order not to interfere with localization. The identification method consists of signal shape tracking, frequency analysis and classification procedures. It is able to distinguish between pre-defined vehicle classes (light versus heavy transport) or identify specific vehicle models using supplied reference patterns.

For tracking we use the discrete Kalman filter (KF) applied to localization results. KF provides the closed form recursive solution for the a linear dynamic system of the form:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_{k-1}\mathbf{x}_k + \mathbf{r}_{k-1} \end{aligned} \quad (5)$$

where \mathbf{x}_k is the system state vector at time step k , \mathbf{y}_k is the measurement vector, \mathbf{A}_{k-1} is the transition matrix of the dynamic model, \mathbf{H}_{k-1} is the measurement matrix, $\mathbf{q}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ is the process noise with covariance \mathbf{Q}_{k-1} and $\mathbf{r}_{k-1} \sim \mathcal{N}(0, \mathbf{R}_{k-1})$ is the measurement noise with covariance \mathbf{R}_{k-1} . KF consists of a prediction step, where the next state of the system is predicted given the previous measurements, and an update step, where the current state is estimated given the measurement at the time. Movement of the acoustic source is described as a discretized Wiener process velocity model [9] with the state vector defined as $\mathbf{x}_k = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]^T$, where (x_k, y_k) is object position and (\dot{x}_k, \dot{y}_k) — the velocity.

Tracking is performed on central nodes of the WSN using the coordinates of acoustically localized regions. For final trajectory estimation the Rauch-Tung-Striebel (RTS) smoother is applied [9]. It removes trajectory irregularities, which appear due to slight errors during localization (see Fig. 4). Additionally we use the KF predictions to determine object position regions for the next localization iteration. This helps to remove redundant regions mentioned in the end of III-A.

Global assessments are made based on acoustic localization and identification, detection results by magnetic and PIR UGS and estimated trajectories. Events are sequenced in time (which is discussed in the next chapter) and estimation of the number of objects, their types and trajectories is performed. While objects remain in the system's FOV, reports on this information are sent to other systems of SoS after every assessment iteration.

IV. NETWORKING AND DATA VALIDATION

Performing computations in a distributed dynamic SoS with wireless communication between the individual systems presents many challenges in terms of guaranteeing data correctness. The data that is used in situation awareness computation must satisfy certain temporal and spatial constraints (i.e. its age must not be greater than a pre-specified value or come from a certain location). It is easy to achieve such guarantees in a system with a fixed configuration, however in a dynamic setting the systems must evaluate these data properties at runtime.

A. Proactive Middleware

Validity checking can be successfully performed with proactive middleware, called ProWare. It relies on the concept of active data mediators that are part of every system to ensure the correctness of data and the resilience of the SoS. ProWare is responsible for ensuring data validity and also mediates the process of finding data providers (in our case local UGS) and bringing them together with data consumers (central UGS). We have shown the viability of this data mediation approach and its ability to ensure temporal and spatial correctness of data in our previous work [10]. Checking temporal constraints of measured data involves estimating the offset of the clocks of data providers and consumers. Our middleware component keeps track of the different clock offsets of system components and regularly checks and updates the change (caused by clock drift, jitter etc.) in these offsets. This temporal information is then used to estimate the time of measurement of the data in local time of the data consumer.

B. Communication and Data Validation

Temporal data validation is very important in our time-critical application. Metadata synchronization plays a major part in global assessment generation. For effective acoustic localization, for example, the DOA calculations ideally must be performed simultaneously. In real conditions a time interval must be specified in which the estimates are considered simultaneous. Due to undefined transmission delays "simultaneous"

data may arrive at different times and not get in this validity interval. Correct data order, on the other hand, is vital for tracking and trajectory estimation. Delayed messages can easily be placed out of order on the receiving side, which will affect the situation assessment.

Our system uses an asynchronous ad-hoc topology of the wireless network. Global time thus is not specified. Each UGS operates in its local time and metadata is broadcasted. Central UGS receive all messages and choose the ones they will use for their procedures (specific localization group, metadata of specific type, etc.). Every metadata message is equipped with a time label specifying data age in milliseconds. Immediately after a signal is acquired, it is given a time label in local time. At the moment of message composition this label is subtracted from the present time reading and the elapsed time label is sent out with the metadata. The receiving UGS then perform transmission time estimation using ProWare and get the data age estimates in their own local time.

V. CONCLUSION

The paper discusses the possibility of using low power sensors in a multimodal ISR system. By using low computational cost signal processing approaches along with distributed computing and special communication schemes, adequate and timely situation assessments can be made. Connecting the sensors through an ad-hoc WSN and using redundant sensors reduces the role of every lone sensor and provides robustness, which is very important in military applications.

REFERENCES

- [1] S. Astapov, J. S. Preden, and J. Berdnikova, "Simplified acoustic localization by linear arrays for wireless sensor networks," in *Proc. Digital Signal Processing, 18th Int. Conf. on*, 2013, pp. 1–6.
- [2] Q. Wang, R. Zheng, A. Tirumala, X. Liu, and L. Sha, "Lightning: A hard real-time, fast, and lightweight low-end wireless sensor election protocol for acoustic event localization," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 5, pp. 570–584, 2008.
- [3] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP*, vol. 1, 2007, pp. 121–124.
- [4] C. T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2009*, 2009, pp. 2027–2032.
- [5] Y. Liu and Z. Yang, *Location, Localization, and Localizability: Location-awareness Technology for Wireless Networks*. Springer, 2010.
- [6] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [7] S. Astapov and A. Riid, "A hierarchical algorithm for moving vehicle identification based on acoustic noise analysis," in *Proc. 19th Int. Conf. Mixed Design of Integrated Circuits and Systems*, 2012, pp. 467–472.
- [8] S. Astapov, J. S. Preden, and E. Suurjaak, "A method of real-time mobile vehicle identification by means of acoustic noise analysis implemented on an embedded device," in *Proc. 13th Biennial Baltic Electronics Conference (BEC)*, Oct 2012, pp. 283–286.
- [9] Y. Bar-Shalom, T. Kirubarajan, and X. R. Li, *Estimation with Applications to Tracking and Navigation*. New York, USA: John Wiley & Sons, Inc., 2002.
- [10] J. S. Preden, J. Llinas, G. Rogova, R. Pahtma, and L. Motus, "On-line data validation in distributed data fusion," in *SPIE Defense, Security and Sensing, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV*, vol. 8742, 2013, pp. 1–12.

Publication 6

Reference

S. Astapov, J. Berdnikova, and J.-S. Preden, “A two-stage approach to 2D DOA estimation for a compact circular microphone array,” in *Proc. Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, Kitakyushu, Japan, Jun 2015, pp. 1–6.

Abstract

Microphone arrays and, specifically, circular arrays have been used for sound source localization and multimedia applications for more than a decade. In recent years the development of compact arrays for implementation in Wireless Sensor Networks (WSN) has risen in popularity. This paper considers a 2D Direction of Arrival (DOA) estimation method for a compact circular array, equipped with additional vertically placed microphones. The proposed method is aimed at reducing the computational cost of DOA estimation for implementation on embedded hardware of WSN smart sensors. The method is compared with a well known localization algorithm of SRP-PHAT and is proven to provide adequate DOA estimates, while being more computationally effective.

A Two-Stage Approach to 2D DOA Estimation for a Compact Circular Microphone Array

Sergei Astapov[#], Julia Berdnikova^{*}, Jürjo-Sören Preden[#]

[#]Laboratory for Proactive Technologies, ^{*}Dept. of Radio and Communication Engineering
Tallinn University of Technology

Ehitajate tee 5, 19086, Tallinn, Estonia

sergei.astapov@ttu.ee, julia.berdnikova@ttu.ee, jurgo.preden@ttu.ee

Abstract—Microphone arrays and, specifically, circular arrays have been used for sound source localization and multimedia applications for more than a decade. In recent years the development of compact arrays for implementation in Wireless Sensor Networks (WSN) has risen in popularity. This paper considers a 2D Direction of Arrival (DOA) estimation method for a compact circular array, equipped with additional vertically placed microphones. The proposed method is aimed at reducing the computational cost of DOA estimation for implementation on embedded hardware of WSN smart sensors. The method is compared with a well known localization algorithm of SRP-PHAT and is proven to provide adequate DOA estimates, while being more computationally effective.

Index Terms—Circular microphone array, Direction of Arrival, SRP-PHAT, Wireless Sensor Networks.

I. INTRODUCTION

Applying an array of microphones for acoustic signal acquisition, instead of a single sensor, provides additional possibilities for signal processing, namely, acoustic source localization and more efficient signal filtering via beamforming. In this lies the reason behind the increasing popularity of multi-sensor approaches to audio processing. Different acoustic array configurations are thoroughly researched and applied to different fields of audio signal processing, the most common being speaker localization for signal quality improvement in multimedia devices. Small microphone arrays are implemented in audio recording devices, such as conference microphones and headsets, which produce clearer speech signals in low Signal to Noise Ratio (SNR) conditions [1].

The application range of acoustic arrays is, however, much broader than just multimedia devices. The increase of productivity of embedded hardware and Analog to Digital Converter (ADC) microsystems paves the way for standalone smart sensors, equipped with acoustic arrays, among other types of sensors. Interconnected through Wireless Sensor Networks (WSN), these smart sensors may perform a large number of different tasks, both civilian and military. In our previous work we have shown the feasibility of a WSN consisting of small and relatively cheap smart sensors performing the tasks of an on-ground Intelligence, Surveillance and Reconnaissance (ISR) system [2]. In this system smart sensors, equipped with linear microphone arrays, are used for the detection and localization of moving enemy ground forces. Civilian applications of our approach include speaker localization and

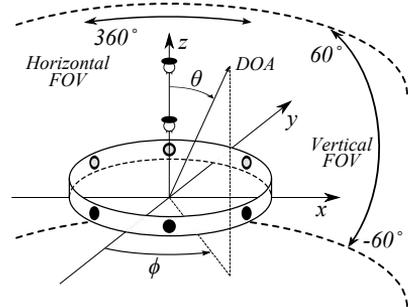


Fig. 1. Configuration of the considered circular microphone array and its desired Field of View.

industrial machinery monitoring [3]. Aside from linear arrays, circular array configuration has been a popular choice for different applications besides speaker localization, ranging from more common ones, like robot sensory systems [4], to human fall detection systems for the elderly [5] and aircraft tracking at airport runways [6]. Circular arrays are appealing because they provide a 360° horizontal Field of View (FOV) with a very simple geometry.

In this paper we consider a circular microphone array, which will be implemented in our WSN systems in the future. This paper discusses a reduced computational cost 2D DOA estimation algorithm for an array configuration presented in Fig. 1. For such a configuration we propose a two-stage approach of azimuth and elevation estimation using cross-correlation in the time domain. For comparison we apply a well known, effective, but computationally expensive localization algorithm of Steered Response Power (SRP-PHAT). Later in the paper we experimentally prove the ability of our approach to provide adequate DOA estimates, while consuming significantly less computational power than SRP-PHAT.

II. ARRAY CONFIGURATION AND APPLICATION FIELD

The task of 2D DOA search consists of estimating the horizontal Angle of Arrival (AOA), or azimuth ϕ , and the polar AOA, or elevation θ , of incoming acoustic waves. The microphone array configuration is designed to meet the desired FOV for localization of on-ground objects. The targeted application

field assumes localization of slow-moving sources (no more than 5 m/s) at a distance of less than 10 m from the array in both indoor and open outdoor environments. The array consists of M_h microphones, equidistantly mounted in a horizontal circular shell, and M_v microphones, placed vertically upwards from the center of the circular shell ($M_h = 6$, $M_v = 2$ in Fig. 1). The flat bottom then allows for mounting the array on different surfaces. All the microphones used in the implementation are omnidirectional, however, the shell mounting reduces the directivity of horizontal microphones. In a closed indoor environment this helps reducing the susceptibility to reverberation. In an open environment, where noise sources may arise at any given time and their number is unpredictable, the microphones, facing the sound source of interest, are less affected by noise sources arising in other directions.

