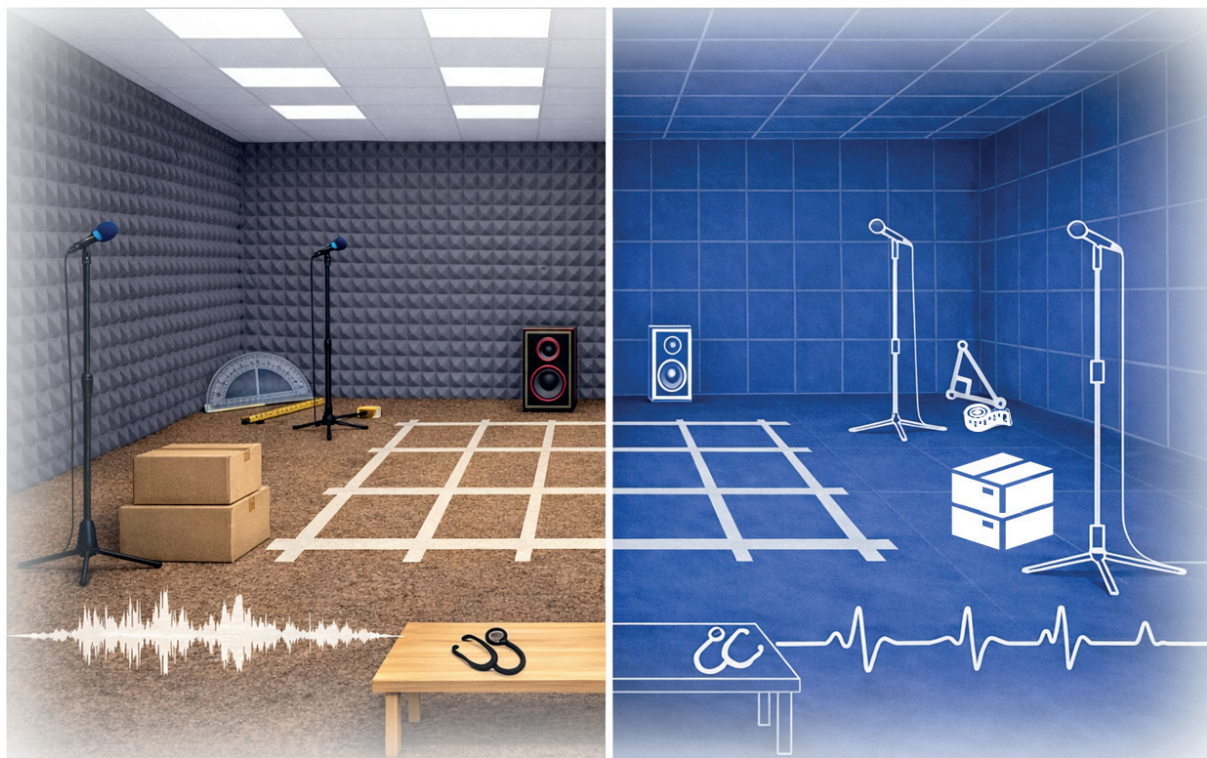


Measurement Quality in Acoustic Sensing with Microphones: From Indoor Localization to Heart Sound Classification

Meng Jiang



Measurement Quality in Acoustic Sensing with Microphones: From Indoor Localization to Heart Sound Classification

Meng Jiang



Mittuniversitetet
MID SWEDEN UNIVERSITY

Department of Computer and Electrical Engineering
Mid Sweden University

Doctoral Thesis No. 446
Sundsvall, Sweden
2026

ISBN 978-91-90017-57-9
ISSN 1652-893X

Mittuniversitetet
Data- och elektroteknik
SE-851 70 Sundsvall
SWEDEN

Akademisk avhandling som med tillstånd av Mittuniversitetet framlägges till offentlig granskning för avläggande av teknologie doktorsexamen den 25 February 2026 klockan 09:00 i sal M108, Mittuniversitetet Holmgatan 10, Sundsvall. Seminarier kommer att hållas på engelska.

©Meng Jiang, January 2026

Tryck: Tryckeriet Mittuniversitetet

To
My Parents "Connie & Ping"
My Husband "Janne"
My Children "Almer & Emma"
My Supervisors "Göran & Mårten & Chibuzo"

饮其流者怀其源，学其成时念吾师。作为我人生中不同时期的老师，你们无时无刻地赐予我灵感，给予我动力，鼓励我在人生的道路上继续。里程碑不仅是一个时期的结束，更是一个新时期的开始。

Those who drink from its flow cherish its source. As teachers at different times in my life, you inspire me all the time, motivate me, and encourage me to continue on my own journey. A milestone is not only an end to a period but also a commencement of a new period.

Abstract

This thesis investigates how measurement design shapes acoustic source localization and classification, with a focus on the interplay between array geometry, device characteristics, and modern signal processing and deep learning. The work is motivated by a persistent gap between theoretically well-understood methods and the practical realities of indoor positioning and biomedical auscultation, where sensor variability, reverberation, and limited control over operating conditions often dominate performance. The overarching aim is to understand how measurement quality in microphone-based sensing constrains and enables what can be inferred from sound under real-world noise, by treating microphones, arrays, and recording protocols as design variables rather than static background assumptions.

Six studies (P1–P6 refer to the list of papers) are presented. The first line of work concerns acoustic fingerprinting. P1 examines how far a single microphone can exploit ambient noise for indoor “silent” object localization, highlighting both the appeal of zero-emission fingerprints and their sensitivity to day-to-day room changes. P6 revisits fingerprinting with active excitation, using exponential sine sweeps and a four-microphone array feeding a convolutional neural network. The comparison between P1 and P6 shows how moving from uncontrolled ambient sound to controlled probing and array-based features improves robustness. Together, they characterize a practical design space for silent object localization, from simple cross-correlation baselines to array-aided deep learning.

The second line of work addresses direction-of-arrival (DoA) estimation with microphone arrays. P2 compares several planar layouts and microphone directivities in a controlled room, using a representative high-resolution DoA estimator to isolate how geometry and sensor pattern affect accuracy and robustness in realistic indoor conditions. P3 focuses on a six-channel uniform circular array and a coherent wide-band pipeline, showing that circular-harmonic focusing can retain MUSIC-level resolution while keeping computational demands compatible with embedded implementations. These studies map how established methods behave when constrained by physically small arrays and practical sensor choices, clarifying when geometry or processing is the main bottleneck.

A third line of work turns to biomedical acoustic classification. P4 evaluates a four-channel electronic stethoscope prototype that combines delay-and-sum beamforming and matched filtering for heart-sound segmentation before classification.

Working with a limited and clinically constrained dataset, the study illustrates how a realistic multi-channel auscultation setup can increase segment quality and support distinguish normal and abnormal sound for murmur detection. Finally, the thesis examines measurement quality more generally. P5 introduces a measurement-quality pipeline that uses existing recordings to extrapolate the benefit of future system upgrades. By fixing a pretrained CNN and synthetically degrading current data to different SNR levels, the study emulates the performance of improved setups, providing a basis for deciding whether it is worthwhile to invest in new measurements and a full round of model retraining and tuning. These results underline that model architecture and measurement quality jointly determine performance, and that metrological upgrades can sometimes deliver rich information without retraining.

Overall, the thesis contributes a set of measurement-driven case studies that make explicit how arrays, excitation signals, and device responses constrain what localization and classification algorithms can realistically achieve. The outcomes include practical recipes for acoustic fingerprinting, design reference points for compact array configurations in indoor DoA tasks, an experimentally grounded path toward reproducible multi-channel auscultation, and empirical guidelines for anticipating how SNR and device variability affect pretrained models. Rather than resolving all trade-offs, the work argues for treating measurement design and algorithm choice as coupled problems.

Sammanfattning

Denna avhandling undersöker hur mätutformning påverkar akustisk källlokalisering och klassificering, med fokus på samspelet mellan arraygeometri, enhetsegenskaper samt modern signalbehandling och djupinlärning. Arbetet motiveras av en bestående klyfta mellan teoretiskt väletablerade metoder och de praktiska realiteterna i inomhuslokalisering och biomedicinsk auskultation, där sensorvariation, efterklang och begränsad kontroll över driftförhållanden ofta dominerar prestandan. Det övergripande målet är att förstå hur mätkvalitet i mikrofonbaserad sensning både begränsas och möjliggör vad som kan utläsas ur ljud under verkligt brus, genom att betrakta mikrofoner, arrayer och mätprotokoll som designvariabler snarare än statistiska bakgrundsantaganden.

Sex studier (P1–P6 i publikationslistan) presenteras. Den första inriktningen handlar om akustisk fingerprinting. P1 undersöker hur långt en enda mikrofon kan utnyttja omgivningsljud för inomhuslokalisering av ett "tyst" objekt, och lyfter fram både fördelarna med nollemissions-fingeravtryck och deras känslighet för dagliga förändringar i rummet. P6 återbesöker fingerprinting med aktiv excitation, genom exponentiella svepsignaler och en kompakt array kopplad till ett konvolutionellt neuralt nätverk. Jämförelsen mellan P1 och P6 visar hur övergången från okontrollerat bakgrundsljud till kontrollerad sondering och arraybaserade egenskaper förbättrar robustheten. Tillsammans karakteriserar de ett praktiskt designutrymme för lokalisering av tysta objekt, från enkla korskorrelationsbaslinjer till arraystödd djupinlärning.

Den andra inriktningen behandlar riktning-till-ankomst (DoA) med kompakta arrayer. P2 jämför flera plana geometrier och mikrofonriktningar i ett kontrollerat rum, med en representativ högupplöst DoA-estimator, för att isolera hur geometri och riktningsskäraktär påverkar noggrannhet och robusthet under realistiska inomhusförhållanden. P3 fokuserar på en sexkanalig uniform cirkulär array och en koherent bredbandskedja, och visar att cirkulärharmonisk fokusering kan behålla MUSIC-nivå på upplösningen samtidigt som beräkningskraven hålls förenliga med inbyggda implementationer. Dessa studier kartlägger hur etablerade metoder beter sig när de begränsas av fysiskt små arrayer och praktiska sensorval, och tydliggör när geometri respektive signalbehandling är den dominerande flaskhalsen.

En tredje inriktning rör biomedicinsk akustisk klassificering. P4 utvärderar en fyrkanalig elektronisk stetoskopprototyp som kombinerar delay-and-sum-strålformning

för matchningsfiltrering för segmentering av hjärtljud före klassificering. Med ett begränsat och kliniskt styrt dataset visar studien hur en realistisk flerkanalig auskultationsuppställning kan förbättra segmentkvaliteten och stödja blåsljuddetektion, samtidigt som den blottlägger öppna frågor kring etikettvariation, placeringskonsekvens och arbetsflöde. Slutligen studerar avhandlingen mätkvalitet mer allmänt. P5 introducerar en metod för att analysera mätkvalitet som utnyttjar befintliga inspelningar för att uppskatta nyttan av framtida systemuppgraderingar. Genom att låsa en förtränad CNN och syntetiskt försämra de befintliga data till olika SNR-nivåer efterliknas prestandan hos förbättrade respektive sämre mätuppsättningar, vilket ger beslutsstöd för om det är värt att samla in nya mätningar och genomföra en full omgång omträning och finjustering av modellen. Dessa resultat understryker att modellarkitektur och mätkvalitet tillsammans bestämmer prestandan, och att metrologiska uppgraderingar ibland kan ge rikare information utan omskolning av modellen.

Sammantaget bidrar avhandlingen med en uppsättning mätstyrda fallstudier och protokoll som tydliggör hur arrayer, excitationssignaler och enhetsvar begränsar vad lokaliserings- och klassificeringsalgoritmer realistiskt kan åstadkomma. Utfallet inkluderar praktiska recept för akustisk fingerprinting, referenspunkter för design av kompakta arraykonfigurationer i inomhus-DoA-uppgifter, en experimentellt förankrad väg mot reproducerbar flerkanalig auskultation, samt empiriska riktlinjer för hur SNR och enhetsvariation påverkar förtränade modeller. I stället för att lösa alla avvägningar argumenterar arbetet för att mätutformning och algoritmval bör behandlas som sammanlänkade problem.

Acknowledgements

First and foremost, I would like to thank my supervisors, Associate Professor Göran Thungström, Prof. Mårten Sjöström and Dr. Chibuzo Nnonyelu, for their guidance and support, and for both their insights and their example of working through the research process.

Secondly, I would like to thank my colleagues and co-authors at the STC Research Centre, Rikard Hamrin, Shan Gao and Marianthi Adamapoulou, for their collaboration, encouragement, and many helpful discussions, including several memorable “last-minute manuscript editing” moments. I am also deeply grateful to our collaborators at the University of Salerno in Italy, Marco Carratù, Vincenzo Gallo, and Valter Laino for many inspiring discussions and for sharing International Instrumentation and Measurement Conference (I2MTC) experiences, presentations, and focused exchanges of ideas. Their feedback and support has shaped both this work and the way I think about research.

I would also like to thank the technical and administrative staff, whose often invisible work keeps infrastructure and logistics around the laboratory making my measurements running smoothly. In particular, I am grateful for the help of all the Vaktmästare from INFRA department, in coordinating building access and control, scheduling and supporting the renovation of the lab room, and arranging temporary ventilation shutdowns and other building-level adjustments that made it possible to carry out measurements when needed. Their patience with repeated “just one more measurement” requests was invaluable.