The array must also be compact for our applications (well integrable and not obstructing movement). For the experimental prototype we use a circular shell with the radius of $r = 0.075$ m. For a uniform array the angle between two successive microphones $\alpha = \angle m_i^h Om_{i+1}^h$, relative to the array center O , is

$$\alpha = \angle m_i^h Om_{i+1}^h = \frac{2\pi}{M_h}, \quad (1 \leq i < M_h). \quad (1)$$

In the case of six microphones $\alpha = \frac{\pi}{3}$. The two vertical microphones m_1^v , m_2^v are set at distances of 0.1 and 0.2 m from the horizontal plane, respectively, to meet the desired vertical FOV of $[-60^\circ, 60^\circ]$.

III. DOA ESTIMATION WITH SRP-PHAT

Steered Response Power with Phase Transform (SRP-PHAT) is a method of DOA estimation for acoustic signals. The SRP $P(\vec{a})$ is a real-valued functional of a spatial vector \vec{a} , defined by the FOV of a specific array. The maxima in $P(\vec{a})$ indicate the estimates of sound source location. $P(\vec{a})$ is computed for each direction in \vec{a} as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) across all pairs of microphones at the theoretical time delays, associated with the chosen direction. Consider a pair of signals $x_k(t)$, $x_l(t)$ of an arbitrary array, consisting of M microphones. The times of soundwave arrival from point a to the two microphones are $\tau(a, k)$ and $\tau(a, l)$, respectively. Hence the time delay between the two signals is $\tau_{kl}(a) = \tau(a, k) - \tau(a, l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega\tau_{kl}(a)} d\omega, \quad (2)$$

where $X_i(\omega)$ is the spectrum (i.e. the Fourier transform) of signal x_i , and $X_i^*(\omega)$ is the conjugate of that spectrum. Ψ_{kl} is the PHAT weight of the inverse of spectral magnitudes:

$$\Psi_{kl} = \frac{1}{|X_k(\omega) X_l^*(\omega)|}. \quad (3)$$

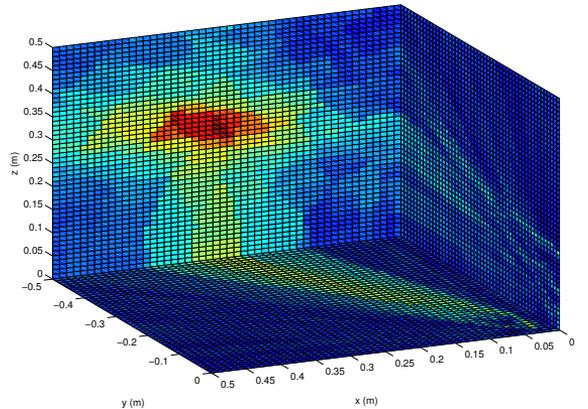


Fig. 2. A volumetric SRP image of an acoustic source with DOA angles equal to $\phi = 300^\circ$ and $\theta = 30^\circ$. Red region indicates the direction to source.

A. Exhaustive DOA Search with SRP-PHAT

SRP-PHAT is a more general tool for acoustic localization and DOA estimation [7]. It is typically used to estimate the SRP values over the whole area or volume of the FOV. This enables localization of multiple sound sources, while also reducing reverberation and masking effects. In this approach the FOV volume is discretized into a spatial grid, with a SRP value computed for every point. Though the approach is most comprehensive, the immense number of computations it requires makes the process extremely slow and demanding.

Consider the example presented in Fig. 2. The figure portrays an SRP image of a part of the considered FOV — the fourth quadrant of the horizontal plane with array center at $(0, 0, 0)$. This volumetric part of the FOV, measuring 0.5 m in length, depth and height, is discretized with a distance space of 0.01 m along every axis. Thus the volume vector \vec{a} consists of $0.5^3/0.01^3 = 125000$ discrete directions, and so the evaluation (2) must be performed the according number of times. The procedure clearly visually indicates the DOA of the acoustic source, however, it presents a challenge for autonomous search of elevated SRP values. There exist numerous propositions of simplifying the SRP-PHAT exhaustive search procedure, e.g. applying a stochastic search for the SRP maxima in the functional [8], or contracting the initial search volume [3], previously proposed by us. In this paper we reduce the number of computations by choosing a reduced functional.

B. Search with SRP-PHAT over a Reduced Functional

To reduce the SRP-PHAT functional we propose to perform horizontal and vertical DOA estimation separately. So, the horizontal plane is divided into n_h and the vertical plane — into n_v possible AOA angles, respectively. A single angle is calculated, similarly to (1), as $\phi_h = \frac{2\pi}{n_h}$ and $\theta_v = \frac{\pi}{n_v}$. The points are chosen in the volumetric FOV along a spherical surface with radius r_{FOV} . The horizontal SRP-PHAT evaluation is performed over the entire circumference $[0, 2\pi)$ for the

points $a_{h,i} = (x_{h,i}, y_{h,i}, 0)$:

$$\begin{aligned} x_{h,i} &= r_{FOV} \cos(i\phi_h), \quad (0 \leq i < n_h), \\ y_{h,i} &= r_{FOV} \sin(i\phi_h), \quad (0 \leq i < n_h). \end{aligned} \quad (4)$$

The azimuth is estimated in the directions of elevated SRP values. For a single source case it is equal to

$$\phi = \frac{i}{\max(P(\bar{a}_h))} \phi_h. \quad (5)$$

Then the vertical SRP-PHAT evaluation is performed over the vertical half-circumference from the positive z -axis downward, i.e. $[0, \pi]$, in the direction of established azimuth ϕ for the points $a_{v,i} = (x_{v,i}, y_{v,i}, z_{v,i})$:

$$\begin{aligned} x_{v,i} &= r_{FOV} \cos(\phi) \sin(i\theta_v), \quad (0 \leq i \leq n_v), \\ y_{v,i} &= r_{FOV} \sin(\phi) \sin(i\theta_v), \quad (0 \leq i \leq n_v), \\ z_{v,i} &= r_{FOV} \cos(i\theta_v), \quad (0 \leq i \leq n_v). \end{aligned} \quad (6)$$

The elevation angle is estimated in the direction of elevated SRP, and also brought to a more comprehensive interval from the positive z -axis downward $[\frac{\pi}{2}, -\frac{\pi}{2}]$:

$$\theta = \frac{\pi}{2} - \frac{i}{\max(P(\bar{a}_v))} \theta_v. \quad (7)$$

IV. PROPOSED APPROACH TO DOA ESTIMATION

Even with a reduced functional, SRP-PHAT requires significant resources and processing time because it operates in the frequency domain and employs cross-correlation between all pairs of microphones. Taking into account the Doppler Effect, signal processing in the frequency domain is advantageous for moving source localization [9]. For our applications we consider velocities, which produce minimal to no frequency shifts. So, we focus on reducing the burden of the Fourier Transform by operating in the time domain, while also reducing the number of microphone pairs for cross-correlation.

Here we review our two-stage approach for a circular array with M_h horizontal and M_v vertical microphones, discussed in Section II. As the horizontal microphones take a directional DOA estimation approach, the pairs of microphones for azimuth estimation are chosen such that their inter-sensor angle is less than $\frac{\pi}{2}$: $\alpha_{ij} = \angle m_i^h Om_j^h < \frac{\pi}{2}$. Let us denote the set of these pairs as

$$A_h = \left\{ (m_i^h, m_j^h) \subseteq S_2^{M_h} \mid \alpha_{ij} < \frac{\pi}{2} \right\}, \quad (8)$$

where $S_2^{M_h}$ is the set of all combinations of horizontal microphone pairs, $|S_2^{M_h}| = \binom{M_h}{2}$. Through a procedure, discussed in Section IV-A, a separate azimuth estimate $\hat{\phi}_{ij}$ is computed for every pair $(m_i^h, m_j^h) \subseteq A_h$, and a final estimate ϕ is made.

For the elevation angle estimation a set of horizontal microphones, situated within the half-circumference of direction ϕ , is chosen: $A_{act} = \left\{ m_i^h \mid -\frac{\pi}{2} \leq \alpha_{\phi, m_i^h} < \frac{\pi}{2} \right\}$, where $\alpha_{\phi, m_i^h} = \phi - \alpha_{m_i^h}$ is the angle between the azimuth estimate and the microphone position. A_{act} here denotes the set of so-called ‘‘active’’ horizontal microphones. The pairs participating in elevation estimation consist of: pairs between every active

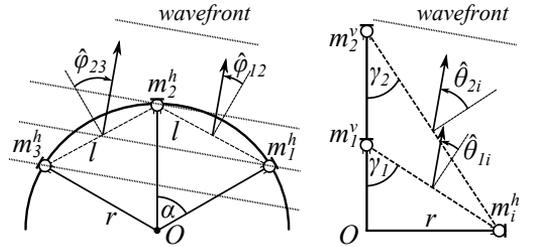


Fig. 3. Horizontal (left) and vertical (right) angle of arrival estimation for a far field acoustic source.

microphone and every vertical microphone; every pair of vertical microphones. Lets denote this set as

$$A_v = \left\{ (m_i^h, m_j^v) \mid m_i^h \in A_{act}, j = [1, M_v]; S_2^{M_v} \right\}, \quad (9)$$

where $S_2^{M_v}$ is the set of all combinations of vertical microphone pairs, $|S_2^{M_v}| = \binom{M_v}{2}$. A separate elevation estimate $\hat{\theta}_{ij}$ is computed for every pair $(m_i^*, m_j^*) \subseteq A_v$, and a final estimate θ is made, as discussed in Section IV-B.

A. Azimuth Angle of Arrival Estimation

For azimuth estimation we assume the far field disposition of the acoustic source (i.e. the spherical wavefront is spread enough to be considered planar). As it was previously stated, the initial azimuth estimates are made for every pair of horizontal microphones $(m_i^h, m_j^h) \subseteq A_h$. Consider Fig. 3 (left). For any pair $(m_i^h, m_j^h) \subseteq A_h$ of consecutive microphones, the azimuth estimate can be obtained by

$$\hat{\phi}_{ij} = \sin^{-1} \left(\frac{\tau_{ij} \cdot c}{l} \right) = \sin^{-1} \left(\frac{\Delta k_{ij} / f_s \cdot c}{l} \right), \quad (10)$$

where c is the speed of sound, l is the distance between two consecutive microphones, calculated as

$$l = 2r \sin \left(\frac{\alpha}{2} \right) = 2r \sin \left(\frac{\pi}{M_h} \right), \quad (11)$$

and τ_{ij} is the Time Difference of Arrival (TDOA) of the wavefront to microphones m_i^h and m_j^h . For non-consecutive microphones, l is calculated by substituting α in (11) with its multiple. In (10), τ_{ij} is also represented in terms of delay in samples Δk_{ij} and the sampling frequency f_s . Note, that the TDOA is always limited to $\tau \in [-\tau_{max}, \tau_{max}]$, where $\tau_{max} = l/c$ is the delay of sound traveling directly from one microphone to the other (i.e. at $\pm \frac{\pi}{2}$). To estimate τ_{ij} we apply cross-correlation to the pair of signals:

$$R(\tau_{ij}) = \sum_{k=0}^{n-1} x_{m_i^h}(k) \cdot x_{m_j^h}(k - \tau_{ij}), \quad (i < j), \quad (12)$$

where n is the length of the signals in samples. The maximum of the cross-correlation $\max(R(\tau_{ij}))$ then defines the TDOA. The quality of the estimate is measured as cross-correlation peak distinctness from its mean level:

$$q_{ij} = \max(R(\tau_{ij})) - \text{mean}(R(\tau_{ij})). \quad (13)$$

Each estimate $\hat{\varphi}_{ij}$ is made for the middle point if the inter-microphone distance and takes the values of $\hat{\varphi}_{ij} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, taking negative values if the source is left, positive — if the source is right, and zero — if the source is in front of the pair. Thus, $\hat{\varphi}_{ij}$ are adjusted to array angle coordinates: $\hat{\varphi}_{ij}^* = \hat{\varphi}_{ij} + ((i-1)\alpha + (j-1)\alpha)/2$. Next, coherent directions are found among the estimates. This is done by applying a partitioning procedure, similar to the one we presented in [3]. The procedure performs the task of clustering the $\hat{\varphi}_{ij}^*$ estimates such, that the coherent estimates must lie within a sector with a central angle no more than φ_{\max} . For example, if $\varphi_{\max} = \frac{\pi}{6}$, then each cluster's coherent estimates must lie no more than $[-\frac{\pi}{12}, \frac{\pi}{12}]$ from the cluster's centroid.

The resulting clusters Φ_p , $p = 1, \dots, P$, where P is the number of clusters, each contain n_p estimates $\hat{\varphi}_k$, $k = [1, n_p]$, and the associated quality q_k . The clusters are evaluated in order to find the final azimuth estimates. Algorithm 1 handles the final azimuth calculation for the single source case. There $\sigma = (0, 1)$ is the lower threshold of tolerance. The final azimuth estimate ϕ cannot be made if there are insufficient coherent estimates or if they are of low quality.

Algorithm 1 Final azimuth ϕ estimation for a single source

Require: Φ_p , $q_k \sim \hat{\varphi}_k \in \Phi_p$, $p = 1, \dots, P$
 get largest cluster size $|\Phi|_{\max}$, maximum quality q_{\max}
if $|\Phi|_{\max} = 1$ **or** $q_{\max} < \text{allowed}$ **then**
 return $\phi \leftarrow \emptyset$
else if Φ_p of size $|\Phi|_{\max}$ contains $\hat{\varphi}_k$ with q_{\max} **then**
 return $\phi \leftarrow \sum_{k=1}^{n_p} q_k \hat{\varphi}_k / \sum_{k=1}^{n_p} q_k \triangleright$ weighted mean
else search in smaller Φ_p , $n_p > 1$ for $q_k \geq \sigma \cdot q_{\max}$
 case found: **return** $\phi \leftarrow \sum_{k=1}^{n_p} q_k \hat{\varphi}_k / \sum_{k=1}^{n_p} q_k$
 case not found: **return** $\phi \leftarrow \emptyset$
end if

B. Elevation Angle of Arrival Estimation

Elevation estimation is performed once the azimuth is set. As the horizontal microphones are not in line with the wavefront, incoming at ϕ , the signals of active microphones must be shifted to meet the TDOA at ϕ . In computing this shift, we maintain the far field assumption, however, its inconsistency produces negligible error. Both the shifting procedure and estimation of near field error are addressed in the Appendix.

After the signals $x_{m_i^h \in A_{act}}$ have been shifted to meet the ϕ AOA, elevation estimates $\hat{\theta}_{ji}$ for all microphone pairs in A_v are made using (10) and (12). The pairs of horizontal and vertical microphones, as portrayed in Fig. 3 (right), are set at angles γ_j , $j = [1, M_v]$, relative to the vertical plane. Thus, the initial estimates are steered by

$$\hat{\theta}_{ji}^* = \gamma_j - \hat{\theta}_{ji} = \tan^{-1} \left(\frac{r}{z_{m_j^v}} \right) - \hat{\theta}_{ji}, \quad (14)$$

where $z_{m_j^v}$ is the distance from O to m_j^v on the z -axis. Final elevation AOA estimation is performed by applying the clustering procedure, addressed in the previous section, and Algorithm 1 for clusters Φ_p and sought-for angle θ .

The array implementation parameters are: $r = 0.075$ m, $M_h = 6$, $\alpha = \frac{\pi}{3}$, $M_v = 2$, $z_{m_j^v} = \{0.1, 0.2\}$ m. We use Vansonic PVM-6052 condenser microphones, connected through a custom-built amplification board to an Agilent U2354A data acquisition device, which samples the signals at $f_s = 48$ kS/s per channel. The data is acquired and processed in the Matlab environment using the Data Acquisition Toolbox. The processing is performed frame-by-frame with a step of 0.1 seconds for both the proposed method and SRP-PHAT.

The experiments for single source DOA estimation are performed indoors under SNR conditions of app. 20 dB. A loudspeaker, reproducing human speech, is placed at certain angles within a distance of 1–2 m from the array. Here we consider four experiments. For the first two the speaker remains at $\phi = 300^\circ$ (i.e. in the direction of the 6-th microphone) and is lifted and lowered to test θ estimation accuracy. The speaker either remains at a certain θ for 3–4 seconds (Experiment I), or moves constantly (II). For the other two experiments the speaker is moved around the array, varying in both ϕ and θ . Here the speaker is also carried either fixing the position with steps (III), or moving constantly (IV).

For the SRP-PHAT approach, reviewed in Section III-B, we choose $r_{FOV} = 0.5$ m and $n_h = 500$, $n_v = 250$, which results in accuracy $\phi_h = \theta_v = \frac{\pi}{250} \simeq 0.72^\circ$. The approach thus requires 750 evaluations (2), as opposed to many thousands in case of exhaustive search. The result of a single frame evaluation is presented in Fig. 4. Both global maxima of $\phi = 300^\circ$ and $\theta = 30^\circ$ are distinctive among local surplus peaks of SRP, likely arising due to reverberation. The elevated θ levels at app. -10° are noteworthy, apparently arising due to sound reflections from the floor surface, as studied in [10].

The proposed approach utilizes 19 pairs of microphones, according to the definitions of A_h and A_v in (8), (9): 6 consecutive pairs $m_i^h m_{i+1}^h$ ((1, 2), (2, 3), etc.); 6 pairs over one microphone $m_i^h m_{i+2}^h$ ((1, 3), (2, 4), etc.); 6 pairs between two vertical m_j^v and three m_i^h ; 1 pair $m_i^v m_{i+1}^v$. Thus the number of pairs is less than for the SRP-PHAT case, which utilizes $\binom{8}{2} = 28$ pairs. This difference will be more evident, if applied to a larger number of microphones. The clustering parameters are set to $\varphi_{\max} = 30^\circ$, $\theta_{\max} = 10^\circ$, $\sigma = 0.6$. The result of DOA estimation with the proposed method for a single frame is presented in Fig. 5. The initial azimuth estimates strongly resemble the SRP peak distribution pattern of Fig. 4, indicating a similar reaction to reverberation. The sought-for DOA is also correctly established.

The DOA estimates of both methods for signal segments of Experiments I and IV are presented in Fig. 6 and Fig. 7, respectively. Static position of a highly elevated source seems to disrupt ϕ estimation for both methods, which is evident in the interval of $\phi = [50^\circ, 60^\circ]$ in Fig. 6. During constant movement this static effect is not noticed, however, certain irregularities do arise during rapid movement, e.g. during the interval of 3–6 s in Fig. 7. Generally, the proposed method is not inferior to SRP-PHAT in the considered FOV,

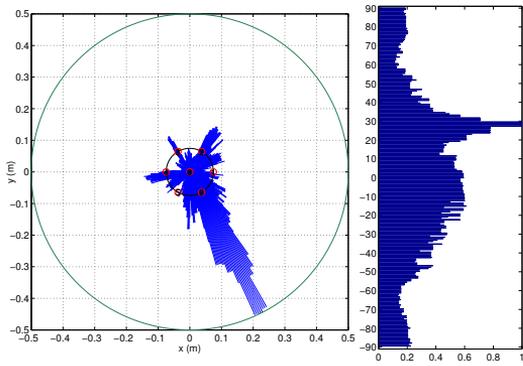


Fig. 4. Result of SRP-PHAT over reduced functional for a frame of Experiment I. Azimuth estimates (left) are normalized to FOV radius. Elevation estimates normalized to interval $[0, 1]$.

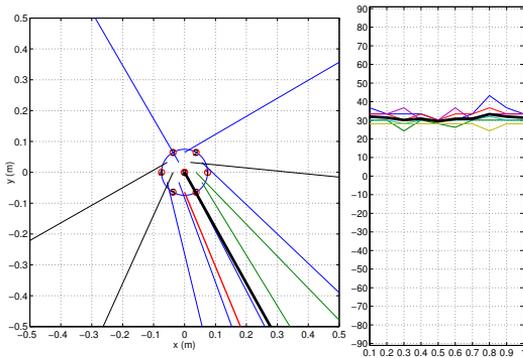


Fig. 5. Result of the proposed method for a frame of Experiment I. Azimuth estimates (left) are colored by quality from worst to best: black, blue, green, red. Thick black line — final estimate. Elevation estimates (right) presented for 10 successive frames. Thick black line — final estimate.

while requiring less processing time. Table I presents the processing time (mean over 10 runs) and DOA estimate Root Mean Square Error (RMSE) for the four experiments. As the processing time is measured in Matlab, running on a PC, it is by no means an adequate measure of computation speed on embedded hardware, however, it shows that the proposed method operates more than 3 times faster than SRP-PHAT over the reduced functional. The RMSE is calculated as a difference between the estimates, produced by SRP-PHAT, and the proposed method: as the testing was performed manually, the error in speaker placement cannot be adequately accounted for. The RMSE remains in reasonable bounds for all experiments, if considering close-range localization.

VI. DISCUSSION AND FUTURE WORK

SRP-PHAT is used in this paper simply for comparison with the proposed method. The handled reduced functional, however, may be reduced further by applying, for example, non-uniform spatial sampling [11], or stochastic search, proposed in [8]. This way, in a WSN of multiple arrays, different DOA

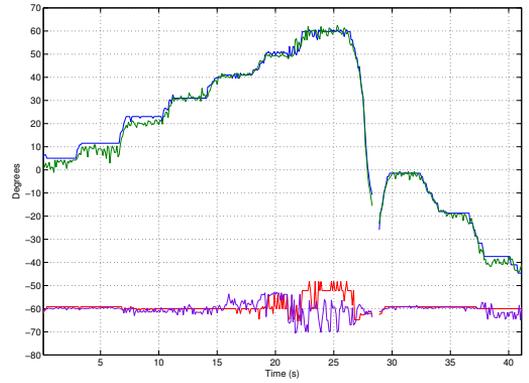


Fig. 6. DOA estimates for Experiment I. Red — azimuth estimated by SRP-PHAT, purple — azimuth estimated by proposed method, blue — elevation estimated by SRP-PHAT, green — elevation estimated by proposed method.

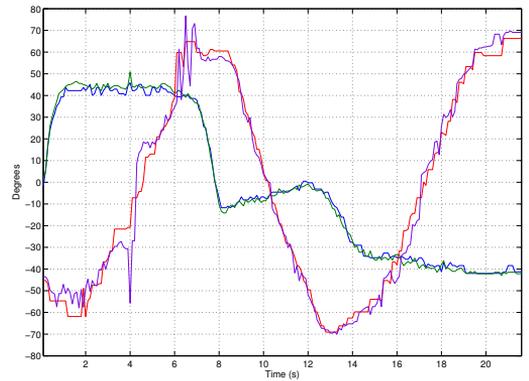


Fig. 7. DOA estimates for Experiment IV. Red — azimuth estimated by SRP-PHAT, purple — azimuth estimated by proposed method, blue — elevation estimated by SRP-PHAT, green — elevation estimated by proposed method.

methods may be applied for mutual reassurance. The MUSIC algorithm is also a good candidate for comparison with our method. In [12], for example, MUSIC is applied to an array of similar configuration with promising results.

The problem of multiple source localization must be thoroughly addressed in the future. Both SRP-PHAT and our method imply multiple DOA estimation, however, the spatial limitations and various acoustic effects must be examined, similarly to [13], [14], in order to make reliable conclusions.