Furthermore, I gratefully acknowledge the financial support from the Swedish Knowledge Foundation and other project partners, as well as the support provided by Mid Sweden University, DET department, and the STC Research Centre. Their commitment to long-term, curiosity-driven research made this work possible.

Finally, I would like to thank my family for their encouragement, understanding, and constant support throughout this journey. To my two children Almer and Emma, who arrived in the middle of this PhD, thank you for the sleepless nights, the loud midnight crying, and the rare “angel baby” moments that somehow made everything feel worthwhile and kept life in perspective. I am deeply grateful to my parents Connie and Ping, who flew all the way from China and stayed through the pandemic, sharing the chaos of two babies and patiently enduring the Nordic cold and darkness so that I could focus on my work. I also wish to thank my husband

Janne and my mother-in-law Ing-Marie, for your patience, emotional stability, and strength in keeping my sanity intact whenever I was overwhelmed by stress and exhaustion. Their belief in me, especially during the less glamorous phases of debugging and rewriting, has been a steady source of motivation. This thesis is as much a result of their support as it is of my own efforts.

Contents

Abstract	v
Sammanfattning	vii
Acknowledgements	ix
List of Papers	xv
Terminology	xxi
1 Introduction	1
1.1 Background and Motivation	1
1.2 Overall Aim	2
1.3 Research Topics and Research Questions	3
1.4 Scope and Delimitations	5
1.5 Author Contributions	5
1.6 Outline of the Thesis	6
2 Theoretical Background	7
2.1 Acoustic propagation basics	8
2.2 Acoustic Fingerprinting	8
2.3 Channel State Information (CSI)	8
2.4 Array-Based Estimation for DoA	9
2.5 Beamforming	10
2.6 Deep learning in audio measurements	12
2.7 Cardiac auscultation	13

2.8	Measurement quality & sensors	14
3	Related Works	15
3.1	Acoustic fingerprinting	15
3.2	Direction of Arrival Estimation	16
3.3	Acoustic classification with heart sound	17
3.4	Measurement quality and sensor factors	18
3.5	Positioning of this thesis	19
4	Methodology	21
4.1	Proposed Measurement Methods	21
4.1.1	Fingerprinting: From Passive to Deep Learning	22
4.1.2	DoA Estimation with Microphone Arrays	22
4.1.3	Classification and Measurement Quality	24
4.2	Experimental Verification	25
4.2.1	Silent object fingerprinting (RQ1.1–RQ1.2)	25
4.2.2	DoA with microphone arrays (RQ2.1–RQ2.2)	26
4.2.3	Biomedical application (RQ3.1–RQ3.2)	27
5	Results	29
5.1	Integrated Results Across Problem Areas	29
5.1.1	Indoor Spatial and Environmental Sensing	29
5.1.2	Biomedical Sound Sensing	31
5.2	Summary	33
6	Discussion	35
6.1	Reflections on Selected Methodology	35
6.1.1	Validity and Reliability	36
6.1.2	Alternatives	36
6.1.3	Generalization and answers to the RQs	39
6.1.4	Scope and trade-offs	39
6.1.5	Technical Limitations	41
6.1.6	Practical Limitations	41
6.2	Research Outcomes	42
6.2.1	Use and Misuse of Outcomes	42

6.2.2	Originality and Novelty	42
6.2.3	Significance and Impact	43
6.3	Risks and Ethical Aspects	44
6.3.1	Uncertainties and measurement bias	44
6.3.2	Ethical aspects in research process	44
7	Conclusions	47
7.1	Work Conclusions	47
7.1.1	Chronological Network of the Study	47
7.1.2	Summary by Problem Area	47
7.1.3	Key Contributions of the Thesis	49
7.1.4	Knowledge Gained	49
7.1.5	Overall Conclusion	50
7.2	Future Work	50
7.2.1	Ongoing and Related Projects	51
7.2.2	Possible Directions	51
7.3	Closing Remark	52
	Bibliography	65
	Biography	67

List of Publications

This thesis is mainly based on the following papers:

- P1. Jiang, M., Lundgren, J., Pasha, S., Carratù, M., Liguori, C., & Thungström, G. "Indoor Silent Object Localization using Ambient Acoustic Noise Fingerprinting" published in *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, IEEE, May 2020.
- P2. Jiang, M., Nnonyelu, C. J., Lundgren, J., Sjöström, M., Thungström, G., & Gao, S. "Performance Comparison of Omni and Cardioid Directional Microphones for Indoor Angle of Arrival Sound Source Localization" published in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, IEEE, May 2022.
- P3. Jiang, M., Nnonyelu, C. J., Lundgren, J., Thungström, G., & Sjöström, M. "A Coherent Wideband Acoustic Source Localization Using a Uniform Circular Array" published in *Sensors*, 23(11):5061, 2023.
- P4. Adamopoulou, M., Jiang, M., Nnonyelu, C. J., Carratù, M., Liguori, C., & Lundgren, J. "Improving Cardiac Auscultation Signal Quality by using 4-Channel Stethoscope Array" published in *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, IEEE, 2024.
- P5. Lundgren, J., Jiang, M., Laino, V., Gallo, V., Carratù, M., & Nnonyelu, C. J. "Accuracy Impact of Increased Measurement Quality when using Pretrained Networks for Classification" published in *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, IEEE, 2024.
- P6. Jiang, M., Nnonyelu, C. J., Adamopoulou, M., Lundgren, J., Carratù, M., V., Gallo, Laino, V.. "Silent Object Localization on a Grid Using RIR-Based Acoustic Fingerprints and EfficientNet" submitted to *2026 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (Accepted).

Other works:

- 1. Nnonyelu, C. J., Jiang, M., & Lundgren, J. "A Lower Bound on the Estimation

- Variance of Direction-of-Arrival and Skew Angle of a Biaxial Velocity Sensor Suffering from Stochastic Loss of Perpendicularity* published in *Sensors*, 22(21), 8464, 2022.
2. Nnonyelu, C. J., Jiang, M., & Lundgren, J. "Spherical-sector harmonics domain processing for wideband source localization using spherical-sector array of directional microphones" Meeting abstract published in *J. Acoust. Soc. Am.* 153, A54, 2023.
 3. Nnonyelu, C. J., Jiang, M., Adamopoulou, M., & Lundgren, J. "A Machine-Learning -based approach to Direction-of-arrival Sectorization using Spherical Microphone Array" published in 2024 IEEE 13rd Sensor Array and Multichannel Signal Processing Workshop (SAM), 2024.
 4. Nnonyelu, C. J., Jiang, M., Adamopoulou, M., & Lundgren, J. "Performance Analysis of Cardioid and Omnidirectional Microphones in Spherical Sector Arrays for Coherent Source Localization" published in *Sensors* 2024, 24(23), 7572, 2024.
 5. Nnonyelu, C. J., Jiang, M., Gallo, V., Laino, V., Carratù, M., & Lundgren, J. "Signal First-Difference as Augmentation Method for CNN-Based Heart Sound Classification" published in 2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2025.
 6. Jiang, M., Nnonyelu, C. J., Carratù, M., Adamopoulou, M., Thungström, G., & Lundgren, J. "A Closed-form Eigenmode-based DoA Estimation using Uniform Circular Array" published in 2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2025.
 7. Gallo, V., Lundgren, J., Jiang, M. "An Embedded Edge Architecture for Real-Time Acoustic Direction Monitoring Using a Four-Microphone Array and GCC-PHAT" submitted to 2026 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (Accepted).

List of Figures

1.1	Overview of the thesis: acoustic sensing under real-world noise, split into three research topics (A–C) and associated research questions (RQs).	3
2.1	Auscultation key aspect picture illustration for heart sound and murmur classification criteria.	14
4.1	Directionality polar pattern of (a) omni-directional microphone (e.g Rode NT45-O capsule), and (b) cardioid-directional microphone (e.g Rode NT45-C capsule).	23
4.2	The illustration of a 3D model of designed multichannel stethoscope. .	24
5.1	Illustration of P1 accuracy result with a radius of 29.9 cm from the reference position (inner dotted circle around the grid center) and a mean radial error of 12.8 cm (example solid circle) when the silent object stands between grids.	30
5.2	Illustration of P4 on real-world auscultation recordings, showing sectional level increases from beamforming. Three samples from co-authors are presented. Sample 2 (middle) has murmur condition, therefore during systole and diastole sound is occurring, translated into a sectional level closer to 0 dB, indicating stronger energy in the murmur-bearing region, while Samples 1 and 3 are healthy.	31
7.1	Illustration of the progression of the studies and connections between all papers.	48

List of Tables

2.1	Terms of the localization methods and their characteristics	11
5.1	Signal sectional level $L^{(s)}$ (dB) by processing stage.	32
5.2	Stage-to-stage gains and normalized dB change.	32
5.3	Illustrative headline results from P1–P6.	34

Terminology

Abbreviations and Acronyms

AMI	Array Manifold Interpolation
AOA	Angle of Arrival
CDM	Cardioid-directional Microphone
CNN	Convolutional Neural Network
CRNN	Convolutional Recurrent Neural Network
CSI	Channel State Information
CSSM	Coherent Signal-Subspace Method
DOA	Direction of Arrival
DSB	Delay-and-Sum Beamforming
ECE	Expected Calibration Error
ESS	Exponential Swept Sine
GCC	Generalized Cross Correlation
HSMM	Hidden Semi-Markov Model
KNN	k-Nearest Neighbour
MAE	Mean Absolute Error
MEMS	Micro-Electro-Mechanical System
MFCC	Mel-Frequency Cepstral Coefficients
MUSIC	Multiple Signal Classification
MVDR	Minimum Variance Distortionless Response
ODM	Omni-directional Microphone
PHAT	Phase Transform
RIR	Room Impulse Response
RMSE	Root-Mean-Squared Error
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRP	Steered Response Power
SSL	Sound Source Localization

SVM	Support Vector Machine
TDOA	Time Difference of Arrival
1D	one-dimensional
2D	two-dimensional

Chapter 1

Introduction

Nowadays, it is more and more common that microphones and acoustic sensors are embedded in buildings, mobile devices, and medical equipment. Beyond recording audio content, they are increasingly used as measurement instruments, for example to infer where sound sources are, how environments respond to acoustic changes, and to assess the state of physiological systems. This thesis focuses on such measurement-driven uses of acoustic sound, by investigating acoustic sensing under real-world noise.

Within this broad perspective, the thesis concentrates on two main application areas. The first is indoor acoustic sensing, where microphones are used to recover spatial information, and the second area is biomedical sound sensing, where heart sound plays a major role. Across these application areas, a common pattern appears: sophisticated algorithms are available, but their performance and usage conditions depends strongly on the quality and stability of the underlying measurements.

1.1 Background and Motivation

Sound source localization and acoustic classification are fundamental challenges in measurement engineering. They enable applications such as indoor positioning, environmental monitoring, and biomedical diagnostics. Traditionally, the focus of research has been on signal processing algorithms, including beamforming [1–4], subspace methods [5,6], and, more recently, machine learning and deep learning approaches [2,7]. However, practical systems are constrained by measurement-related factors: the type of microphones used, the geometry and aperture of the array, device variability, and the presence of real-world noise [2,6,8].

A recurring limitation in prior research is the insufficient attention to these measurement aspects. For example, many studies benchmark algorithms under idealized conditions, under-reporting sensor characteristics, calibration, or room conditions. Furthermore, reviews and challenge reports note that many experimental

papers provide limited detail about their measurement setups, which hinders verification and replication [6, 8]. The LOCATA¹ initiative was explicitly motivated by this reproducibility gap and the lack of a common, well-documented dataset for fair comparison [9, 10]. Its emphasis on carefully specified arrays, rooms, and measurement protocols mirrors the motivation of this thesis, where measurement design and transparent reporting are treated as central parts of acoustic sensing rather than secondary details. Likewise, passive localization approaches such as acoustic fingerprinting have been explored, but questions remain about their robustness under uncontrolled conditions [8, 11]. In biomedical acoustics, advances in machine learning have driven classification performance, yet relatively little has been done to systematically evaluate how multi-channel recording strategies improve the measurement quality at the source [7, 12]. Addressing these gaps by foregrounding measurement design, reporting, and their link to algorithm behavior is the motivation of this thesis.

1.2 Overall Aim

The overall aim of this dissertation is to investigate acoustic sensing under real-world noise. The research builds upon a sequence of studies carried out between 2020 and 2025, spanning acoustic fingerprinting, direction-of-arrival (DoA) estimation, acoustic classification, and measurement quality. The work examines how far theoretical models can be carried into practical deployment and where measurement constraints become limiting by focusing on real-world measurement conditions, microphone choice and array design, and on how established methods behave under these conditions.

In this thesis, measurement quality is understood broadly as the suitability of acoustic data for extracting the spatial or diagnostic information of interest. This includes sensor configuration (geometry, spacing, and directionality), excitation and frequency content, signal-to-noise ratio and reverberation, and the choice of signal representation. In what follows, the term “measurement design” refers both to how microphones, arrays, excitations, and recording protocols are chosen, and to how these measurement conditions and observations shape the design of the algorithms and experiments themselves. The studies are measurement-driven in the sense that results from earlier measurement activities inform which models are used, how they are tuned, and how later experiments are structured.

As illustrated in Figure 1.1, the work is organized into three research topics that span two application areas. In spatial and environmental sensing, there are two related research topics: Topic A focuses on grid-based localization via acoustic fingerprinting for silent object localization, while Topic B investigates array-based sound source localization with DoA estimation. Another application area is biomedical sensing, where Topic C explores the application of cardiac auscultation with the goal of fully utilizing heart sound signals to extract useful information. These topics are

¹LOCATA – The IEEE Audio and Acoustic Signal Processing (AASP) Challenge on acoustic source Localization And TrAcking (LOCATA)

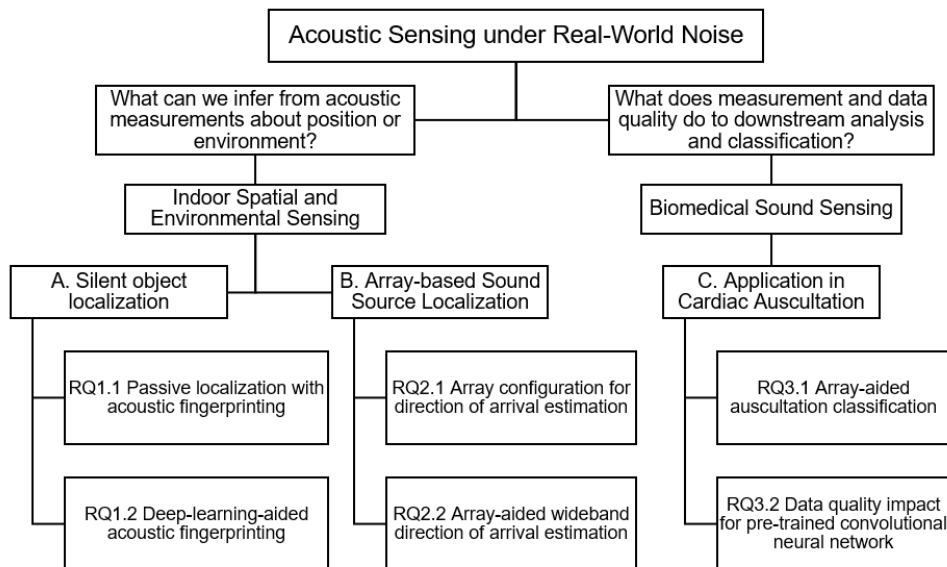


Figure 1.1: Overview of the thesis: acoustic sensing under real-world noise, split into three research topics (A–C) and associated research questions (RQs).

further specified through research questions in Section 1.3.

Together, these studies support the overarching theme of the thesis: how microphones and measurement design shape what can be reliably inferred from sound under real-world noise, from indoor localization to heart sound classification.

1.3 Research Topics and Research Questions

As mentioned in Figure 1.1, three research topics later derived into six specific research questions investigated in each paper, addressing the two measurement-driven problem areas. This section presents the topics and their related research questions (RQs).

Topic A. Silent Object Fingerprinting: Passive fingerprinting allows for localization of silent objects, where early approaches were mainly based on radio frequency identification (RFID), or WiFi. Under some conditions, active emitters, tags or extra devices on the object being localized are undesired. It is essential to investigate the scalability of passive acoustic methods in high-noise environments and evaluate how structured arrays, integrated with modern machine learning, can enhance performance. Note that in this thesis, the definition of “Passive” in passive fingerprinting refers to the object being located remains passive and silent rather than actively making sound.

RQ1.1 To what extent can ambient noise be exploited for reliable acoustic fingerprinting to achieve localization without active emission?

RQ1.2 How does the integration of spatial features extracted from a microphone array with convolutional neural network (CNN) enhance acoustic fingerprinting for grid-based indoor localization?

Topic B. Array-based Sound Source Localization: DoA estimation methods are well established in controlled environments. However, their performance under realistic measurement conditions and practical microphone array designs remains under-explored and worth investigating. Existing literature often neglects the systematic evaluation of how array geometry, microphone directivity, and processing methods interact in complex acoustic fields, leaving a gap in understanding optimal measurement configurations for high-accuracy indoor localization.

RQ2.1 To what extent can a change of measurement configuration (as in array geometry selection and microphone directionality) improve DoA estimation accuracy in indoor environments?

RQ2.2 How can coherent wideband processing methods, specifically the coherent signal subspace method (CSSM), improve sound source localization with a uniform circular array in a reverberant measurement environment?

Topic C. Application in Cardiac Auscultation: In biomedical applications such as cardiac auscultation, acoustic measurements are easily degraded by environmental noise, motion artifacts, and sensor limitations. While the fundamental quality of measurements significantly shapes the performance of downstream localization and classification systems, most existing studies prioritize algorithm design over the systematic evaluation of device-level factors. The potential of multi-channel stethoscope arrays, to improve the signal-to-noise ratio (SNR) of heart sound data for subsequent classification accuracy to support diagnostic tasks, remains to be explored.

RQ3.1 How can a multi-channel digital stethoscope array be leveraged to enhance the signal quality of heart sound acquisition and aid heart sound segmentation by diagnostic classification?

RQ3.2 Measurements taken at different times or under varying conditions will result in differing data quality (e.g., varying SNR). How can we utilize this information to extrapolate useful insights for future development beyond the data's original task?

Each research question is systematically addressed across the included publications. Papers P1–P3 and P6 investigate how measurement design affects indoor localization with single microphone and small arrays, specifically through room fingerprints, microphone types, array manifolds, and frequency–spatial constraints in wideband circular arrays. P1 and P6 address RQ1.1 and RQ1.2, while P2 and P3 address RQ2.1 and RQ2.2, respectively. Regarding RQ3.1, P4 applies similar methodologies to cardiac auscultation, utilizing a digital stethoscope and beamforming to

increase the signal quality of heart sound recordings. Finally, P5 addresses RQ3.2 by quantifies how variations in measurement quality, in particular signal-to-noise ratio(SNR), impact the performance of pre-trained convolutional neural networks.

1.4 Scope and Delimitations

The thesis is limited to measurement-driven investigations of acoustic source localization and classification. The focus is on indoor environments and biomedical scenarios where sound-based real-world measurement with noise plays a central role. Hardware selections and experimental environments are provided by Mid Sweden University, which most of the work in this thesis was based on. Outdoor propagation environments, strong multipath scenarios, and high-mobility tracking are outside the scope. Algorithmic contributions (e.g., coherent wideband DoA, deep learning for fingerprinting) are presented, but always in the context of how measurement setups enable or limit their performance. The included publications are P1–P6. Future work is presented in Section 7.2.

1.5 Author Contributions

The thesis is based on six main publications (P1–P6). I am the first author on P1–P3 and P6, and co-author on P4–P5. Detailed author contributions are listed below ¹.

P1: Indoor Silent Object Localization using Ambient Acoustic Noise Fingerprinting

I was responsible for Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review & Editing, and Visualization.

P2: Performance Comparison of Omni and Cardioid Directional Microphones for Indoor Angle of Arrival Sound Source Localization

I contributed in Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review & Editing, and Visualization.

P3: A Coherent Wideband Acoustic Source Localization Using a Uniform Circular Array

My work involved Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing-Original Draft, Writing-Review & Editing, and Visualization.

¹According to CRediT author statement from ELSEVIER [13]

P4: Improving Cardiac Auscultation Signal Quality by using 4-Channel Stethoscope Array

My role was in Conceptualization, Validation, Investigation, Resources, Writing-Original Draft, and Writing-Review & Editing.

P5: Accuracy Impact of Increased Measurement Quality when using Pretrained Networks for Classification

My contribution included Conceptualization, Methodology, Investigation, Writing-Original Draft, Writing-Review & Editing, and Visualization.

P6 (Accepted): Silent Object Localization on a Grid Using RIR-Based Acoustic Fingerprints and EfficientNet

My work involved Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Resources, Writing-Original Draft, Writing-Review & Editing, and Visualization.

1.6 Outline of the Thesis

This thesis is organised as follows: Chapter 1 introduces the aim, problem area, purposes, scope, and contributions. Chapter 2 provides background concepts for indoor acoustics, DoA estimation, fingerprinting, acoustic classification, and measurement quality. Chapter 3 discusses prior-art based on the related concepts and highlights the knowledge gaps in the field. Chapter 4 outlines the research methodology in each projects, including experimental verification, describe the datasets (real and simulated) and method used throughout the thesis projects. Chapter 5 presents the results per purpose and research question. Chapter 6 is a discussion covers reflections, limitations, research outcomes, originality and novelty, significance and impact, risks and ethical aspects, and targeted directions for continued research. Chapter 7 concludes the thesis with chronological progression and cross-links between publications, a summary by per purpose is presented, key contributions and knowledge gain are mentioned. After the comprehensive summary (Chapter 1 through 7) the bibliography is given.

Chapter 2

Theoretical Background

Sound source localization (SSL) and acoustic classification rest on three tightly coupled layers: propagation (how signals travel in rooms), sensing (what microphones or arrays measure), and inference (how algorithms convert measurements into positions or labels).

From a propagation standpoint, indoor environments introduce multipath, reverberation, spectral coloration, and time variation (opening and closing doors, moving people). On the sensing side, choices of microphone directivity, array geometry, synchronization and calibration directly shape spatial resolution, sidelobe behavior, and robustness towards additive noise such as heating, ventilation, and air conditioning (HVAC). For inference, classical geometric methods such as steered response and generalized cross correlation with phase transform weighting (GCC-PHAT), emphasize time-delay cues and are often favored for their simplicity under reverberation. Subspace methods, for example multiple signal classification (MUSIC), exploit spatial covariance structure for high-resolution direction-of-arrival (DoA) estimation, while wideband formulations use interpolation or modal processing to gather information across frequency. In parallel, learning-based approaches such as convolutional neural network (CNN) and convolutional recurrent neural network (CRNN) have become standard in acoustic scene and event work. A complementary thread is fingerprinting, where locations are inferred by matching measurements to a database of site-specific signatures rather than solving geometry explicitly. Acoustic fingerprints may be passive (ambient sound) or active (known probes), and can use spectral and/or phase features.

Across these layers, measurement quality often limits system performance more than algorithmic choice. This chapter presents the concepts which are covered by the thesis study over the problem area. More detailed prior-art study is presented in Chapter 3, and comparative method choice referring with prior-art is discussed in Chapter 6 section 6.1.2.

2.1 Acoustic propagation basics

Indoor sound propagation contains rich signal information for the localization and classification method [7, 14]. A path from the sound source to the receiver typically contains a direct component, early reflections, and late reverberation [10, 15, 16]. The room impulse response (RIR) generated by such propagation affects both the magnitude spectrum and the phase spectrum, and varies with source–array geometry (multiple signal) and surface materials [17–19]. These effects influence the time-difference and phase-difference that many DoA and fingerprinting methods exploit. Furthermore, they set practical limits on resolution and stability across time [6].

2.2 Acoustic Fingerprinting

Acoustic fingerprinting localizes by matching measurements to a database of specific location references as signatures rather than solving geometry explicitly [20]. Thus, it can be seen as an identification or classification approach for region recognition. Signatures may be extracted from ambient recordings passively or from responses to a designed excitation actively, and may include spectral magnitudes, phases, or higher-level features [8, 11].

The definition here is that passive localization schemes minimize instrumentation complexity, no active acoustic signal, but are sensitive to dynamic environments [21], while active localization schemes usually involve an acoustic signal, and can offer better repeatability. Both are grid-based or region-based localization that consists of measurement at each grid point [22]. In active schemes, multichannel room impulse responses (RIRs) across frequencies can be treated as an ‘acoustic CSI’ tensor, analogous to radio frequency channel state information (CSI), and used as rich fingerprints.

2.3 Channel State Information (CSI)

In radio-based localization, CSI refers to complex channel coefficients measured per subcarrier and antenna, which describe how a transmitted signal is filtered by the environment and the hardware chain [22–24]. Compared to Received Signal Strength Indicator (RSSI), CSI retains phase and frequency-selective fading structure, which makes it a richer fingerprint of the propagation conditions. Modern indoor positioning systems often treat CSI matrices as high-dimensional feature maps and use machine learning to regress positions or classify grid cells [22, 25].

A similar idea can be formulated for acoustics. Multichannel RIRs between a probe source and a microphone array encode how sound propagates along direct and reflected paths [17, 18]. If these responses are transformed to the frequency domain, each microphone has a frequency response that varies across frequency and location. When such responses are measured on a spatial grid and stacked across

microphones, frequencies, and possibly time frames, they form an “acoustic CSI” representation, which can be seen as a multi-dimensional description of how the room, the source, and the microphone array jointly shape the recorded sound.

Passive fingerprinting implicitly uses a degraded version of this information, since the ambient sound already carries the imprint of the room but with unknown excitation and limited control over SNR [8, 11]. Active schemes that employ known probes, such as swept-sine excitation and deconvolution, recover the RIRs more directly, hence provide better-conditioned acoustic CSI for learning-based localization. Treating these RIR-derived feature tensors as acoustic CSI makes the role of measurement quality explicit: stability of the microphone setup, excitation, and room conditions directly controls how consistent and discriminative the fingerprints are, which is central to the reliability of indoor acoustic sensing systems.