TABLE I
EXPERIMENTAL SIGNAL PROCESSING TIMES AND RMSE OF DOA

Experiment	Length (s)	t. proc. (s) SRP-PHAT	t. proc. (s) proposed	RMSE ϕ (deg)	RMSE θ (deg)
I	70.0	283.45	84.25	4.51	2.29
II	62.0	248.77	74.63	5.22	1.86
III	86.0	343.30	103.68	9.43	2.84
IV	52.0	210.97	62.88	13.67	1.99

TABLE II
INFLUENCE OF NEAR FIELD ERROR AT DIFFERENT d AND f_s

$c = 343$ (m/s), $r = 0.075$ (m), $\alpha_{\phi, m_i^h} = \frac{\pi}{2}$				
d (m)	f_s (S/s)	$R - d$ (cm)	Δ_{nf} (cm)	Δk_{nf} (S)
0.05	48000	9.51	2.06	3
0.05	32000	9.51	2.06	2
0.05	24000	9.51	2.06	1
0.1	48000	8.97	1.52	2
0.1	32000	8.97	1.52	1
0.1	24000	8.97	1.52	1
0.2	48000	8.44	0.99	1
0.7	48000	7.81	0.36	1

VII. CONCLUSION

The proposed method for 2D DOA estimation is shown to produce results, similar to the ones of SRP-PHAT, while requiring less than a third of the processing time. The final estimates are very similar to the results of SRP-PHAT for acoustic source localization in the considered FOV. Further research is required to estimate the effectiveness of the approach for multiple source DOA estimation and the ability to operate in real-time on embedded hardware.

APPENDIX

SIGNAL SHIFTING AND INFLUENCE OF NEAR FIELD ERROR

As for elevation estimation the azimuth estimate is fixed, the signal of the active horizontal microphones in A_{act} must be shifted to meet with the ϕ AOA. Consider Fig. 8. The far field planar acoustic wave reaches microphone m_i^h in the horizontal plane with a spatial delay h , which is defined as

$$h = a \cos \beta = 2r \sin \left(\frac{\alpha_{\phi, m_i^h}}{2} \right) \cos \left(\frac{\pi - \alpha_{\phi, m_i^h}}{2} \right). \quad (15)$$

The signal of microphone m_i^h must be thus left-shifted for $\Delta k_f = \left\lfloor h \frac{f_s}{c} \right\rfloor$ samples.

Let us now review the case, where the far field assumption is not met. The spatial delay between a far field planar and a near field spherical acoustic wave is defined in Fig. 8 as Δ_{nf} . Assuming a near field source, situated at distance d from the point of ϕ AOA, the microphone m_i^h will receive the wave from the distance R with a spatial delay of $R - d$. While dealing with a spherical wavefront, $h + \Delta_{nf} = R - d$, and thus $\Delta_{nf} = R - d - h$, where R is defined as

$$R = \sqrt{d^2 + a^2 - 2ad \cos(\pi - \beta)}. \quad (16)$$

In turn, the delay error in samples is

$$\Delta k_{nf} = \left\lfloor (R - d - h) \frac{f_s}{c} \right\rfloor = f \left(r, \alpha_{\phi, m_i^h}, d, c, f_s \right). \quad (17)$$

Now let us evaluate the influence of d and f_s on the near field error with the fixed: sound speed, r (by array geometry) and α_{ϕ, m_i^h} (by requirement of belonging to A_{act}). Table II presents several cases of error at different d and f_s . For our array geometry the error produced by a sound source at a distance larger than 0.1 m is negligible, even at $f_s = 48$ kS/s.

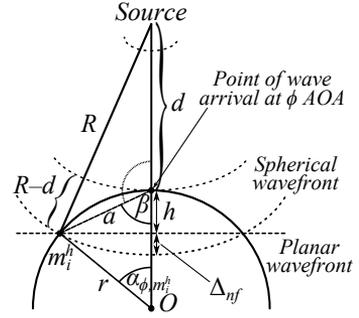


Fig. 8. Signal delay estimation for the near and far field assumption.

REFERENCES

- [1] C. Zhang, D. Florencio, D. Ba, and Z. Zhang, "Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 538–548, April 2008.
- [2] S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, "Object detection for military surveillance using distributed multimodal smart sensors," in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP 2014)*, Aug 2014, pp. 366–371.
- [3] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [4] Y. Tamai, S. Kagami, Y. Amemiya, Y. Sasaki, H. Mizoguchi, and T. Takano, "Circular microphone array for robot's audition," in *Sensors*, 2004. *Proc. of IEEE*, vol. 2, Oct 2004, pp. 565–570.
- [5] Y. Li, Z. Zeng, M. Popescu, and K. Ho, "Acoustic fall detection using a circular microphone array," in *Proc. 2010 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2010, pp. 2242–2245.
- [6] I. Hafizovic, C. Nilsen, and M. Kjolerbakken, "Acoustic tracking of aircraft using a circular microphone array sensor," in *Proc. IEEE Int. Symp. on Phased Array Systems and Technology (ARRAY)*, Oct 2010, pp. 1025–1032.
- [7] J. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Brown University, Providence, RI, USA, 2000.
- [8] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP*, vol. 1, 2007, pp. 121–124.
- [9] D. Lindgren, G. Hendeby, and F. Gustafsson, "Distributed localization using acoustic Doppler," *Signal Processing*, vol. 107, pp. 43–53, 2015.
- [10] Y. Song, S. Rahardja, and R. Yu, "Beamforming performance of circular microphone array on spherical platform near bottom boundary," in *Proc. 2010 IEEE Int. Conf. on Multimedia and Expo (ICME)*, July 2010, pp. 743–747.
- [11] M. Kleider, B. Rafaely, B. Weiss, and E. Bachmat, "Golden-ratio sampling for scanning circular microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2091–2098, Nov 2010.
- [12] Y. Albagory and A. Ashour, "MUSIC 2D-DOA estimation using split vertical linear and circular arrays," *Int. J. of Computer Network and Information Security (IJCNIS)*, vol. 5, no. 8, p. 12, 2013.
- [13] F. Ribeiro, D. Florencio, D. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1449–1460, July 2012.
- [14] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2193–2206, Oct 2013.

Publication 7

Reference

S. Astapov, J. Berdnikova, J. Ehala, and J. S. Preden, “Shooter localization by networked multichannel acoustic ground sensors,” in *Proc. Int. Conf. on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, Oct 2015, pp. 332–337.

Abstract

Gunshot acoustic localization for military and urban security systems has long been an important topic of research. In recent years the development of independent Unmanned Ground Sensors (UGS), interconnected through Wireless Sensor Networks (WSN), performing distributed cooperative localization, has grown in popularity. This paper proposes an asynchronous method of gunshot localization, performed by UGS, equipped with circular microphone arrays. Each UGS in the WSN estimates the Direction of Arrival (DOA) of acoustic events and the time delay between these events. Fusion nodes perform event identification, accounting for outliers (e.g. target hit noise), and shooter localization, based on gathered event information and WSN geometry. The approach is tested on real signals, acquired at a shooting range, and succeeds in localizing the shooter’s position with a mean accuracy of 0.87 meters for 30 shots.

Shooter Localization by Networked Multichannel Acoustic Ground Sensors

Sergei Astapov[#], Julia Berdnikova^{*}, Johannes Ehala[#], Jürjo-Sören Preden[#]

[#]Laboratory for Proactive Technologies, ^{*}Dept. of Radio and Communication Engineering
Tallinn University of Technology
Ehitajate tee 5, 19086, Tallinn, Estonia
sergei.astapov@ttu.ee, julia.berdnikova@ttu.ee, johannes.ehala@ttu.ee, jurgo.preden@ttu.ee

Abstract—Gunshot acoustic localization for military and urban security systems has long been an important topic of research. In recent years the development of independent Unmanned Ground Sensors (UGS), interconnected through Wireless Sensor Networks (WSN), performing distributed cooperative localization, has grown in popularity. This paper proposes an asynchronous method of gunshot localization, performed by UGS, equipped with circular microphone arrays. Each UGS in the WSN estimates the Direction of Arrival (DOA) of acoustic events and the time delay between these events. Fusion nodes perform event identification, accounting for outliers (e.g. target hit noise), and shooter localization, based on gathered event information and WSN geometry. The approach is tested on real signals, acquired at a shooting range, and succeeds in localizing the shooter’s position with a mean accuracy of 0.87 meters for 30 shots.

I. INTRODUCTION

Active development of shooter acoustic localization systems has continued for more than three decades. Numerous different gunshot detection and direction estimation systems are currently available for military applications of sniper and covert enemy force positioning, and are also used in law enforcement for gun violence reduction and forensics [1]. The devices currently available are generally standalone systems, composed of a single microphone array, e.g. the vehicle-mountable Boomerang system [2]. Individual gunshot detectors, developed for military and law enforcement personnel [3], consist of compact shoulder-carried, helmet or uniform mounted sensors. Such individual systems increase local situation awareness, however, for large area coverage a different approach is required.

Modern Military Intelligence, Surveillance and Reconnaissance (ISR) systems apply distributed Unmanned Ground Sensors (UGS) interconnected through a Wireless Sensor Network (WSN) for large area coverage. UGS perform local situation assessment, and through data fusion a global assessment over the whole monitored area is made. A distributed system configuration expands UGS collective Field of View (FOV) and thus is well suitable for shooter localization. The state of the art in this area suggests either synchronous [4], or asynchronous [5] gunshot acoustic event detection and subsequent shooter localization based on UGS collective information. The majority of the proposed approaches work with single-sensor UGS, computing the gunshot event parameters (e.g. time delays between gunshot events) under different initial assumptions, e.g. the caliber of the fired projectile [4], or a certain ballistic shockwave acoustic model [6]. However, initial assumption

inconsistency and the presence of acoustic event outliers may significantly reduce localization accuracy [7].

Employing multichannel smart sensors for gunshot localization allows to additionally estimate the Direction of Arrival (DOA) of gunshot event acoustic waves. In this paper we propose a method of shooter localization based on gunshot event DOA and Time Difference of Arrival (TDOA) information. In our approach each UGS performs gunshot event detection, computes the DOA and fixates event occurrence time in its own local time. The fusion node gathers DOA and time information from all UGS, performs gunshot event identification, calculates the TDOA between events of interest and estimates the shooter position, based on the known UGS positions. The TDOA are calculated per each UGS and no cross-UGS delays are used, thus node synchronization is not required (however, node clock divergence still needs to be roughly estimated for the fusion node to be able to distinguish between shot instances).

The proposed method is tested on signals, acquired by four circular microphone array UGS during a live shooting exercise with a shooter-target distance of 35 m. For DOA estimation we apply a reduced computational cost approach, presented by us in [8], and the well-known method of Steered Response Power (SRP-PHAT) for comparison. The experimental results indicate the feasibility of the proposed localization method in terms of gunshot event detection, outlier elimination and shooter position estimation.

II. GUNSHOT ACOUSTIC COMPONENTS

A gunshot is characterized by the shockwave (SW), produced by a supersonic projectile, and the muzzle blast (MB) of the fired weapon. Fig. 1 portrays the acoustic events, produced by a gunshot at point Z , as observed at point O . SW produces a conical wavefront at an angle θ to the bullet’s trajectory. The angle θ depends on the speed of sound in air c and the bullet velocity v :

$$\theta = \sin^{-1} \frac{c}{v}. \quad (1)$$

The initial bullet velocity is equal to the muzzle velocity v_0 , which depends on the bullet caliber and cartridge type, and decreases with distance. The waves of MB, on the other hand, propagate spherically at c in all directions. As c depends on the ambient temperature, for an open environment it is calculated as

$$c = 331.45 \sqrt{1 + t^\circ/273}, \quad (2)$$

where t° is the temperature in Celsius.

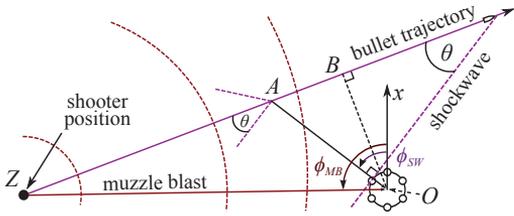


Fig. 1. Gunshot acoustic event geometry.