2.4 Array-Based Estimation for DoA

In array-based localization, the core task is to infer the direction from which an acoustic wavefront arrives at a set of microphones. For most methods this starts with time delay estimation (TDE) between sensor signals and ends with a direction-of-arrival (DoA) estimate obtained by combining those delays with the known array geometry [3, 26–30].

Under a far-field plane-wave assumption, the difference in arrival time between two microphones reflects the path-length difference to the two sensors and is the basic cue for direction finding [5, 31, 32]. For a fixed pair of microphones, each possible time difference of arrival (TDOA) corresponds to a set of source directions consistent with the distance between the sensors and the speed of sound [5, 31, 32]. With an array of more than two microphones, multiple TDOAs across different pairs can be combined so that only one direction satisfies all the delay constraints, yielding a DoA estimate [33, 34]. In this sense, TDOA is an intermediate quantity computed from the signals, whereas DoA is the final geometric estimate of the source direction.

Practically, TDE and TDOA are obtained by measuring how similar two sensor signals are when one is shifted in time. Cross-correlation is the standard tool for this, originally formulated in the time domain but often implemented in the frequency domain for efficiency and to allow frequency-dependent weighting [35, 36]. The generalized cross-correlation (GCC) framework introduces such weightings, for example phase transform (PHAT) weighting, which normalizes by the cross-spectrum magnitude to emphasize phase information [1, 37]. Steered-response power (SRP) methods extend this idea by summing evidence over many microphone pairs or over the full array for each candidate direction [35, 36].

Subspace and beamforming methods often describe the array response to a plane wave from a given direction using a steering vector or array manifold. A steering vector describes, for each microphone in the array, the relative phase and amplitude that an ideal source from a given direction would produce. In other words, it is a compact way of encoding how a plane wave from that direction should appear

across all channels. DoA estimators such as multiple signal classification (MUSIC) search for directions whose steering vectors are most consistent with the measured spatial covariance structure [5, 32]. Beamforming-based approaches, such as SRP and minimum variance distortionless response (MVDR), can be interpreted as scanning the output of direction-dependent spatial filters constructed from these steering vectors [2, 3, 38].

As the word “estimation” already tells there will be errors, factors that affect TDE and TDOA (e.g noise, reverberation) will also affect the inferred DoA in the following step. Different weighting functions and processing strategies have therefore been developed over the years to improve robustness, especially in reverberant rooms and at low SNR. Table 2.1 summarizes several common methods and their characteristics. Thus, DoA pipelines often apply frequency-domain weighting before inverse transforming, to emphasize reliable frequency bands and to mitigate noise and reverberation.

2.5 Beamforming

Beamforming treats a microphone array as a spatial filter. Sensor signals are delayed, weighted, and summed so that sound from a desired direction is preserved while noise and interference from other directions are attenuated [2, 3]. In this thesis, beamforming is used in a broadband acoustic setting and implemented in the short-time Fourier transform (STFT) domain. For each time frame k and frequency bin f , the M microphone signals are collected in a vector $\mathbf{x}(k, f) \in \mathbb{C}^M$. A linear beamformer for the incident angle $\boldsymbol{\theta}$ then produces

$$y(k, f; \boldsymbol{\theta}) = \mathbf{w}(\boldsymbol{\theta}, f)^H \mathbf{x}(k, f), \quad (2.1)$$

where $\mathbf{w}(\boldsymbol{\theta}, f) \in \mathbb{C}^M$ are complex weights and $(\cdot)^H$ denotes Hermitian transpose. The weights are typically designed using the steering vector $\mathbf{a}(\boldsymbol{\theta}, f)$, which describes the phase and gain pattern of a plane wave arriving from $\boldsymbol{\theta}$ at frequency f [3]. In broadband applications, the beamformer output is computed per frequency bin and then combined across frequency, either by inverse STFT (for reconstruction) or by summing power across bins (for localization measures).

One of the design is delay-and-sum beamforming (DSB), where each channel is time-aligned to the incident angle and then averaged. In the STFT domain this can be expressed as

$$\mathbf{w}_{\text{DSB}}(\boldsymbol{\theta}, f) = \frac{1}{M} \mathbf{a}(\boldsymbol{\theta}, f). \quad (2.2)$$

DSB is easy to implement, robust, and widely used in embedded systems, but its angular resolution and sidelobe levels are limited by the array aperture and geometry [2, 27]. Many SRP methods, including SRP-PHAT, can be interpreted as scanning the output power of a DSB-type beamformer across a grid of candidate directions and aggregating this power over frequency using different frequency-domain weightings [35, 36].

Table 2.1: Terms of the localization methods and their characteristics

Method	Characteristics
TDE	Estimates the relative arrival-time delay between two sensors from their signals, typically via cross-correlation [3, 26].
TDOA	The estimated time offset in seconds or samples that serves as a basic cue to infer the incident angle of the source when combined with array geometry and speed of sound [3, 26, 33, 34].
AOA / DOA	Source direction inferred from TDOAs and known array geometry. Could also be obtained by geometric methods, beamforming, or subspace techniques [5, 31, 32].
GCC	Frequency-domain cross-correlation framework that applies a weighting to the cross-spectrum before inverse transforming, improving robustness in noise and reverberation [1, 35, 36].
PHAT	A GCC weighting that normalizes by cross-spectrum magnitude to emphasize phase cues and reduce sensitivity to spectral coloration [1, 37].
SRP	Beamform-and-scan approach, which sums pairwise (or array) evidence over an angle grid (e.g., SRP-PHAT uses PHAT weighting) [35, 36, 39].
MUSIC	Subspace-based high-resolution DoA estimator, which exploits the spatial covariance structure to separate signal and noise subspaces and searches over candidate steering vectors [5, 32, 40].
MVDR	Adaptive beamformer that minimizes the output noise-and-interference variance subject to a distortionless constraint in the desired look direction. In theory, this uses the noise covariance matrix, while practical implementations often approximate it with the total covariance [2, 3, 38].

Adaptive beamformers further exploit the spatial covariance of the received microphone signals. The MVDR design chooses weights that minimize the output noise-and-interference variance while keeping a distortionless response in the desired direction [3]. In its theoretical form, MVDR at frequency f is written using the noise covariance matrix $\mathbf{R}_n(f)$:

$$\mathbf{w}_{\text{MVDR}}(\boldsymbol{\theta}_0, f) = \frac{\mathbf{R}_n(f)^{-1} \mathbf{a}(\boldsymbol{\theta}_0, f)}{\mathbf{a}(\boldsymbol{\theta}_0, f)^H \mathbf{R}_n(f)^{-1} \mathbf{a}(\boldsymbol{\theta}_0, f)}, \quad (2.3)$$

which minimizes the output noise variance subject to the distortionless constraint $\mathbf{w}^H \mathbf{a}(\boldsymbol{\theta}_0, f) = 1$. In many practical acoustic scenarios, a separate noise-only estimate is not available, and the total covariance of the received signals is used as an approximation, which leads to a minimum power distortionless response formulation often still referred to as MVDR in the literature [2, 3]. This family of designs can sharpen angular selectivity and suppress interference, but is sensitive to covariance estimation errors and model mismatch. Superdirective and constrained MVDR variants push this trade-off further, improving low-frequency directivity at the price of increased sensitivity to sensor mismatch and noise [3, 41].

Array geometry and microphone characteristics play a central role in beamforming performance. Linear arrays are attractive for their simplicity, but their response is strongly dependent on the chosen incident direction and elevation [3]. Uniform circular arrays (UCAs) admit a modal-domain formulation, where beamformers are built in the circular-harmonic domain and then projected back to the sensors [3, 41]. This geometry-aware processing can provide frequency-independent beam patterns over a useful range and enables efficient combinations with subspace methods or deep learning models [42]. In all cases, beamforming is tightly coupled to measurement quality: microphone directivity, calibration, SNR, and array layout determine how well the spatial filter can suppress noise and interference, and thus how informative the beamformed signal is for downstream localization or heart-sound classification tasks.

2.6 Deep learning in audio measurements

Machine learning systems learn patterns from data rather than being specified only by hand-designed rules [43]. Classical machine learning in signal processing typically relies on two stages: first extracting features from the raw signals, and then training a separate classifier or regression model on those features [23, 43]. Deep learning (DL) extends this idea by using multi-layer neural networks that can learn both the feature representation and the decision function jointly from examples [25, 44]. Convolutional, recurrent, and hybrid convolutional–recurrent (CRNN) architectures are now common choices in audio tasks [7, 12, 44].

Before DL became standard, many audio pipelines used engineered descriptors such as Mel-Frequency Cepstral Coefficients (MFCCs) together with classical classifiers like Gaussian mixture models or support vector machines [23, 45]. These approaches can work well when the features capture the relevant structure, but they

require substantial manual design and often struggle when conditions differ from those seen during development [23, 43].

In the audio applications considered in this thesis, DL is used mainly for classification and regression tasks. A basic example is audio classification, where the goal is to assign a label such as “speaker,” “room,” or “murmur” to a recording or to a short time frame [46]. Community challenges in sound event detection and localization have shown a shift from hand-crafted features to DL-based systems that operate on time–frequency representations such as spectrograms, learning both what happened and, in some cases, where it came from [7, 12]. For direction finding and localization with microphone arrays, DL models typically consume multi-channel time–frequency maps and learn to exploit inter-channel phase and amplitude differences as spatial cues [42, 47, 48]. In challenging urban audio, convolutional neural networks on spectrograms have been used to denoise, detect alarms, and regress source direction under heavy noise and reverberation [4, 16].

In biomedical acoustics, DL models have been applied to heart sound recordings to segment S1 and S2, assess signal quality, and classify normal versus pathological patterns [46, 49, 50]. In this setting the measurement front end (for example, a multi-channel stethoscope with beamforming) and the DL classifier are tightly linked, because the network inherits the noise, artefacts, and variability present in the recorded signals. For fingerprinting and indoor localization, DL has also reduced survey effort by learning location-dependent structure from radio CSI and related features [22–24]. The acoustic fingerprinting work in this thesis follows a similar pattern, but applied to multichannel room responses instead of radio channels.

Overall, DL provides flexible models that can absorb complex measurement distortions and spatial patterns, but their performance and reliability remain strongly dependent on the quality and diversity of the underlying audio measurements [25, 43, 44, 49]. This dependence motivates the thesis focus on measurement quality, array design, and data conditions alongside the choice of network architecture.

2.7 Cardiac auscultation

Cardiac auscultation listens to heart sounds: the first and second heart sounds (S1, valve closure at systole onset and S2, valve closure at diastole onset) carry most energy at low frequencies (roughly 20–150 Hz), while murmurs may extend into several hundred hertz. The Figure 2.1 highlights the main auscultation landmarks ¹.

S1 is heard the loudest at the apex over the mitral and tricuspid areas, S2 is loudest at the base over the aortic and pulmonic areas, and Erb’s point S1 and S2 have similar prominence. Those sites act as anatomical anchors for classifying heart sounds and murmurs by location and timing, as summarized by [50]. Recordings are easily contaminated by respiration, movement, and ambient noise [4, 54–56]. Mod-

¹Figure material is a reproduction modified with the mixture images from human body in [51], its derivative work by Ickle and Vinne2 in [52], and the heart illustration from [53]

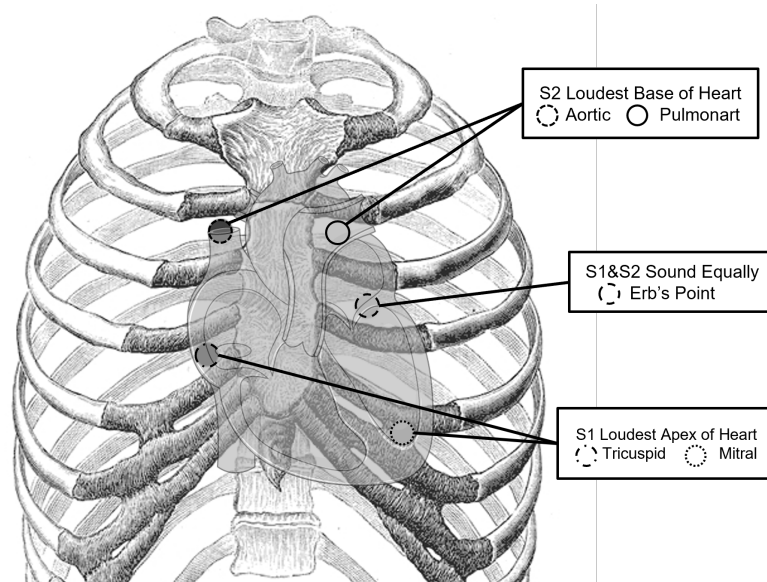


Figure 2.1: Auscultation key aspect picture illustration for heart sound and murmur classification criteria.

ern DL contributes through improving segmentation of S1 and S2 and of systolic and diastolic intervals. By providing signal quality assessment and denoising, DL can improve classification of normal vs. abnormal heart sound patterns, as demonstrated by [46], and extended by [49, 50].

2.8 Measurement quality & sensors

Accurate localization and classification depend as much on how audio is measured as on the algorithm that follows, because SNR, microphone directivity, array aperture and geometry, channel synchronization, and device response all determine the usable spatial and spectral evidence [41, 57, 58]. Differences in capsule response or small gain or phase drifts change the effective array manifold, so simple, documented calibration and repeatable procedures are important for reliability (e.g., level and phase checks, delay alignment) [59, 60]. Community datasets such as LOCATA were created to standardize sensors, layouts, motion protocols, and reporting, which makes results verifiable and comparable across studies [9, 10]. Learning-based systems inherit the statistics of their measurements, as explained by [61]. As reviewed by [25], accuracy can degrade when SNR, device traits, or room conditions shift. Evaluations should therefore cover realistic variations and report the measurement context alongside model details, as recommended by [12]. Taken together, careful sensor choice, calibration, and transparent reporting are essential for reliable acoustic sensing with microphones and form one of the main themes of this thesis.

Chapter 3

Related Works

A broad body of research has addressed acoustic source localization and classification, motivated by applications ranging from robotics and smart cities to healthcare and human-machine interfaces. In the following, the literature is reviewed along the three research topics defined in Chapter 1: silent object fingerprinting for indoor localization (Topic A), array-based sound source localization with compact microphone arrays (Topic B), and biomedical applications in cardiac auscultation with measurement quality for heart-sound analysis (Topic C). For each area, prior works are considered, remaining knowledge gaps are identified, and the positioning of the present research (P1–P6) within these gaps is clarified, with particular emphasis on how microphones and measurement design shape what can be reliably inferred from sound under real-world noise.

3.1 Acoustic fingerprinting

Fingerprinting has long been explored for indoor positioning, typically in wireless and acoustic domains. Traditional works using antenna arrays such as Wi-Fi based approaches rely on received signal strength indicator (RSS/RSSI), or on spectral signatures or CSI features [8, 24, 62–64].

Early acoustic fingerprinting established feasibility and addressed how to build and match different databases under reverberation [11]. This was extended to multidevice self-localization using audio fingerprints [21]. Later, room and background acoustic-signatures and auditory scene features were used for room or sub-area discrimination [65]. More recently, fingerprinting has evolved with machine learning and deep learning techniques [66], but robustness to noise and scalability remain challenges. Surveys from [6, 8] also pointed out that fingerprinting approaches often require heavy offline training and are sensitive to environmental changes. With microphone arrays, voice and acoustic fingerprinting has been paired with modern classifiers to improve robustness in typical homes [67, 68].

A common operational form is grid-based classification: the area is partitioned into cells, a labeled fingerprint is collected for each cell, and inference predicts the most likely cell (optionally followed by temporal smoothing or majority voting). Classifiers range from k-nearest neighbors (kNN), support vector machines (SVM), trees and boosting (e.g., multiple weighted decision trees) to CNNs that ingest magnitude–phase spectrograms or CSI tensors, trading survey effort and grid resolution against accuracy [8,24,62,67,68]. Trajectory-aware decoding can stabilize predictions under orientation changes and day-to-day variability [8,24]. When the search space is large, grid search can be accelerated (e.g., heuristic pruning or quantum-assisted search over candidate grid points), though this presumes a reliable fingerprint–cell mapping [63,69]. Robustness remains coupled to grid granularity, re-survey needs, and environmental drift, motivating designs that improve distinctiveness per cell (array features, phase cues) while keeping acquisition practical [8,24,67,68,70].

In this thesis, P1 [71] establishes passive, single-microphone ambient-noise fingerprinting, while P6 explores array-aided, deep learning based fingerprinting with active excitation. Together, these studies respond to the identified gaps by explicitly linking microphone configuration, excitation choice, and room conditions to fingerprint robustness, rather than treating the fingerprints as abstract feature vectors. They form part of the first research topic on indoor acoustic localization, and illustrate how moving from minimal hardware to structured array capture changes what can be reliably inferred about position under real-world noise.

3.2 Direction of Arrival Estimation

DoA, also known as angle of arrival (AoA), estimates the direction from which a wavefront reaches a sensor array. The same mathematical tools appear in radar, sonar, and wireless communications, but this thesis focuses on acoustic DoA with microphone arrays for indoor environments [2,6,14,40].

Early and influential work on generalized cross-correlation with phase transform weighting (GCC–PHAT) showed how robust time-delay estimates can be obtained from microphone pairs in reverberant noise [1,26]. Building on the same principle, steered-response power approaches such as SRP–PHAT treat the array as a delay-and-sum beamformer and scan over a grid of candidate directions, looking for peaks in summed energy [2,27]. These methods are widely used in room-acoustic localization, but surveys emphasise that their resolution is limited by array aperture and that many studies report results under fairly controlled conditions or with large arrays [2,6,28].

Beyond delay-based techniques, subspace methods such as MUSIC form a spatial covariance matrix, estimate its signal and noise subspaces by eigen-decomposition, and then search for steering vectors that are orthogonal to the noise subspace [5]. In free-field or mildly reverberant conditions these methods can deliver very fine angular discrimination [6]. However, in strongly reverberant rooms, coherent reflections reduce the effective rank of the covariance matrix, which can cause narrowband subspace DoA estimators to degrade unless additional processing such as spatial

smoothing is used [2, 6, 28]. For wideband signals, methods like coherent subspace focusing and array manifold interpolation (AMI) first align the array responses at different frequencies to a common reference, and then combine them. This lets the estimator use information across frequency while still behaving like a narrowband high-resolution subspace method [72, 73]. Reviews note that such coherent wideband schemes can be comparatively robust to reverberation, but they increase computational and calibration demands [2, 28]. Adaptive beamformers such as MVDR offer another route, trading sharper angular selectivity for sensitivity to covariance-estimation errors and model mismatch [2, 38, 74].

Due to these methods assuming sensor arrays, array geometry and microphone characteristics play a central role. Uniform circular arrays (UCAs) support circular-harmonic (modal) processing, where microphone signals are transformed into angular modes that naturally encode azimuth structure for scanning and beamforming [3, 75, 76]. In this modal domain, constrained MVDR designs provide geometry-aware processing with sidelobe control and robustness mechanisms [41]. Practical implementations such as SoundCompass [77] and the LOCATA challenge [9, 10] demonstrate how microphone arrays can be deployed for localization tasks, yet often require large apertures, favourable mounting, or controlled environments. Other advanced wideband and modal formulations [32, 78] have shown potential but still face limitations in real-world scalability and in how clearly they characterise the impact of array geometry and microphone type.

Within this landscape, the main gaps relevant to this thesis concern the role of microphone type, compact array geometry, and coherent wideband processing under realistic acoustic conditions. P2 [79] addresses part of this gap by fixing a representative high-resolution estimator and experimentally isolating the influence of planar array layout and microphone directivity (omnidirectional versus cardioid) under controlled indoor measurements. P3 [80] complements this by focusing on a small UCA and a coherent wideband pipeline based on circular harmonics and modified AMI, demonstrating how UCA structure can be exploited to achieve MUSIC-like resolution at practical bandwidth and computational cost. Together, these studies contribute to the second research topic on array-based localization, and help bridge the gap between algorithm-centric DoA formulations and their behaviour on measurement-constrained microphone arrays.

3.3 Acoustic classification with heart sound

Another important research strand concerns classification of biomedical and environmental sounds. Reviews such as [4, 6, 28] underline the potential of machine learning in improving recognition performance, but also highlight issues of measurement quality and sensor configuration. In cardiac signal analysis, prior works mainly focused on single-sensor stethoscopes with limited robustness to noise.

Traditional cardiac auscultation relies on acoustic stethoscopes with a passive chest piece and air-tube transmission. The diagnostic cues of such devices include the first and second heart sounds (S1, S2) and systolic/diastolic murmurs that mainly

occupy sub-kHz bands (often < 500 Hz for principal components), but practical use is affected by placement, contact pressure, ambient noise and user variability [54]. Electronic or digital stethoscopes add on-board amplification, filtering and recording, and often adopt contact sensors (e.g., piezoelectric discs) to capture body vibrations at low frequencies while reducing airborne noise. However, handling noise and device response still shape signal quality and repeatability [6, 54]. With digital recordings available, early pipelines used hand-crafted time–frequency features and sequence models, for example hidden semi-Markov models (HSMMS) for robust S1/S2 segmentation, establishing baselines for quality assessment and pathology classification [12, 54]. Recent work has shifted toward deep learning (CNN/CRNN) on spectrogram inputs, mirroring trends in audio event detection and localization where learned representations replace manual features [7, 12].

Remaining challenges include sensitivity to measurement quality (SNR, device response, placement and pressure variability), limited diversity of labeled datasets compared with clinical breadth, and domain drift across devices or environments, issues repeatedly emphasized in audio and measurement literature [6, 7, 12, 28]. These gaps motivate designs that pair robust hardware (e.g. arrays or contact sensors and low-noise front-ends) with standardized measurement protocols and models that explicitly account for acquisition variability [6, 28]. P4 [81] contributes to this third research topic by introducing a four-channel stethoscope array for spatial filtering and reporting measurable improvements in SNR for auscultation tasks. This measurement-driven approach illustrates how array concepts from DoA research can directly benefit classification problems in biomedical acoustics, while also revealing practical limitations in dataset size, pathology diversity, and long-term stability that motivate future work.

3.4 Measurement quality and sensor factors

Across sensing domains, system accuracy is bounded by measurement quality: SNR, sensor frequency response and directivity, array aperture and geometry, calibration and synchronization, placement, and environmental variability. Reviews in indoor positioning and audio consistently note that these “front-end” factors shape achievable error more than algorithmic tweaks when conditions are realistic [6, 8, 14, 28, 62].

Long-term drift (furniture moves, HVAC cycles, occupancy) challenges fingerprinting and scene models unless re-survey or adaptation protocols are in place [82–84]. For array processing, microphone pattern (omni vs. directional), mounting, and spacing determine spatial resolution and sidelobe behavior, while device response and clocking affect comparability across sessions and sites [2, 27, 85]. In other words, device and environment generalization still affect DoA and classification performance greatly despite algorithmic advances [74, 78, 85]. For circular arrays, modal beamforming and constrained MVDR provide geometry-aware control of sidelobes, but again depend on calibrated steering vectors and stable device responses [76]. Similar concerns regarding measurement dependencies also appear in radio fingerprinting. For instance, device or system differences and temporal drift

force frequent re-survey unless the measurement protocol is stabilized [86]. More broadly, the robustness of machine learning models against degraded inputs remains an open issue [61]. Practical systems (e.g., an audio-wearable for vehicle detection on construction sites) show how measurement protocols and low-power array hardware must be engineered together to remain reliable in heavy noise [87].

Motivated by a need for joint design of measurement and method, combined with the dependencies mentioned earlier, this thesis includes dedicated studies of measurement quality. P2 [79] makes these dependencies concrete by isolating microphone directivity and array geometry under identical room conditions (linear, square, arc; omni vs. cardioid), showing how much DoA performance can change when only the measurement configuration is altered. P5 [88] complements this from a learning perspective by fixing a verified, pretrained CNN and systematically varying SNR. The study provides quantitative evidence on how these measurement changes affect accuracy of classification, and offers a reference point for the practical dilemma of whether to invest in collecting a completely new dataset for an existing model when earlier recordings were made with so-called “outdated” data that gathered by suboptimal setups. In contexts with privacy constraints, rare edge cases, or data that require long-term collection. The proposed method demonstrates how existing datasets can be used to extrapolate the future performance of a current model by simulating controlled degradations in data quality. This provides decision-making options before committing to the collection of a full dataset with optimal setups. Ultimately, this approach helps determine whether the potential performance gains worth a full cycle of model redesign, retraining, and fine-tuning.

3.5 Positioning of this thesis

The overall aim of this dissertation is to investigate how measurement quality in microphone-based sensing shapes what can be reliably inferred from sound under real-world noise. As described in Chapter 1 and illustrated in Figure 1.1, the work is organised into three measurement-driven research topics: silent object fingerprinting for indoor localization (Topic A), array-based sound source localization with compact arrays (Topic B), and cardiac auscultation for heart-sound analysis (Topic C). Each topic is framed by 2 specific research questions, but all share the same underlying concern: how do microphone choice, array geometry, excitation design, and recording protocols constrain or enable acoustic sensing in practice?

Across all three topics, the included publications P1–P6 position the thesis as a contribution to measurement-driven acoustic sensing. Rather than treating microphones, arrays, and recording protocols as background implementation details, the work treats them as central components of the scientific questions and results. The related works in this chapter are selected and discussed with this perspective in mind, so that the later chapters can build on a clear understanding of how the proposed methods and experiments sit within the broader field of acoustic localization and heart-sound classification under real-world noise.

Chapter 4

Methodology

This chapter presents the measurement-centered methodology that underpins the thesis contributions. The primary design principle is to answer the research questions stated in Chapter 1 through purpose-built measurement configurations and verification protocols. Each configuration was expressly chosen to answer RQ 1.1–1.2 (passive and DL-aided fingerprinting), RQ 2.1–2.2 (DoA array configuration and array-specific coherent wideband localization), RQ 3.1 (array-aided auscultation), and RQ 3.2 (data quality impacts on pretrained models). Choices of experiment design including microphone type, array configuration, simulation of data acquisition result, and evaluation metrics were made to both fill in the knowledge gaps mentioned in Chapter 3, and expose how measurement parameters affect outcome quality.