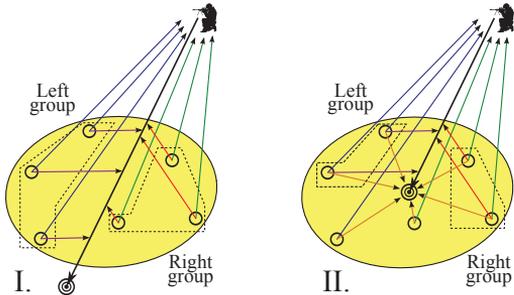


Fig. 2. Two considered shot scenarios. Bullet passes the UGS cluster (I); bullet hits the target in the vicinity of the UGS cluster (II).

At line-of-sight, the sensor at O detects MB at the time

$$t_{MB} = t_{shot} + \frac{d_{O,Z}}{c}, \quad (3)$$

where t_{shot} is the time of shot, and $d_{O,Z} = \|O - Z\|$ is the Euclidean distance between points O and Z . Acoustic waves of SW originate from the bullet itself and not from the muzzle. As the bullet has reached point A (see Fig. 1), the SW wave propagates at c , at an angle θ , and reaches point O at the time

$$t_{SW} = t_{shot} + \frac{d_{Z,A}}{v} + \frac{d_{A,O}}{c}. \quad (4)$$

The TDOA between SW and MB acoustic events can then be expressed as

$$\Delta t = t_{MB} - t_{SW} = \frac{d_{O,Z}}{c} - \frac{d_{Z,A}}{v} - \frac{d_{A,O}}{c}. \quad (5)$$

The DOA of MB and SW are defined in the horizontal plane as azimuth values ϕ_{MB} , ϕ_{SW} , relative to the sensor's local coordinate system (x -axis in Fig. 1).

III. PROBLEM STATEMENT

Knowing t_{SW} and t_{MB} , gunshot acoustic localization may be performed by estimating the angle θ and the miss distance (i.e. the distance from the sensor to the bullet's trajectory, or $d_{O,B}$ in Fig. 1). Angle θ may be estimated by applying a shockwave acoustic model to the duration of SW signal perturbation [6], or calculated under known bullet caliber assumption [4]. Then, using multiple measurements from K UGS, the miss distance and point Z are estimated via an optimization procedure. If UGS clocks are sufficiently synchronized, a mutual reference moment t_{shot} can be established for all UGS via (4), and Z — estimated by multilateration, using time delays t_{MB} from (3).

Unfortunately, if gunshot events include outliers, such as reflections and target hit (TH) noise, MB cannot be unambiguously selected from numerous events, following SW. Consider, for example, Fig. 2. The majority of state of the art approaches assumes Scenario I, where the bullet passes the UGS cluster, and no TH is detected. Outliers can be eliminated by identifying MB and SW by their acoustic properties [9], however, that does not solve the masking problem. We assume Scenario II of Fig. 2, where a target is hit inside the UGS cluster (which is more likely, if UGS surround the potential target). In this case, firstly, either SW, or MB may be corrupted or masked by TH, and, secondly, UGS situated behind the target may not detect SW at all. Acoustic event detection also poses a problem. At a sufficient range the TDOA between SW and MB acoustic pulses makes the events well distinguishable [10]. We rather study a close range case, where event separation is not straightforward due to short TDOA.

IV. PROPOSED APPROACH TO GUNSHOT LOCALIZATION

The proposed approach is intended for application in WSN with a dynamic ad-hoc topology. This implies node synchronization complications and a varying number of active nodes at any given time. Thus we focus on an asynchronous, size-invariant solution. The approach consists of the following steps:

- 1) Each UGS detects a shot, separates gunshot events, marks the time and computes a DOA value per event.
- 2) Per each detected shot, each UGS sends a packet to the fusion node, containing its position, steering angle and acoustic event parameters $\{\mathbf{x}, \beta, t, \Phi\}$.
- 3) Fusion node performs event identification and shooter localization, based on the information of active UGS.

The packet of UGS $k = 1, \dots, K$ contains: UGS coordinates $\mathbf{x}_k = (x_k, y_k)$; UGS steering angle β_k ; gunshot event times $\mathbf{t}_k = [t_1, \dots, t_{EV_k}]$; event DOA $\Phi_k = [\phi_1, \dots, \phi_{EV_k}]$, where EV_k is the number of detected events of k -th UGS. As each UGS operates in its own coordinate system, the steering angle β_k is used to specify UGS local coordinate system steering from a global zero-rotation angle (which can be defined by Earth's magnetic north).

While receiving packets from UGS, the fusion node maintains a validity interval, beginning at the moment of arrival of the first packet. This way the expired packets, or the ones corresponding to another shot, are dealt with separately. Data validation is discussed in our previous works [11] and will not be discussed in this paper along with routing and other network-related problems.

For sensor configuration we choose Uniform Circular Arrays (UCA) because they provide full horizontal FOV with a simple geometry. Each array consists of $M = 6$ microphones with an angle between two successive microphones $\alpha = \angle m_i O m_{i+1}$, relative to the array center O , of

$$\alpha = \angle m_i O m_{i+1} = \frac{2\pi}{M}, \quad (1 \leq i < M). \quad (6)$$

As the UGS need to be covert, if hidden in the monitored environment, the array must also be compact. For the UCA experimental prototype we use a circular shell with the radius of $r = 7.5$ cm.

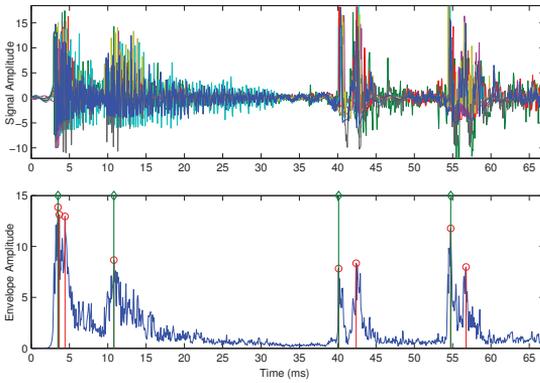


Fig. 3. Gunshot acoustic components acquired by UCA at 48 kS/s (top). Collective envelope and times of detected events (bottom). Red stems — results of peak detection; green stems — event establishing peaks.

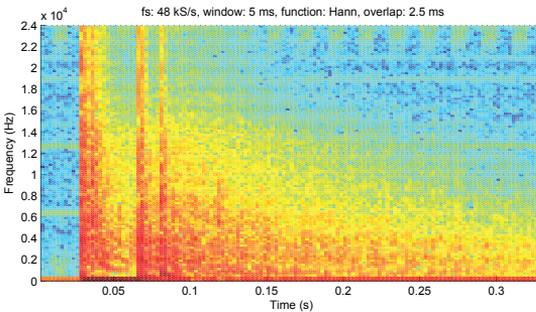


Fig. 4. Spectrogram of the acquired gunshot signal. Acoustic components, presented in Fig. 3, are located at approximately 30–95 ms.

A. Gunshot Event Detection and Separation

Acoustic event detection at close range poses a problem due to an extremely short TDOA between SW and MB [12]. Fig. 3 presents an example of a gunshot acquired at 16 m from the shooter. Here the TDOA between SW (at 4 ms) and MB (at 11 ms) is 7 ms. Neglecting post-blast events (starting at the 40th ms) may seriously harm the detection process. For example, an AK-47 can fire 600 rounds per minute and an M-4 — at 950 rpm, which constitutes approximately 1 bullet in 100 ms and 63.2 ms, respectively. In this case consecutive SW and MB may be mistaken for post-blast events or outliers, and vice versa for a single shot case.

We establish all acoustic events by the following procedure. First, a collective envelope is computed using the signals from all microphones. At sampling time n , the envelope of samples $x_1[n], \dots, x_M[n]$ is $s_{env}[n] = \max(|x_1[n]|, \dots, |x_M[n]|)$. For reliable thresholding and peak detection, the envelope's low-frequency components are filtered out, which does not affect gunshot signatures, as they possess highly uniform spectral densities (see Fig. 4). Peaks within an interval of $t_W/2$ seconds are grouped together and one (the first) peak per event is chosen. An example of separation of four different events is presented in Fig. 3 (lower). One frame of duration t_W is retrieved from the multichannel signal buffer for each event peak such, that event beginning is included in the frame and adjacent events are strictly excluded.

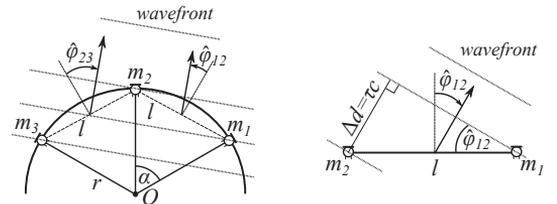


Fig. 5. Azimuth estimation in the far field for consecutive microphones of the circular array (left). Geometry of a single microphone pair (right).

B. Direction of Arrival Estimation

At the time of shot detection, k -th UGS produces EV_k frames of length $N = f_s t_W$, where f_s is the sampling frequency. A separate DOA estimate is then computed per each frame by applying SRP-PHAT (for reference) and our proposed lightweight method [8].

1) *SRP-PHAT*: To reduce the number of SRP-PHAT computations we divide the horizontal plane into n_h possible azimuth angles. A single angle increment is calculated, similarly to (6), as $\phi_h = \frac{2\pi}{n_h}$. The evaluation points are chosen in the planar FOV along a circle with a radius r_{FOV} . The SRP-PHAT evaluation is performed over the entire circumference $[0, 2\pi)$ for the points $a_{h,i} = (x_{h,i}, y_{h,i})$:

$$\begin{aligned} x_{h,i} &= r_{FOV} \cos(i\phi_h), \quad (0 \leq i < n_h), \\ y_{h,i} &= r_{FOV} \sin(i\phi_h), \quad (0 \leq i < n_h). \end{aligned} \quad (7)$$

The azimuth is estimated in the direction of elevated SRP values $P(\vec{a}_h)$. For a single source case the final azimuth is equal to $\phi = \underset{\max(P(\vec{a}_h))}{\hat{i}} \cdot \phi_h$.

2) *Proposed*: The method takes a directional DOA estimation approach. The pairs of microphones for azimuth estimation are chosen such, that their inter-sensor angle is less than $\frac{\pi}{2}$: $\alpha_{ij} = \angle m_i O m_j < \frac{\pi}{2}$. The set of these pairs is

$$A = \left\{ (m_i, m_j) \subseteq S_2^M \mid \alpha_{ij} < \frac{\pi}{2} \right\}, \quad (8)$$

where S_2^M is the set of all combinations of microphone pairs, $|S_2^M| = \binom{M}{2}$. A separate azimuth estimate $\hat{\phi}_{ij}$ is made under the far field assumption for every pair of microphones $(m_i, m_j) \subseteq A$. For any pair (m_i, m_j) of consecutive microphones (see Fig. 5), the azimuth estimate can be obtained by

$$\hat{\phi}_{ij} = \sin^{-1} \left(\frac{\tau_{ij} \cdot c}{l} \right) = \sin^{-1} \left(\frac{\Delta n_{ij} / f_s \cdot c}{l} \right), \quad (9)$$

where l is the distance between two consecutive microphones, calculated as

$$l = 2r \sin \left(\frac{\alpha}{2} \right) = 2r \sin \left(\frac{\pi}{M} \right), \quad (10)$$

τ_{ij} is the TDOA of the wavefront between microphones m_i and m_j , and Δn_{ij} is the delay in samples at f_s . In case of non-consecutive microphones, l is calculated by substituting α in (10) with its multiple. To estimate τ_{ij} , we apply cross-correlation to the pair of signals:

$$R(\tau_{ij}) = \sum_{n=0}^{N-1} x_{m_i}(n) \cdot x_{m_j}(n - \tau_{ij}), \quad (i < j), \quad (11)$$

and locate the index of its maximal value.