4.1 Proposed Measurement Methods

This section presents the detailed motivation of proposed measurement methods, explaining why certain choices were made during the project design. From gaps (Chapter 3) to solutions (this chapter): Silent object localization with acoustic fingerprinting using one microphone (P1) and exponential sine sweep (ESS) based classification with microphone array using deep learning (P6) address localization feasibility and robustness limitations [11,24]. Microphone-array-based DoA studies with real-world measurements in P2 and P3 address the lack of directivity- and geometry-controlled experiments in prior work [2,6,10]. Multi-channel auscultation and quality analysis (P4, P5) fill the missing link between measurement quality and classifier reliability [7,12]. Subsections are organised by research topic and their associated research questions.

4.1.1 Fingerprinting: From Passive to Deep Learning

Acoustic fingerprinting typically uses active beacons [14, 21, 28, 30, 89] while radio fingerprinting uses RSS, Wi-Fi, CSI etc. and often requires heavy site surveys [11, 62, 66, 86]. In addition, fully passive acoustic fingerprints using ambient sound remain less explored and are sensitive to dynamic environments [8, 11, 24, 90, 91].

RQ1.1 asks whether no-emission localization is feasible with minimal hardware for silent objects. A single condenser microphone maximizes feasibility and tests the lower bound of sensing complexity. By recording ambient noise at each candidate position, each defined as the center of a grid cell, and applying subtraction and normalization, the grid-based room signatures are extracted by emphasizing position-dependent acoustic absorption due to the object-room interaction. Matching in the frequency domain (with band-limiting to remove fluctuating unusable parts) provides a transparent, reproducible baseline that exposes the measurement contribution without conflating it with large models [11]. The design of P1 therefore verifies whether ambient acoustics carry sufficient discriminative information for different positions.

Due to many passive localization methods having restricted prerequisites on environmental conditions, RQ1.2 took the active probe approach and extended P1 by asking how a structured array combined with deep learning improves robustness and scalability for silent object localization using acoustic fingerprinting. Since P2 had shown that a simple four-microphone planar rectangular array provides a reasonable compromise between spatial coverage and reverberation sensitivity, P6 reused a four-microphone rectangular array to increase spatial diversity for silent object localization. Exponential swept-sine (ESS) was used to obtain repeatable, high-SNR linear room impulse responses and to reduce sensitivity to transient ambient sounds that limit passive fingerprints [92, 93]. Inspired by recent work on multi-channel CNNs for sound event localization and detection, which shows that CNNs can exploit inter-channel spatial cues from stacked time-frequency features [7, 48], P6 adopted a CNN architecture that consumes stacked magnitude and phase features of the room's linear impulse responses' spectrograms as input. Because survey effort is a central limitation in fingerprinting (well documented for RF/CSI in [22]), P6 evaluated whether the added modeling complexity can achieve the target accuracy on a denser labeled grid than in P1.

4.1.2 DoA Estimation with Microphone Arrays

Prior work and benchmarks often focus on ideal omnidirectional sensors, with large apertures and controlled conditions. In addition, fewer studies quantify sensor directivity and compact geometry trade-offs under practical indoor acoustics [2, 6, 10].

Cardioids are commonly adopted in practice for their off-axis attenuation, whereas many DoA studies (and classical models) assume omnis for analytical convenience. The directivity pattern alters the array manifold and the effective spatial filtering. Therefore, keeping the array size and aperture fixed and swapping only the capsule

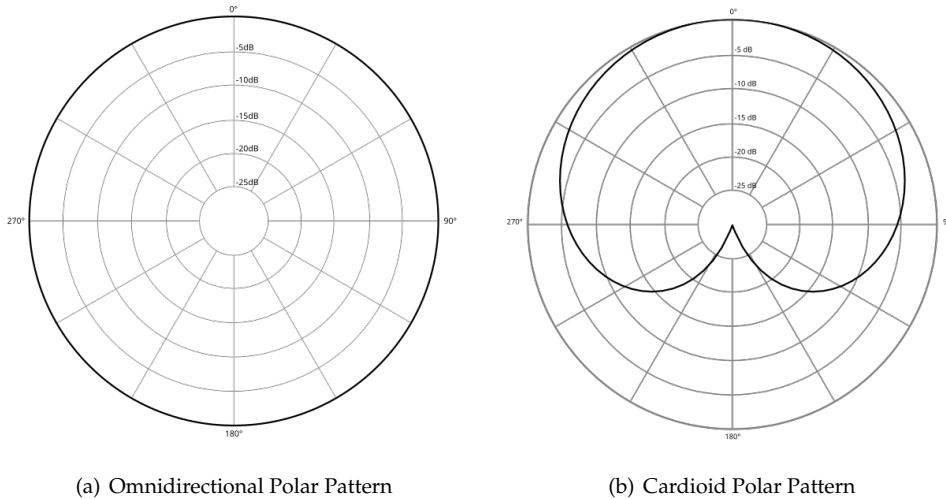


Figure 4.1: Directionality polar pattern of (a) omni-directional microphone (e.g Rode NT45-O capsule), and (b) cardioid-directional microphone (e.g Rode NT45-C capsule).

type isolates the measurement effect of directivity on DoA accuracy.

RQ2.1 focuses on how measurement configuration influences DoA estimation. To investigate this, P2 was designed as an experiment that systematically varies array geometry and microphone directivity. The microphones used during the experiment are Rode NT55 with its omni-directional original capsule NT45-O, and the interchangeable cardioid capsule NT45-C. Figure 4.1 shows a general illustration of two types of microphone directivity polar pattern [94, 95]. It also reflects field constraints—cardioids are frequently chosen to suppress diffuse noise and reverberation, while omnis are inexpensive and widely available [2, 6]. In exploring the efficacy of various microphone array configurations for indoor sound source localization, a true-experimental design has been deployed. This approach uses existing indoor environments to assess the impact of different microphone array configurations on localization accuracy and allows for practical evaluation within real-world settings. To deploy the microphone arrays, three popular array geometries were chosen [58, 96]: a planar 2D aperture (rectangular array), a curved aperture (quarter-circle, arc array), and a 1D aperture (linear array) with the same channel count. The four-microphone limit deliberately constrains hardware complexity and cost. A peak-selection weighting threshold was employed to mitigate reverberant false peaks in the steered response, scaled as a percentage of the highest peak per scan, enabling fair comparisons across capsules and geometries. Localization results were evaluated using mean error distance and standard deviation, computed per geometry and per microphone type. This yields a measurement-only view on the DoA impact of directivity and geometry (Topic B; cf. Chapter 3).

Motivated by the rotational symmetry and uniform bearing coverage of circular manifolds, the measurement in P3 used a uniform circular array for further research to address RQ2.2 with a coherent wideband method. A pre-designed UCA with 6 MEMS microphones was chosen to keep the aperture small and cost modest while meeting spatial-aliasing limits over the target band. To address the computational burden noted in coherent focusing literature [73], a UCA-specific focusing matrix was derived, which reuses circular-harmonics structure and avoids repeated Bessel calculations [5, 80]. Processing was structured to reduce computational demand, aiming to make embedded systems with real-time implementation at low frequencies feasible. Localization accuracy was evaluated using root-mean-square error (RMSE) in degrees rather than standard deviation, allowing a direct angular performance measure under wideband conditions.

4.1.3 Classification and Measurement Quality

Biomedical pipelines often optimize models while overlooking how multi-channel acquisition and input SNR support downstream accuracy [7, 12].

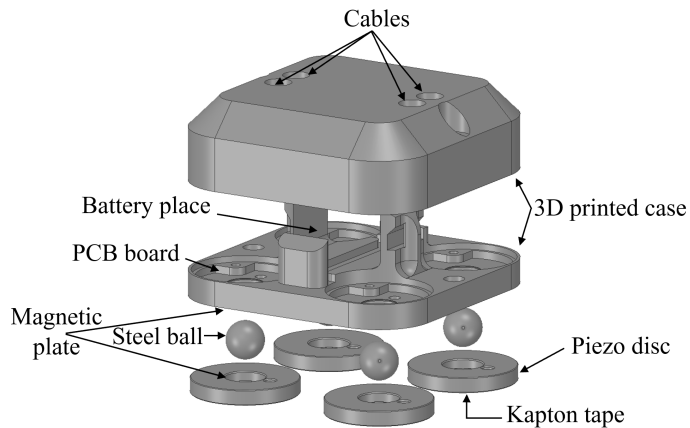


Figure 4.2: The illustration of a 3D model of designed multichannel stethoscope.

RQ3.1 asks whether multi-channel acquisition can improve the signal quality before classification. To address this research question, a design of a prototype digital stethoscope was reported in P4 with a four-piezo-disc sensor array, which balances contact area, patient comfort, and spatial diversity, see Figure 4.2 [81]. The battery supply and shielding provide a quiet electrical environment, and per-channel matching enables the idea of using delay-and-sum beamforming across the four elements. The preamp topology follows the typical low-noise high-impedance (Hi-Z) piezo front-end pattern (inspired by community designs for contact microphones in [97]). In aggregate, this yields array gain (SNR improvement) and the ability to emphasize the cardiac sound field while attenuating diffuse ambient noise, which provide a measurement-quality boost before classification. The following steps were deployed

to the prototype digital stethoscope: First of all, using DSB increases SNR for directional heart sounds. After that, a matched filter around S1/S2 further enhances periodic components prior to feature extraction. Thus, this measurement-first design improves the input that any classifier would receive, rather than relying on a more complex classifier to overcome poor SNR [98]. Evaluation uses clinician-annotated segments to connect signal enhancement to clinical interpretability. P4 originally used the “Dangerous Heartbeat Dataset (DHD)” hosted on Kaggle by user Mersico, which is a re-packaging of the PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) dataset [99] (classes: normal, murmur, extra heart sound, artifact, extrasystole). PASCAL is the abbreviation of “Pattern Analysis, Statistical Modelling and Computational Learning.”, which was an EU-funded Network of Excellence (NoE) in machine learning.

RQ3.2 asks how measurements taken at different times with different qualities can be utilized to extrapolate useful insights for future development beyond the original task. P5 uses fixed, pretrained CNNs to eliminate confusion from architecture changes and isolate the effect of acquisition quality. By augmenting the dataset with controlled noise levels, successive measurement-system upgrades are emulated. With the pretrained model held fixed, performance trends toward higher-quality acquisitions can be extrapolated, and the benefit of each upgrade can be estimated [19, 59, 88]. This provides an experimental basis for judging when it might be worthwhile to invest in new measurements for an already deployed model.

4.2 Experimental Verification

This section summarises how each research question was experimentally verified, grouping real-world recordings, simulations, and controlled-quality studies according to the topics in Chapter 1. The emphasis is on how the chosen measurement setups support the research questions, rather than on repeating all implementation details from the individual papers.

4.2.1 Silent object fingerprinting (RQ1.1–RQ1.2)

P1: To test RQ1.1 in a realistic but controlled setting, ambient noise was recorded with the silent object positioned at grid-based positions using a single condenser microphone in an empty room. After that, subtraction and normalization produced spectral absorption signatures that were matched against a reference set, with cross-validated trials to assess lateral localization feasibility without beacons [23, 25, 89, 100]. The silent object used during the measurement is a real human, standing inside a 320×350 cm room with 9 grids for such passive localization [71]. Localization results were quantified via a resemblance score computed from the average standard deviation of the difference between smoothed spectral fingerprint vectors. This experiment directly probes how much positional information can be recovered from ambient sound alone with minimal hardware.

P6: To extend fingerprinting to active, array-based sensing for RQ1.2, a tetragonal array was used for grid-based silent object fingerprinting localization, collected in target rooms. During each measurement, a short, band-limited exponential swept-sine (ESS) [92, 93] was played as excitation to obtain high-SNR, repeatable responses within a few seconds per location while keeping the sweep spectrum in a comfortable frequency range (approximately 200 Hz–15 kHz at 48 kHz sampling). The silent object used during the measurement is two moving paper boxes stacked vertically, placed at different positions on a 4×4 grid, giving $G = 16$ grid cells of size 80×80 cm that covered a 320×320 cm area around the array. Measurements were conducted in the same conference room as P2, but after renovation with additional sound insulation. The multichannel responses were converted into STFT-based magnitude–phase tensors of size $316 \times 799 \times 8$ (frequency \times time \times channel) and used as input to a modified EfficientNet-B0 classifier with an 8-channel input stem and a 16-way softmax output over grid IDs. The primary evaluation metrics were classification accuracy for grid ID, confusion matrices, and distance error as used in P1. Together with P1, this verification set shows how moving from single-microphone measurements to array-based measurements changes the achievable resolution and robustness in silent object localization.

4.2.2 DoA with microphone arrays (RQ2.1–RQ2.2)

P2: For RQ2.1, synchronized four-channel recordings were collected under identical room and source placement, so that geometry and microphone type were the only changing factors. Three geometries \times two capsule types were evaluated. Using hand-clap as signal makes the experimental sound source more diverse [101]. The enclosed environment of the measurement is referring to a conference room with the area of interest is 300×400 cm. Accuracies by distance error were reported per condition [79]. This design directly supports RQ2.1 by comparing DoA accuracy across array manifolds and directivities under matched acoustic conditions.