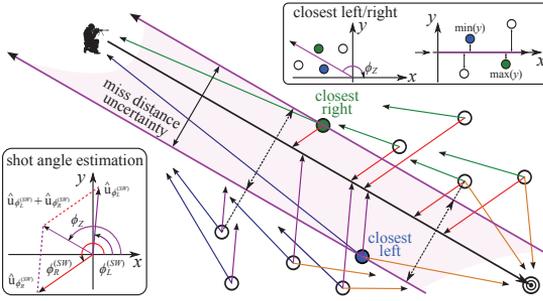


Fig. 6. Shot angle and miss distance uncertainty interval estimation by UGS groups, situated to the left and to the right from the bullet's trajectory.

Initial estimates $\hat{\varphi}_{ij}$ are adjusted to the array's coordinate system: $\hat{\varphi}_{ij}^* = \hat{\varphi}_{ij} + ((i-1)\alpha + (j-1)\alpha)/2$, and coherent directions are found among these estimates by applying a clustering procedure, presented in [13]. The clusters are evaluated to identify the largest cluster that contains the estimates, which gave the sharpest correlation peaks. The final azimuth estimate ϕ is computed as the weighted mean of estimates belonging to that cluster [8].

C. Information Fusion and Shooter Localization

As a result of shot detection, the fusion node receives K packets $\{\mathbf{x}, \beta, \mathbf{t}, \Phi\}_k$, $k = 1, \dots, K$, where K is the number of active UGS, which have detected at least one gunshot event. The number of detected events EV_k may vary per UGS. The DOA estimates Φ_k are first steered to the global coordinate system, $\Phi_k = \Phi_k - \beta_k$, and information fusion is then conducted in the following steps.

1) *DOA Coherency*: Consensual DOA are established by analyzing all $\Phi = \{\Phi_k \mid k = 1, \dots, K\}$ estimates. Angular values in Φ are clustered in such a manner, that the coherent estimates must lie within a sector with a central angle no more than φ_{\max} . For example, if we use $\varphi_{\max} = \frac{\pi}{9}$, then each cluster's coherent estimates must not lie farther than $[-\frac{\pi}{18}, \frac{\pi}{18}]$ from the cluster's centroid. If coherent estimates exist, we obtain P clusters Φ_p , $p = 1, \dots, P$, each containing n_p estimates ϕ_i , $i = [1, n_p]$.

Assuming Scenario II (Fig. 2), $\bar{\Phi}_p$ will contain SW DOA, corresponding to the UGS situated to the left and to the right from the bullet's trajectory, different MB DOA, and outliers, like TH DOA and various reflection noise. The DOA of SW vary only slightly (due to DOA estimation error and natural variation of angle θ) and do not depend on the distance to shooter; MB DOA, on the other hand, depend on the distance to shooter and UGS cluster dimensions. If the distance to shooter is significantly larger than the width of the UGS cluster, MB DOA will be roughly parallel for all UGS. At a closer distance the UGS situated on the opposite sides of the bullet's trajectory will have their MB DOA significantly skewed towards the trajectory in the shooter's direction.

2) *Event Identification and Grouping*: To reduce the error of individual DOA estimates, event identification is performed on $\bar{\Phi}_p$ clusters' mean values: $\bar{\Phi}_p = \frac{1}{n_p} \sum \Phi_p$, $p = 1, \dots, P$. SW DOA are identified under the assumptions that SW events are detected first, and at least one SW DOA was detected to the left and to the right from the trajectory.

$\bar{\Phi}_p$ is searched for $\bar{\phi}_i, \bar{\phi}_j$, $i = [1, P-1]$, $j = [i+1, P]$, that meet the conditions

$$\begin{aligned} \frac{\pi}{2} - \varphi_{\max}^{(SW)} &< \frac{\bar{\phi}_i + \bar{\phi}_j}{2} < \frac{\pi}{2} + \varphi_{\max}^{(SW)}, \\ \forall \text{ind}_{t_k}(\phi_k \mid \phi_k \in \Phi_i) &= 1, \\ \forall \text{ind}_{t_k}(\phi_k \mid \phi_k \in \Phi_j) &= 1. \end{aligned} \quad (12)$$

Here $\varphi_{\max}^{(SW)}$ is the upper tolerance of SW DOA variation. If the conditions are met, $\bar{\phi}_i, \bar{\phi}_j$ and $\phi_k \in \Phi_i \cup \Phi_j$ are labeled $\bar{\phi}_i^{(SW)}, \bar{\phi}_j^{(SW)}$ and $\phi_k^{(SW)}$, respectively. For $\bar{\phi}_i^{(SW)}, \bar{\phi}_j^{(SW)}$ condition (12) also implies that they were measured on the opposite sides of the bullet's trajectory. So, the shot angle ϕ_Z is derived as the angular component of the sum of their corresponding unit vectors $\hat{\mathbf{u}}_{\bar{\phi}_i} + \hat{\mathbf{u}}_{\bar{\phi}_j}$ (see Fig. 6).

Having estimated ϕ_Z , the UGS S_k that have detected SW are placed either into the "left", or "right" groups G_L, G_R :

$$\begin{aligned} \phi_k^{(SW)} < \phi_Z &\Rightarrow S_k \in G_L, \\ \phi_k^{(SW)} > \phi_Z &\Rightarrow S_k \in G_R. \end{aligned} \quad (13)$$

To estimate the miss distance, $S_k \in G_L \cup G_R$ closest to the bullet's trajectory are first located. This is done by steering the S_k coordinates \mathbf{x}_k by ϕ_Z towards the x -axis around the UGS overall spatial centroid $\bar{\mathbf{x}} = \frac{1}{K} \sum \mathbf{x}_k$ as

$$\begin{pmatrix} x'_k \\ y'_k \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} + \begin{pmatrix} \cos(\phi_Z) & \sin(\phi_Z) \\ -\sin(\phi_Z) & \cos(\phi_Z) \end{pmatrix} \begin{pmatrix} x_k - \bar{x} \\ y_k - \bar{y} \end{pmatrix}. \quad (14)$$

Then, as portrayed in Fig. 6, "closest left" and "closest right" UGS \bar{S}_L, \bar{S}_R are defined as

$$\begin{aligned} \bar{S}_L &= S_i, \quad i = \text{ind min}(y'_k), \quad S_k \in G_L, \\ \bar{S}_R &= S_i, \quad i = \text{ind max}(y'_k), \quad S_k \in G_R, \end{aligned} \quad (15)$$

and the distance between them, perpendicular to the shot angle, $\phi_Z - \frac{\pi}{2}$, is referred to as the miss distance uncertainty interval. Inside this interval the exact miss distance cannot yet be estimated at this point. We approximate it at a later stage of shooter localization.

The DOA, corresponding to MB, are searched for in $\bar{\Phi}_p$ for such $\bar{\phi}_i$, $i = [1, P]$, that abide

$$|\phi_Z - \bar{\phi}_i| < \varphi_{\max}^{(MB)}, \quad \bar{\phi}_i \neq \bar{\phi}_i^{(SW)}, \quad (16)$$

where preference is given to $S_k \in G_L \cup G_R$, because SW detection implies that the bullet has passed the UGS and thus TH DOA will likely not resemble MB. This way TH DOA will most certainly be avoided. Noise outliers, on the other hand, are seldom acquired at similar DOA by a significant number of UGS, and thus their corresponding clusters $\bar{\Phi}_p$ are significantly smaller and estimates — more dispersed. At this stage they are easily separable from the estimates, considered for the MB label. (Incidental acoustic sources, arising in the FOV, can be identified and excluded from analysis by general acoustic monitoring and source tracking techniques, e.g. [13].) After outlier separation, $\phi_k \in \Phi_i$ meeting condition (16) are labeled $\phi_k^{(MB)}$.

3) *Distance to Shooter and Localization*: Having identified $\phi_k^{(SW)}$ and $\phi_k^{(MB)}$, $k = 1, \dots, K$, where K is now the number of UGS with both detected events, it is possible to accurately compute the TDOA Δt_k as

$$\begin{aligned} \Delta t_k &= t_{k,i} - t_{k,j}, \\ i &= \text{ind}_{\Phi_k}(\phi_k^{(MB)}), \quad j = \text{ind}_{\Phi_k}(\phi_k^{(SW)}). \end{aligned} \quad (17)$$

Based on Δt_k and the miss distance estimates $\hat{d}_{miss}^{(k)}$, it is possible to estimate the distance to shooter from the k -th UGS using a closed form solution, proposed in [4]:

$$d_{S_k,Z} = \frac{1}{2(c^4 - v^4)} (A - 2\sqrt{B}), \quad (18)$$

where

$$\begin{aligned} A &= -2v^3 \hat{d}_{miss}^{(k)} \sqrt{v^2 + c^2} - 2\Delta t_k c^3 v^2 \\ &\quad + 2c^2 \hat{d}_{miss}^{(k)} v \sqrt{v^2 + c^2} - 2\Delta t_k c v^4, \\ B &= -2c^4 v^4 \left(\hat{d}_{miss}^{(k)} \right)^2 + 2(\Delta t_k)^2 c^6 v^4 \\ &\quad + 2(\Delta t_k)^2 c^4 v^6 - 2c^7 \hat{d}_{miss}^{(k)} \Delta t_k v \sqrt{v^2 + c^2} + c^8 (\Delta t_k)^2 v^2 \\ &\quad + 2c^8 \left(\hat{d}_{miss}^{(k)} \right)^2 + 2v^5 \hat{d}_{miss}^{(k)} \sqrt{v^2 + c^2} \Delta t_k c^3. \end{aligned}$$

Projectile velocity can be empirically estimated by inverting equation (1) as $\hat{v} = c/\sin(\hat{\theta})$ and applying it to $\hat{\theta}$, which is computed as $\hat{\theta} = \bar{\phi}_L^{(SW)} - (\pi - \phi_Z)$, where $\bar{\phi}_L^{(SW)}$ is the mean value of cluster of estimates, labeled as SW and belonging to the left group. For $\hat{d}_{miss}^{(k)}$ estimation, a minimal and maximal miss distance interval $[d_{min}^{(k)}, d_{max}^{(k)}]$ is first established. For every S_k , its $d_{min}^{(k)}$ spans from its coordinates \mathbf{x}_k in the direction towards the bullet's trajectory (perpendicularly to ϕ_Z) up to the point, where miss distance ambiguity starts (see Fig. 6 dashed lines); the maximal distance $d_{max}^{(k)}$ spans further, up to the point, where miss distance ambiguity ends. Equation (18) suggests, that $d_{S_k,Z}$ rises with $\hat{d}_{miss}^{(k)}$, therefore $S_k \in G_L$ will give larger, and $S_k \in G_R$ — smaller estimates if $\hat{d}_{miss}^{(k)}$ is at the ambiguity start of group G_R , and vice versa if it is at the ambiguity start of G_L . So, the ambiguity interval is iteratively passed from $d_{min}^{(k)}$ to $d_{max}^{(k)}$ with a step of d_{step} , the miss distances for K UGS are estimated as $\hat{d}_{miss}^{(k)} = d_{min}^{(k)} + i \cdot d_{step}$, and distance estimates to shooter $\hat{d}_{S_k,Z}(i)$ at each step are obtained using (18). A shooter position estimate $\hat{Z}_k(i)$ is computed per each UGS, using \mathbf{x}_k , $\phi_k^{(MB)}$ and $\hat{d}_{S_k,Z}(i)$. The fitness of $\hat{Z}_k(i)$ point estimates is measured by the their mean distance from their centroid $\bar{Z}(i)$:

$$f_{fit}(i) = \frac{1}{K} \sum_k \left\| \bar{Z}(i) - \hat{Z}_k(i) \right\|. \quad (19)$$

The minimum of f_{fit} suggests that $\hat{d}_{miss}^{(k)} \simeq d_{miss}^{(k)}$, and the final shooter's position estimate is selected as $Z = \bar{Z}(i)$, where $i = \text{ind min}(f_{fit})$.