P3 (real-world): For RQ2.2, a pre-designed UCA with 6 MEMS microphones recorded signals from 4 known loudspeaker azimuths to validate a UCA-specific coherent wideband method comparing against baselines [73]. The signal used were real-world recordings playing via the loudspeaker including car engine, water fall, hovering drone, electric fan, and train engine. For the proposed method, the average estimation errors for each source type are compared with the original AMI [80]. These experiments test whether the coherent wideband processing pipeline maintains high angular resolution on realistic acoustic content.

P3 (simulations): Monte Carlo trials modeled a single far-field source at a designated direction with band-limited, zero-mean Gaussian signals, SNR swept from -10 to 40 dB as in most of the literature. The proposed method was compared against iMUSIC, WS-TOPS, and the original AMI method over computational complexity, all of which avoid initial DoA pre-estimation. The primary evaluation metrics were angular RMSE (degrees) over trials. These simulations complement the real-world recordings by allowing controlled exploration of noise conditions and by checking whether the advantages of the coherent wideband UCA processing persist across a

realistic SNR range.

4.2.3 Biomedical application (RQ3.1–RQ3.2)

P4: For RQ3.1, a prototype digital stethoscope with a 4-channel piezoelectric array was built for the experimentation, aiming to improve signal quality by isolating and applying DSB beamforming to the systole and diastole part of the sound. The proposed algorithm was first tested with a dataset, which is “The PhysioNet/Computing in Cardiology Challenge 2016” training-b dataset, that contain real heart sound assessed with clinician auscultation annotations [102], and then verified with real auscultation of volunteered co-authors by the designed prototype digital stethoscope [81]. The performance was compared against single-channel baselines using clinician-annotated segments. This two-step verification links the array-based measurement design to both benchmark data and real recordings, and connects SNR gains to classification and interpretability outcomes.

P5: For RQ3.2, different SNR levels were augmented into the Dangerous Heart-beat Dataset (DHD) from Kaggle and tested on two pretrained models. By augmenting the dataset with controlled noise at varying SNR levels, the successive early upgrade-iterations of measurement systems under improved conditions effectively simulate by “degrading” the signal quality. This experimental design allows for the estimation of extrapolation trends—predicting how future upgrades in measurement quality would influence classification accuracy [19, 59, 88, 98]. In contrast to P4, which changes the hardware, P5 keeps the model fixed and manipulates the data, so that the specific impact of measurement quality on model behaviour can be observed and used to inform decisions about future data collection and model updates.

Chapter 5

Results

This chapter summarizes results across the included publications (P1–P6), organized by the two application areas and their research questions (RQs) as defined in Chapter 1. Detailed metrics, experimental protocols, and dataset specifics are provided in the individual papers. Here, the focus is on combined findings and patterns, and on how the outcomes relate to the measurement choices.

5.1 Integrated Results Across Problem Areas

The emphasis of this summary is on linking outcomes back to the measurement decisions motivated in Chapters 1–3, and on maintaining reproducibility through explicit sensor and geometry descriptions and shared processing steps. Throughout the thesis, one principle recurs: performance is tightly linked to how microphones and arrays are deployed, to the excitation and measurement protocol, and to how well the resulting data match the assumptions of the downstream method.

5.1.1 Indoor Spatial and Environmental Sensing

Addressing **RQ1.1** and **RQ1.2**, P1 and P6 together show a progression from minimal-hardware acoustic fingerprinting to array-aided, deep-learning-based fingerprinting classification.

In P1 [71], single-microphone ambient-noise measurements were used to build spectral fingerprints for a silent person standing on a 3×3 grid in a 3.2×3.5 m room. Subtraction and band-limited spectral comparison yielded feasible lateral localization of the silent object without active beacons or tags. The study reported that the method localized correctly for all tested positions within a radius of 29.9 cm around each reference grid point, and that the mean radial error was 12.8 cm when the person stood between grid centers (see Figure 5.1). In the original paper, this was described as “absolute position”, meaning the distance between the true position and

the nearest reference grid center. The main conclusion is that position-dependent absorption features in ambient noise can be sufficiently discriminative when spectra are stabilized by subtraction and frequency selection.

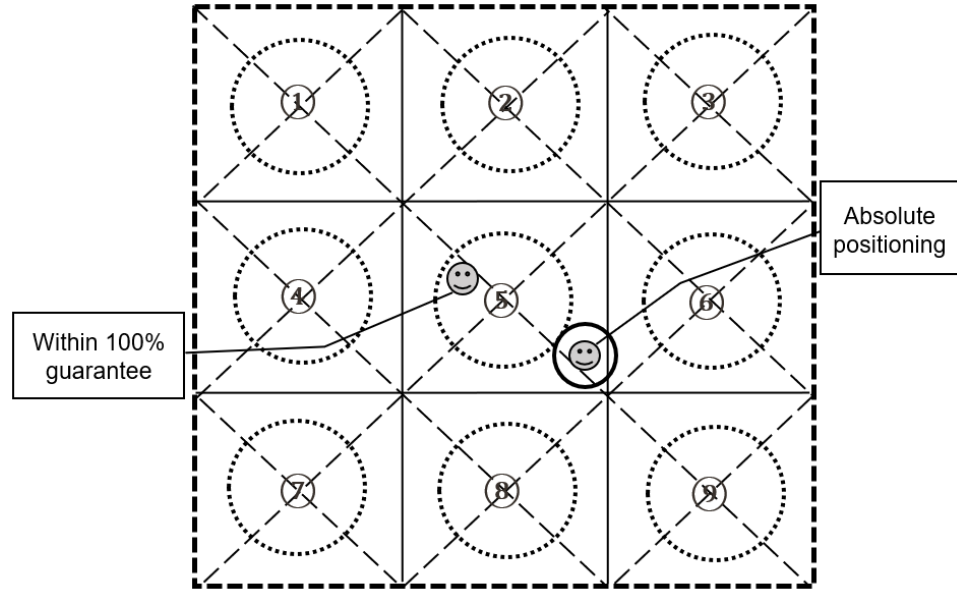


Figure 5.1: Illustration of P1 accuracy result with a radius of 29.9 cm from the reference position (inner dotted circle around the grid center) and a mean radial error of 12.8 cm (example solid circle) when the silent object stands between grids.

P6 extends this line of work and answers **RQ1.2** by moving to a structured four-microphone rectangular array and active probing with exponential swept-sine (ESS) excitation. Multichannel room impulse responses were measured on a 4×4 grid in a 3.2×3.2 m area of interest and converted into magnitude-phase tensors using a short-time Fourier transform. A modified EfficientNet-B0 CNN was trained on these tensors. In the reported experiments, grid-ID accuracy reached up to 97.7% and the mean distance between the predicted and true grid centers was below 11.8 cm. Compared with the single-microphone, passive baseline in P1, the spatial diversity and learned features in P6 improved robustness to small scene changes.

For **RQ2.1**, P2 [79] quantified how microphone type and array geometry affect indoor angle-of-arrival outcomes. Under identical room and source conditions, a four-channel microphone array was deployed in three geometries (linear, rectangular, and arc) using both omnidirectional and cardioid capsules. Classic AOA estimation with GCC and a proposed weighting factor were applied. The average error distances for each combination were reported. Before weighting, the linear array gave the lowest average error distance for both microphone types. For both rectangular and arc arrays, the difference in localization accuracy before and after applying the proposed weighting method shows that these two array manifolds are sensitive

to room reverberation. Overall, cardioid capsules yielded higher accuracy than omnis when combined with the proposed processing, and the study showed that array geometry and directivity can shift errors by more than a meter under the same room conditions.

P3 [80] addressed **RQ2.2** by implementing a uniform circular array (UCA) with six MEMS microphones and a coherent signal subspace method (CSSM) tailored to circular harmonics. A UCA-specific focusing matrix was derived that exploits the modal structure and avoids repeated Bessel function evaluations. In simulations with a single far-field source and SNR between -10 and 40 dB, the proposed method achieved slightly lower root-mean-square angular error than the original array manifold interpolation (AMI), while maintaining an advantage over iMUSIC and WS-TOPS at low SNR. In terms of computational effort, the focusing reduced processing time by up to 30% compared with AMI. Real recordings with different wideband source types confirmed that the proposed UCA-based pipeline preserves MUSIC-like resolution at modest computational cost and aperture, in line with the measurement-driven constraints targeted by the thesis.

5.1.2 Biomedical Sound Sensing

For **RQ3.1**, P4 developed a four-channel stethoscope array and applied delay-and-sum beamforming (DSB) for matched filtering around the S1 and S2 patterns to improve SNR in cardiac auscultation prior to classification. Compared with a single-sensor baseline, the array-based processing improved SNR and the perceived quality of heart sounds on clinician-annotated segments, which translated into more stable features for downstream classifiers. Figure 5.2 shows the sectional level gains obtained on real auscultation recordings, corresponding to Table I in [81].

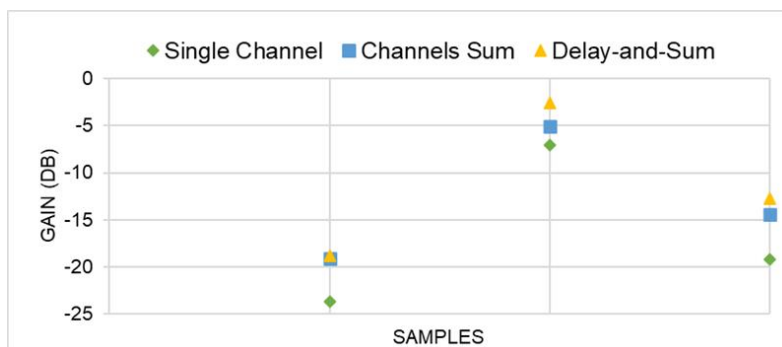


Figure 5.2: Illustration of P4 on real-world auscultation recordings, showing sectional level increases from beamforming. Three samples from co-authors are presented. Sample 2 (middle) has murmur condition, therefore during systole and diastole sound is occurring, translated into a sectional level closer to 0 dB, indicating stronger energy in the murmur-bearing region, while Samples 1 and 3 are healthy.

Table 5.1: Signal sectional level $L^{(s)}$ (dB) by processing stage.

Sample	Single	Channels Sum	Delay-and-Sum
1	-23.62	-19.05	-18.82
2	-7.09	-5.05	-2.57
3	-19.18	-14.41	-12.68

Table 5.2: Stage-to-stage gains and normalized dB change.

Sample	$\Delta L_{\text{Single} \rightarrow \text{Sum}}$	$\Delta L_{\text{Sum} \rightarrow \text{DSB}}$	$\Delta L_{\text{Single} \rightarrow \text{DSB}}$	Gain ^{log}
1	+4.57 dB	+0.23 dB	+4.80 dB	20.32
2	+2.04 dB	+2.48 dB	+4.52 dB	63.75
3	+4.77 dB	+1.73 dB	+6.50 dB	33.89

The strengthening of murmur-bearing intervals relative to S1/S2 across the three processing stages is summarized by the three legends in Figure 5.2: Single, Channels Sum, and Delay-and-Sum. For clarity, Table 5.1 reports the sectional levels, and Table 5.2 shows the derived gains.

For each stage s , the mean-square (power) is computed on the murmur sections ($P_S^{(s)}$) and on the S1/S2 baseline ($P_B^{(s)}$), and the sectional level in dB is defined as

$$L^{(s)} = 10 \log_{10} \left(\frac{P_S^{(s)}}{P_B^{(s)}} \right).$$

Levels are typically negative (as shown in TABLE I from [81]) because $P_S^{(s)} < P_B^{(s)}$; less-negative values indicate murmur in the signal, as the energy in that region is relatively close to the energy level of S1/S2.

Improvements between stages are reported as dB differences Δ :

$$\Delta L_{\text{Single} \rightarrow \text{Sum}} = L^{(\text{Sum})} - L^{(\text{Single})}, \quad (5.1)$$

$$\Delta L_{\text{Sum} \rightarrow \text{DSB}} = L^{(\text{DSB})} - L^{(\text{Sum})}, \quad (5.2)$$

$$\Delta L_{\text{Single} \rightarrow \text{DSB}} = L^{(\text{DSB})} - L^{(\text{Single})}. \quad (5.3)$$

For completeness, the original paper also reported a normalized dB change,

$$\text{Gain}^{\log} = \frac{\Delta L_{\text{Single} \rightarrow \text{DSB}}}{|L^{(\text{Single})}|} \times 100,$$

which expresses the dB improvement relative to the magnitude of the baseline level. This is a normalized dB ratio rather than a linear percentage in power or amplitude and is included here only for consistency with [81]; the physically meaningful metric is the dB gains ΔL .

In the intermediate “Channels Sum” stage, the four channels are combined without time alignment, so heart-sound events that occur at slightly different times across channels only partly reinforce, whereas noise contributions combine more fully. In the final DSB stage, the channels are time-aligned before combining, which increases the coherent summation of the heart-sound components and yields an additional improvement in sectional level. This confirms that multi-channel acquisition can serve as a first-step enhancement that benefits downstream classifiers regardless of the specific network choice, which is consistent with the measurement-driven focus of this thesis (cf. Chapter 3).