V. EXPERIMENTAL RESULTS

The signals were acquired by 4 UGS during a live experiment at a shooting range. The shooter fired 30 shots from a standing position at a target 35 m away. The range was surrounded by scattered trees; a bullet-catching sand mound was situated approximately 5 m behind the target. The signals were acquired from four points in direct line-of-sight to the shooter. Layout coordinates in meters: target at (0, 0); shooter at (0, 35); UGS 1 at (4, 6); UGS 2 at (-5.5, 7); UGS 3 at (-6, 20); UGS 4 at (14, 7.5). The firearm used was the Husqvarna 8x57JS rifle with the cartridge muzzle velocity of 780 m/s, thus the shockwave is expected to spread at $\theta \simeq 25.8^\circ$ relative to bullet trajectory.

For the UCA implementation we have used a circular shell with the radius of $r = 7.5$ cm and $M = 6$ Vansonic PVM-6052 condenser microphones. The signals were acquired at $f_s = 48$ kS/s per channel and processed offline in MATLAB with an event window length of $t_W = 10$ ms. During signal acquisition each UGS recorded the signals only from its own UCA and no synchronization with other UGS was performed. DOA estimation with SRP-PHAT was performed with the following parameters: $r_{FOV} = 0.5$ m and $n_h = 500$. Data fusion parameters were the following: $\varphi_{max} = 20^\circ$, $\varphi_{max}^{(SW)} = 5^\circ$, $\varphi_{max}^{(MB)} = 45^\circ$.

An example of gunshot event detection was presented in Fig. 3. During the experiment all 30 shots were detected by all UGS, however, UGS 4 failed to detect seven MB events. Close analysis of signals acquired by UGS 4 has shown, that the number of detected events was equal to the number of signal perturbations per shot. This indicates the masking effect of outliers, such as TH. Other UGS have detected both SW and MB per every shot; TH was detected in the majority of cases. There have also been 13 cases of TH detection before MB by UGS 1 and 2, the reason being their close disposition to the target. These facts clearly indicate the need of gunshot event identification prior to shooter localization.

The two considered DOA estimation methods succeed in establishing a single distinct direction among outliers. Fig. 7a presents the DOA estimation results for an MB event, detected by UGS 1. SRP-PHAT values are scaled to the maximal value of 0.2; the individual pair-wise estimates of the proposed method are colored by their cross-correlation peak distinctness from the less to the most sharp as: black, blue, green, red. It can be seen that both methods produce one distinct beam and four lesser beams, corresponding to outliers. Fig. 7b presents the DOA estimates of four consecutive shots, computed by the proposed method. SW, MB and TH events are well distinguishable, however, results for UGS 4 are significantly worse due to its larger miss distance and the masking effect of TH. A more detailed comparison of SRP-PHAT and the proposed method is presented in [14]. In general, the DOA estimation quality is sufficient for the task of shooter localization.

The intermediate results of localization and the final shooter location estimate for a single shot are presented in Fig. 7c. UGS {2, 3} and {1, 4}, as expected, form clusters of coherent DOA estimates and group into G_L and G_R , respectively. Mean estimates of clustered DOA values are drawn in Fig. 7c from the points of these clusters' centroids. The shot angle $\phi_Z \simeq 90^\circ$ is estimated with high accuracy; $\check{S}_L = S_2$, $\check{S}_R = S_1$ are correctly assigned, and thus the miss distance uncertainty interval is properly computed.

Final shooter position estimates are presented in Fig. 8a. The mean error of position estimation for 30 shots is equal to 0.87 m. In Fig. 8a a congestion of remote points in the top left corner results from UGS 4 not detecting several MB. Bullet velocity estimation by $\hat{\theta}$ resulted in $\hat{v} \simeq 740$ m/s, which is consistent with the cartridge parameters. The values of the fitness function f_{fit} are presented in Fig. 8b. The function's minimum is situated at ± 1 m from the actual miss distance. As one global minimum of f_{fit} exists for every shot, miss distance estimation can be performed by a gradient descent method rather than by iterative search.

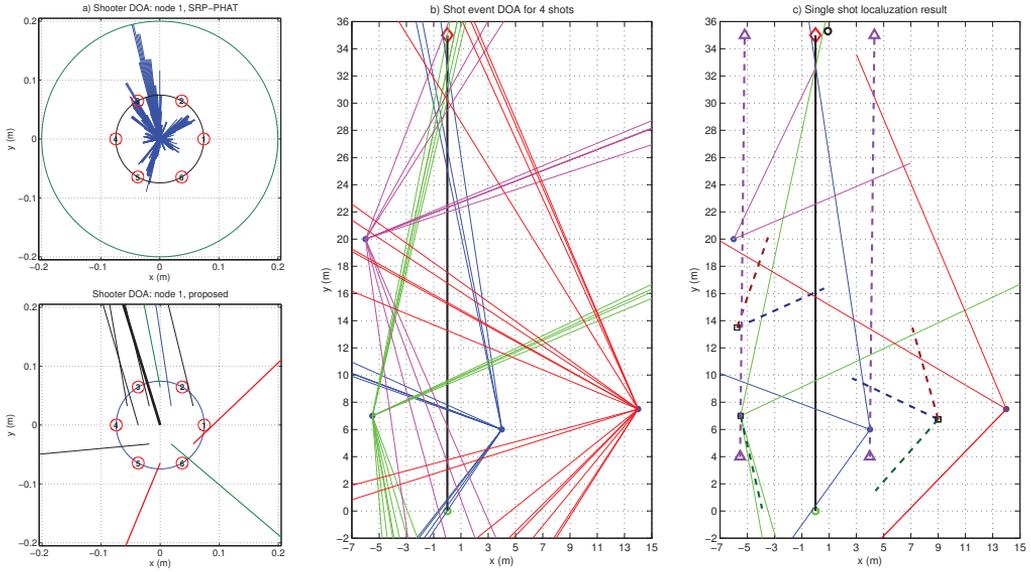


Fig. 7. a) $\hat{\phi}_1^{(MB)}$ estimation for the same event using SRP-PHAT and the proposed method (thick black — final estimate). b) DOA estimates for four consecutive shots using the proposed method (red diamond — shooter true position; green circle — target; blue dots — UGS positions; blue, green, purple, red lines — DOA estimates of UGS 1–4, respectively). c) Localization result for single shot (red, blue and green dotted — $\hat{\phi}^{(SW)}$, $\hat{\phi}^{(MB)}$ and outlier DOA of clusters Φ_p ; purple dotted — ϕ_Z and miss distance uncertainty; black circle — final estimated shooter position).

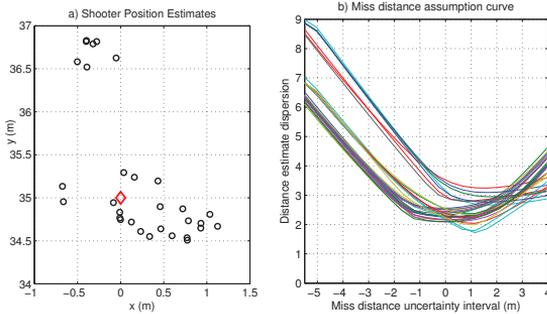


Fig. 8. a) Estimated shooter position for 30 shots (red diamond — shooter true position). b) Values of f_{fit} for the miss distance uncertainty interval.

VI. CONCLUSION

The proposed method of shooter localization by distributed grounded sensors is shown to provide adequate estimates in real life conditions with presence of outliers. Further experiments will be conducted to estimate the minimal degree of node spatial homogeneity, required for the method's robust operation. Also other gunshot scenarios and the methods of distinction between them need to be studied.

REFERENCES

- [1] J. Aguilar, "Gunshot location systems the transfer of the sniper detection technology from military to civilian applications," in *Proc. 47th Int. Carnahan Conf. on Security Technology (ICCST)*, Oct 2013, pp. 1–6.
- [2] J. A. Mazurek, J. E. Barger, M. Brinn, R. J. Mullen, D. Price, S. E. Ritter, and D. Schmitt, "Boomerang mobile counter shooter detection system," in *Proc. SPIE 5778, Sensors, and Command, Control, Communications, and Intelligence (C3I)*, vol. 5778, 2005, pp. 264–282.
- [3] J. George and L. Kaplan, "Shooter localization using soldier-worn gunfire detection systems," in *Proc. 14th Int. Conf. on Information Fusion (FUSION)*, July 2011, pp. 1–8.
- [4] J. Sallai, P. Volgyesi, K. Pence, and A. Ledeczki, "Fusing distributed muzzle blast and shockwave detections," in *Proc. 14th Int. Conf. on Information Fusion (FUSION)*, July 2011, pp. 1–8.
- [5] T. Damarla, L. Kaplan, and G. Whipps, "Sniper localization using acoustic asynchronous sensors," *Sensors Journal, IEEE*, vol. 10, no. 9, pp. 1469–1478, Sept 2010.
- [6] J. R. Aguilar, R. A. Salinas, and M. A. Abidi, "Acoustical model of small calibre ballistic shock waves in air for automatic sniper localization applications," *Proc. SPIE 6538, Sensors, and Command, Control, Communications, and Intelligence (C3I)*, vol. 6538, pp. 65381D–65381D–8, 2007.
- [7] J. Ash, G. Whipps, and R. Kozick, "Performance of shockwave-based shooter localization under model misspecification," in *2010 IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 2694–2697.
- [8] S. Astapov, J. Berdnikova, and J.-S. Preden, "A two-stage approach to 2D DOA estimation for a compact circular microphone array," in *Proc. 4th Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, 2015, in press.
- [9] U. Libal and K. Spyra, "Wavelet based shock wave and muzzle blast classification for different supersonic projectiles," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5097–5104, 2014.
- [10] A. Borzino, J. Apolinario, and M. de Campos, "Estimating direction of arrival of long range gunshot signals," in *Int. Telecommunications Symp. (ITS)*, Aug 2014, pp. 1–5.
- [11] J. S. Preden, J. Llinas, G. Rogova, R. Pahtma, and L. Motus, "On-line data validation in distributed data fusion," in *SPIE Defense, Security and Sensing, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV*, vol. 8742, 2013, pp. 1–12.
- [12] I. Freire and J. Apolinario, "GCC-based DoA estimation of overlapping muzzleblast and shockwave components of gunshot signals," in *IEEE Second Latin American Symp. on Circuits and Systems (LASCAS)*, Feb 2011, pp. 1–4.
- [13] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [14] S. Astapov, J. Ehala, J. Berdnikova, and J.-S. Preden, "Gunshot acoustic component localization with distributed circular microphone arrays," in *Proc. 2015 IEEE Int. Conf. on Digital Signal Processing (DSP)*, 2015, pp. 1186–1190.

**DISSERTATIONS DEFENDED AT
TALLINN UNIVERSITY OF TECHNOLOGY ON
INFORMATICS AND SYSTEM ENGINEERING**

1. **Lea Elmik**. Informational Modelling of a Communication Office. 1992.
2. **Kalle Tammemäe**. Control Intensive Digital System Synthesis. 1997.
3. **Eerik Lossmann**. Complex Signal Classification Algorithms, Based on the Third-Order Statistical Models. 1999.
4. **Kaido Kikkas**. Using the Internet in Rehabilitation of People with Mobility Impairments – Case Studies and Views from Estonia. 1999.
5. **Nazmun Nahar**. Global Electronic Commerce Process: Business-to-Business. 1999.
6. **Jevgeni Riipulk**. Microwave Radiometry for Medical Applications. 2000.
7. **Alar Kuusik**. Compact Smart Home Systems: Design and Verification of Cost Effective Hardware Solutions. 2001.
8. **Jaan Raik**. Hierarchical Test Generation for Digital Circuits Represented by Decision Diagrams. 2001.
9. **Andri Riid**. Transparent Fuzzy Systems: Model and Control. 2002.
10. **Marina Brik**. Investigation and Development of Test Generation Methods for Control Part of Digital Systems. 2002.
11. **Raul Land**. Synchronous Approximation and Processing of Sampled Data Signals. 2002.
12. **Ants Ronk**. An Extended Block-Adaptive Fourier Analyser for Analysis and Reproduction of Periodic Components of Band-Limited Discrete-Time Signals. 2002.
13. **Toivo Paavle**. System Level Modeling of the Phase Locked Loops: Behavioral Analysis and Parameterization. 2003.
14. **Irina Astrova**. On Integration of Object-Oriented Applications with Relational Databases. 2003.
15. **Kuldar Taveter**. A Multi-Perspective Methodology for Agent-Oriented Business Modelling and Simulation. 2004.
16. **Taivo Kangilaski**. Eesti Energia käiduhaldussüsteem. 2004.
17. **Artur Jutman**. Selected Issues of Modeling, Verification and Testing of Digital Systems. 2004.
18. **Ander Tenno**. Simulation and Estimation of Electro-Chemical Processes in Maintenance-Free Batteries with Fixed Electrolyte. 2004.

19. **Oleg Korolkov**. Formation of Diffusion Welded Al Contacts to Semiconductor Silicon. 2004.
20. **Risto Vaarandi**. Tools and Techniques for Event Log Analysis. 2005.
21. **Marko Koort**. Transmitter Power Control in Wireless Communication Systems. 2005.
22. **Raul Savimaa**. Modelling Emergent Behaviour of Organizations. Time-Aware, UML and Agent Based Approach. 2005.
23. **Raido Kurel**. Investigation of Electrical Characteristics of SiC Based Complementary JBS Structures. 2005.
24. **Rainer Taniloo**. Ökonoomsete negatiivse diferentsiaaltakistusega astmete ja elementide disainimine ja optimeerimine. 2005.
25. **Pauli Lallo**. Adaptive Secure Data Transmission Method for OSI Level I. 2005.
26. **Deniss Kumlander**. Some Practical Algorithms to Solve the Maximum Clique Problem. 2005.
27. **Tarmo Veskiõja**. Stable Marriage Problem and College Admission. 2005.
28. **Elena Fomina**. Low Power Finite State Machine Synthesis. 2005.
29. **Eero Ivask**. Digital Test in WEB-Based Environment 2006.
30. **Виктор Войтович**. Разработка технологий выращивания из жидкой фазы эпитаксиальных структур арсенида галлия с высоковольтным р-п переходом и изготовления диодов на их основе. 2006.
31. **Tanel Alumäe**. Methods for Estonian Large Vocabulary Speech Recognition. 2006.
32. **Erki Eessaar**. Relational and Object-Relational Database Management Systems as Platforms for Managing Softwareengineering Artefacts. 2006.
33. **Rauno Gordon**. Modelling of Cardiac Dynamics and Intracardiac Bio-impedance. 2007.
34. **Madis Listak**. A Task-Oriented Design of a Biologically Inspired Underwater Robot. 2007.
35. **Elmet Orasson**. Hybrid Built-in Self-Test. Methods and Tools for Analysis and Optimization of BIST. 2007.
36. **Eduard Petlenkov**. Neural Networks Based Identification and Control of Nonlinear Systems: ANARX Model Based Approach. 2007.
37. **Toomas Kirt**. Concept Formation in Exploratory Data Analysis: Case Studies of Linguistic and Banking Data. 2007.
38. **Juhan-Peep Ernits**. Two State Space Reduction Techniques for Explicit State Model Checking. 2007.

39. **Innar Liiv**. Pattern Discovery Using Seriation and Matrix Reordering: A Unified View, Extensions and an Application to Inventory Management. 2008.
40. **Andrei Pokatilov**. Development of National Standard for Voltage Unit Based on Solid-State References. 2008.
41. **Karin Lindroos**. Mapping Social Structures by Formal Non-Linear Information Processing Methods: Case Studies of Estonian Islands Environments. 2008.
42. **Maksim Jenihhin**. Simulation-Based Hardware Verification with High-Level Decision Diagrams. 2008.
43. **Ando Saabas**. Logics for Low-Level Code and Proof-Preserving Program Transformations. 2008.
44. **Ilja Tšahhиров**. Security Protocols Analysis in the Computational Model – Dependency Flow Graphs-Based Approach. 2008.
45. **Toomas Ruuben**. Wideband Digital Beamforming in Sonar Systems. 2009.
46. **Sergei Devadze**. Fault Simulation of Digital Systems. 2009.
47. **Andrei Krivošei**. Model Based Method for Adaptive Decomposition of the Thoracic Bio-Impedance Variations into Cardiac and Respiratory Components. 2009.
48. **Vineeth Govind**. DfT-Based External Test and Diagnosis of Mesh-like Networks on Chips. 2009.
49. **Andres Kull**. Model-Based Testing of Reactive Systems. 2009.
50. **Ants Torim**. Formal Concepts in the Theory of Monotone Systems. 2009.
51. **Erika Matsak**. Discovering Logical Constructs from Estonian Children Language. 2009.
52. **Paul Annus**. Multichannel Bioimpedance Spectroscopy: Instrumentation Methods and Design Principles. 2009.
53. **Maris Tõnso**. Computer Algebra Tools for Modelling, Analysis and Synthesis for Nonlinear Control Systems. 2010.
54. **Aivo Jürgenson**. Efficient Semantics of Parallel and Serial Models of Attack Trees. 2010.
55. **Erkki Joasoon**. The Tactile Feedback Device for Multi-Touch User Interfaces. 2010.
56. **Jürgo-Sören Preden**. Enhancing Situation – Awareness Cognition and Reasoning of Ad-Hoc Network Agents. 2010.
57. **Pavel Grigorenko**. Higher-Order Attribute Semantics of Flat Languages. 2010.
58. **Anna Rannaste**. Hierarcical Test Pattern Generation and Untestability Identification Techniques for Synchronous Sequential Circuits. 2010.

59. **Sergei Strik**. Battery Charging and Full-Featured Battery Charger Integrated Circuit for Portable Applications. 2011.
60. **Rain Ottis**. A Systematic Approach to Offensive Volunteer Cyber Militia. 2011.
61. **Natalja Sleptšuk**. Investigation of the Intermediate Layer in the Metal-Silicon Carbide Contact Obtained by Diffusion Welding. 2011.
62. **Martin Jaanus**. The Interactive Learning Environment for Mobile Laboratories. 2011.
63. **Argo Kasemaa**. Analog Front End Components for Bio-Impedance Measurement: Current Source Design and Implementation. 2011.
64. **Kenneth Geers**. Strategic Cyber Security: Evaluating Nation-State Cyber Attack Mitigation Strategies. 2011.
65. **Riina Maigre**. Composition of Web Services on Large Service Models. 2011.
66. **Helena Kruus**. Optimization of Built-in Self-Test in Digital Systems. 2011.
67. **Gunnar Pihõ**. Archetypes Based Techniques for Development of Domains, Requirements and Software. 2011.
68. **Juri Gavšin**. Intrinsic Robot Safety Through Reversibility of Actions. 2011.
69. **Dmitri Mihhailov**. Hardware Implementation of Recursive Sorting Algorithms Using Tree-like Structures and HFSM Models. 2012.
70. **Anton Tšertov**. System Modeling for Processor-Centric Test Automation. 2012.
71. **Sergei Kostin**. Self-Diagnosis in Digital Systems. 2012.
72. **Mihkel Tagel**. System-Level Design of Timing-Sensitive Network-on-Chip Based Dependable Systems. 2012.
73. **Juri Belikov**. Polynomial Methods for Nonlinear Control Systems. 2012.
74. **Kristina Vassiljeva**. Restricted Connectivity Neural Networks based Identification for Control. 2012.
75. **Tarmo Robal**. Towards Adaptive Web – Analysing and Recommending Web Users` Behaviour. 2012.
76. **Anton Karputkin**. Formal Verification and Error Correction on High-Level Decision Diagrams. 2012.
77. **Vadim Kimlaychuk**. Simulations in Multi-Agent Communication System. 2012.
78. **Taavi Viilukas**. Constraints Solving Based Hierarchical Test Generation for Synchronous Sequential Circuits. 2012.

79. **Marko Kääramees**. A Symbolic Approach to Model-based Online Testing. 2012.
80. **Enar Reilent**. Whiteboard Architecture for the Multi-agent Sensor Systems. 2012.
81. **Jaan Ojarand**. Wideband Excitation Signals for Fast Impedance Spectroscopy of Biological Objects. 2012.
82. **Igor Aleksejev**. FPGA-based Embedded Virtual Instrumentation. 2013.
83. **Juri Mihhailov**. Accurate Flexible Current Measurement Method and its Realization in Power and Battery Management Integrated Circuits for Portable Applications. 2013.
84. **Tõnis Saar**. The Piezo-Electric Impedance Spectroscopy: Solutions and Applications. 2013.
85. **Ermo Täks**. An Automated Legal Content Capture and Visualisation Method. 2013.
86. **Uljana Reinsalu**. Fault Simulation and Code Coverage Analysis of RTL Designs Using High-Level Decision Diagrams. 2013.
87. **Anton Tšepurov**. Hardware Modeling for Design Verification and Debug. 2013.
88. **Ivo Mürsepp**. Robust Detectors for Cognitive Radio. 2013.
89. **Jaas Ježov**. Pressure sensitive lateral line for underwater robot. 2013.
90. **Vadim Kaparin**. Transformation of Nonlinear State Equations into Observer Form. 2013.
92. **Reeno Reeder**. Development and Optimisation of Modelling Methods and Algorithms for Terahertz Range Radiation Sources Based on Quantum Well Heterostructures. 2014.
93. **Ants Koel**. GaAs and SiC Semiconductor Materials Based Power Structures: Static and Dynamic Behavior Analysis. 2014.
94. **Jaan Übi**. Methods for Coopetition and Retention Analysis: An Application to University Management. 2014.
95. **Innokenti Sobolev**. Hyperspectral Data Processing and Interpretation in Remote Sensing Based on Laser-Induced Fluorescence Method. 2014.
96. **Jana Toompuu**. Investigation of the Specific Deep Levels in p -, i - and n -Regions of GaAs p^+pin-n^+ Structures. 2014.
97. **Taavi Salumäe**. Flow-Sensitive Robotic Fish: From Concept to Experiments. 2015.
98. **Yar Muhammad**. A Parametric Framework for Modelling of Bioelectrical Signals. 2015.
99. **Ago Mõlder**. Image Processing Solutions for Precise Road Profile Measurement Systems. 2015.

100. **Kairit Sirts**. Non-Parametric Bayesian Models for Computational Morphology. 2015.
101. **Alina Gavrijaševa**. Coin Validation by Electromagnetic, Acoustic and Visual Features. 2015.
102. **Emiliano Pastorelli**. Analysis and 3D Visualisation of Microstructured Materials on Custom-Built Virtual Reality Environment. 2015.
103. **Asko Ristolainen**. Phantom Organs and their Applications in Robotic Surgery and Radiology Training. 2015.
104. **Aleksei Tepljakov**. Fractional-order Modeling and Control of Dynamic Systems. 2015.
105. **Ahti Lohk**. A System of Test Patterns to Check and Validate the Semantic Hierarchies of Wordnet-type Dictionaries. 2015.
106. **Hanno Hantson**. Mutation-Based Verification and Error Correction in High-Level Designs. 2015.
107. **Lin Li**. Statistical Methods for Ultrasound Image Segmentation. 2015.
108. **Aleksandr Lenin**. Reliable and Efficient Determination of the Likelihood of Rational Attacks. 2015.
109. **Maksim Gorev**. At-Speed Testing and Test Quality Evaluation for High-Performance Pipelined Systems. 2016.
110. **Mari-Anne Meister**. Electromagnetic Environment and Propagation Factors of Short-Wave Range in Estonia. 2016.
111. **Syed Saif Abrar**. Comprehensive Abstraction of VHDL RTL Cores to ESL SystemC. 2016.
112. **Arvo Kaldmäe**. Advanced Design of Nonlinear Discrete-time and Delayed Systems. 2016.
113. **Mairo Leier**. Scalable Open Platform for Reliable Medical Sensorics. 2016.
114. **Georgios Giannoukos**. Mathematical and Physical Modelling of Dynamic Electrical Impedance. 2016.
115. **Aivo Anier**. Model Based Framework for Distributed Control and Testing of Cyber-Physical Systems. 2016.