For **RQ3.2**, P5 [88] emulated the effect of different acquisition quality on two pre-trained CNNs for heart-sound classification. Different SNR levels were obtained by adding controlled noise to a fixed dataset, and the networks were evaluated without retraining. The study showed that a 1D CNN operating directly on raw audio is strongly affected by SNR changes. When test data were “upgraded” to higher SNR, classification accuracy moved close to that of a hypothetical network trained on high-SNR data. A 2D CNN operating on spectrograms was more robust to SNR degradation, but also showed smaller gains when SNR improved, which indicates that time–frequency features absorb part of the noise variability. Overall, P5 demonstrated that this evaluation pipeline can be used to approximate how future improvements in measurement SNR would translate into performance gains, without retraining the model or collecting an entirely new high-SNR dataset, thereby making fuller use of existing data when planning sensor or protocol upgrades.

5.2 Summary

Table 5.3 summarizes one illustrative result from each publication to highlight how measurement design and outcomes are tied together.

The three research topics A–C span two application areas and form a coherent progression. In indoor silent object localization, the results move from minimal single-microphone fingerprinting (P1) to ESS-based, array-supported deep learning fingerprints (P6), and from empirical comparisons of array geometries and microphone directivities (P2) to a UCA-specific coherent wideband method (P3). In biomedical sensing, the work progresses from single-channel heart-sound recordings to array-enhanced auscultation (P4), and then to an explicit study of how SNR and device response influence pretrained classifiers (P5). Across **RQ1.1–RQ1.2** (silent object acoustic fingerprinting), **RQ2.1–RQ2.2** (array-based DoA estimation), **RQ3.1** (heart-sound classification with improved acquisition), and **RQ3.2** (measurement-quality effects on pretrained models), the results consistently show that measurement design is a primary lever for robustness and accuracy, and that compact, deployable solutions benefit from both spatial diversity and explicit control of data quality.

Table 5.3: Illustrative headline results from P1–P6.

PaperID	Illustrative result
P1	Single-microphone ambient-noise fingerprints localized a silent person on a 3×3 grid with 100 % correct cell decisions within a 29.9 cm radius around each reference point and a mean radial error of 12.8 cm when the person stood between grid centers [71].
P2	In a conference room with the area of interest of 3.0×4.0 m, the cardioid linear array achieved the lowest average error distance (29.9 cm) after weighting, while rectangular and arc arrays improved from errors above 50 cm (and up to 237.8 cm) to 34.8–36.1 cm, demonstrating the strong impact of geometry and directivity on AOA-based localization [79].
P3	A UCA-specific coherent wideband method for a six-microphone uniformed circular array slightly reduced root-mean-square angular error compared with the original AMI and reduced computation time by up to 30 %, enabling MUSIC-like resolution with lower computational cost [80].
P4	A four-channel digital stethoscope combined with DSB increased murmur sectional levels by 4.5–6.5 dB relative to S1/S2 and yielded logarithmic gains between 20 % and 64 % in three representative real-world recordings, indicating clearer murmurs before classification [81].
P5	For a pretrained 1D CNN on heart sounds, keeping current data as the “highest SNR” while degrading different levels of SNR as “old dataset” emulated a system performance upgrading trend, while a 2D spectrogram-based CNN showed smaller accuracy changes, highlighting different sensitivities of model families to measurement quality [88].
P6	ESS-based acoustic fingerprints from a four-microphone array, processed by an EfficientNet-B0 CNN, achieved up to 97.7 % grid-ID accuracy and a mean grid-center distance error below 11.8 cm on a 3.2×3.2 grid for silent object localization.

Chapter 6

Discussion

This chapter steps back from the individual studies in P1–P6 and considers what they collectively say about the overall aim of the thesis: to view acoustic localization and classification through the lens of measurement. Rather than re-stating the results, the discussion reflects on how the chosen methodologies worked in practice—how validity and reliability were supported or strained, where alternative designs might have led to different conclusions, and how far the findings can be generalized beyond the specific setups and domains studied. From there, the chapter turns to the outcomes themselves and asks what they are good for, and where they might be misused, including the ethical implications of deploying measurement-based models in technical and biomedical contexts. Finally, the discussion situates the originality, significance, and impact of the work alongside its risks and limitations, including uncertainties, possible measurement biases, and ethical aspects of the research process. In this way, the chapter treats the thesis not only as a collection of answers to well-posed research questions, but also as a set of methodological and ethical choices that shape how those answers should be interpreted in the wider field.

6.1 Reflections on Selected Methodology

The methods chosen were designed to match the research questions and define the measurement factors: array configurations for direction of arrival (DoA), single microphone versus microphone arrays for fingerprint recognition, a dedicated stethoscope prototype with beamforming approach for measuring heart sound signals, and controlled noise for measurement quality investigation. Array geometry, microphone type, sampling, and error metrics remained fixed across all studies to ensure interpretable comparative results.

6.1.1 Validity and Reliability

In this thesis, validity was supported by aligning processing with the measurement setup. Magnitude-only passive fingerprinting in P1 tested the feasibility of ambient-based silent object localization. Array configurations in P2 investigated the role of directivity and geometry, results were reported respectively. Circular-harmonic focusing matrix in P3 matches the UCA array manifold. P4 and P5 both shows from an improvement of signal quality perspective, before classification, claims feasibility of the method based on data acquisition rather than model choice. Known excitation and multichannel magnitude with phase information in P6 increase construct validity for “location uniqueness”.

Reliability was supported through repeated captures, fixed grids and azimuths angles of the ground truth, synchronized channels, and accuracy report (average error distance by standard deviation in P2 and RMS angular error reported in P3). In these controlled conditions, the results show consistent performance across repetitions and sessions performed within the same setup, indicating good short-term reliability within each project. Where environment or human factors could affect reproducibility (room changes in fingerprinting; stethoscope placement), these were bounded operationally but not exhaustively re-tested over long periods. As a result, long-term reliability and robustness to major environmental changes remain only partially evaluated and should be confirmed in future studies.

6.1.2 Alternatives

There are alternatives to achieve fingerprinting localization with radio features (RSSI/CSI), which are mature and can deliver high accuracy [8,24]. These methods were not chosen as the main line of work here, partly because the thesis is centered on acoustic instrumentation and partly because radio pipelines introduce their own metrological issues, including device diversity, drift, and continuous survey upkeep. For the overall aim of this thesis, focusing on sound-based fingerprints made it possible to study how array geometry, excitation, and room acoustics together shape performance. At the same time, systematically contrasting RSSI/CSI-based fingerprints with the acoustic methods in P1 and P6 would have clarified which robustness limits are specific to acoustics and which belong to fingerprinting as a general strategy, and remains a natural extension.

An active acoustic approach with short, band-limited excitations (e.g., exponential swept-sine) can increase SNR and reduce contamination from foreground sounds, which is a well-established approach for robust room measurements [92]. In P1, these were deliberately not used; the goal was to test the limits of purely passive, ambient fingerprints for RQ1.1, even if this made the method more vulnerable to day-to-day variation. P6 then moved towards an ESS-based, actively probed design, but still with relatively short sweeps tuned to grid-level localization rather than full room impulse response estimation. Longer or denser excitation schemes could have produced more detailed channel responses and cleaner labels, at the cost of increased measurement burden and user intrusion, and might have shifted the bal-

ance between practicality and accuracy in a way that is relevant for large surveys or continuous monitoring. Phone-based active methods such as RoomSense and EchoTag [20, 103] represent another alternative. They rely on commodity hardware and are convenient to deploy, but their performance is strongly influenced by device placement and orientation, which is difficult to control in a metrological sense. These approaches were not adopted as the primary experimental platform because the thesis emphasizes controlled array design and repeatable setups rather than opportunistic sensing. Nevertheless, incorporating such phone-based schemes as a comparison point would have highlighted the trade-off between formal measurement control and the scalability and convenience that mobile devices offer.

In order to align with the thesis focus on acoustic measurement design, P1 and P6 answered RQ1.1 and RQ1.2 by examining how measurement choices in sound-based fingerprinting (excitation, array geometry, and recording protocol) influence downstream localization performance. Within acoustic methods more broadly, similar combinations of array-aided fingerprints and learned classifiers have been explored in home and voice localization, where they have shown promising results [11, 67].

For direction-of-arrival estimation, classical alternatives include GCC-PHAT and SRP-PHAT for robust delay-based localization, as well as subspace methods such as MUSIC and ESPRIT for higher resolution [1, 5, 35]. SRP-PHAT is attractive because it is robust under reverberation and conceptually simple, but on small apertures it trades angular resolution for robustness and can become computationally heavy when scanning dense spatial grids.

P2 did not attempt an exhaustive comparison of all these options. Instead, the study focused on a representative high-resolution DOA estimator and used it to compare different microphone directivities and planar array layouts in the same room, to test how measurement choices affect performance in realistic rooms for RQ2.1. A more extensive SRP-PHAT versus MUSIC/ESPRIT benchmark would have broadened the algorithmic picture, but at the cost of diluting the focus on geometry and microphone directivity, which directly answers research question RQ2.1 and is consistent with the objectives of the thesis.

In wideband processing, incoherent wideband MUSIC (often referred to as IMUSIC), where a narrowband MUSIC spectrum is computed at each frequency bin and then combined across the band, without introducing an explicit focusing matrix [32, 104]. This makes IMUSIC relatively straightforward to implement and numerically stable, but it also means that each frequency is treated largely independently, so small-aperture arrays may not fully exploit the fact that the source direction is common across frequencies. Coherent subspace methods, which apply a focusing or alignment step to enforce a shared signal subspace over frequency, can in principle achieve higher angular resolution or better robustness when suitable focusing transformations are available [32, 104]. Coherent approaches, such as CSSM and related focusing schemes, are better aligned with the goal of preserving high-resolution behaviour while reducing per-band search cost [31, 73, 104]. Modal MVDR beamforming is another appealing alternative, especially for UCAs, but it demands careful calibration and can be sensitive to model mismatch [3]. For uniform circular arrays specifically, modal-domain beamforming and constrained MVDR de-

signs provide geometry-aware processing that keeps sidelobe levels low, reducing sensitivity to interferers from other directions [3,41], and surveys summarise these trade-offs between complexity and robustness [2].

P3 adopted a UCA-specific coherent focusing matrix to maintain MUSIC-level resolution while reducing computational cost for embedded targets, in line with these trade-offs [2,3,41]. Exploring IMUSIC or fully established modal MVDR as primary methods could have shifted the emphasis towards either implementation simplicity or interference suppression for a given UCA, rather than on how well the circular geometry and its modal structure can be exploited in a coherent subspace framework. That alternative focus could have led to different conclusions about whether localization performance is mainly limited by the physical array or by how completely the processing pipeline makes use of its theoretical properties, which is central to the thesis theme of bridging array theory and measurement.

In biomedical sound analysis and sound-event-localization-and-detection (SELD) tasks, a common alternative is a model-first pipeline: single-channel recordings combined with denoisers, extensive augmentation, and increasingly complex neural architectures [7,12]. This route was not taken in P4, where the priority was a measurement design based on a four-channel stethoscope with delay-and-sum beamforming and matched filtering, so that the contribution of improved acquisition could be quantified before increasing model complexity. This choice aligns with the view that measurement quality is often the primary lever for downstream reliability [61]. A more model-centric path, starting from stronger baselines, larger architectures, or end-to-end learning, could have pushed raw classification scores higher and positioned the work closer to benchmark-oriented studies, but it would also have shifted the emphasis away from the measurement side and towards incremental gains in network design. Considering both strategies side by side in future work would clarify how far multi-channel measurement improvements can carry performance before they are overtaken by gains from more sophisticated models.

Similarly, there are classifier families that include explicit denoising or enhancement stages within the model, blurring the line between front-end processing and inference. Due to the purpose of RQ3.1, such a classifier with built-in denoiser is excluded from the study in P5, where the aim was to examine how changes in raw input data affect a fixed pretrained CNN, rather than to adjust the CNN itself. For studying measurement quality, an obvious alternative would have been to train new networks on very large, noisy dataset or to rely primarily on popular benchmarks [7,12]. In such a setting, measurement effects, model architecture, and training strategy become tightly entangled. However, the core idea in P5 is to offer a pipeline that fully exploit the information in existing data, even when its quality is known to be suboptimal or “outdated.” The study is intended to extrapolate expected performance during the gap between a measurement-system upgrade and full model retraining, providing reference points for front-end upgrade decisions and for planning when and how to retrain and tune the models. A limitation is that only two CNN architectures were tested, so the conclusions about upgrade benefits may not generalize to other model families.

6.1.3 Generalization and answers to the RQs

The findings generalize best to settings close to the tested ones: P1/P6 to rooms of similar size and survey granularity (with re-survey for larger changes), P2/P3 to small-aperture indoor arrays with the studied capsules and far-field sources, and P4 to multi-channel auscultation with similar placement discipline. Within those bounds, the results answer the questions as follows:

Topic A (RQ1.1, RQ1.2). Ambient noise can support position discrimination when conditions are stable (P1), and adding an array with known excitation using CNN improves distinctiveness and robustness (P6).

Topic B (RQ2.1, RQ2.2). Microphone directivity and array geometry materially affect DoA accuracy in rooms (P2); a UCA with coherent focusing attains low RMS angular error at reduced computational cost relative to generic focusing (P3).

Topic C (RQ3.1, RQ3.2). A four-channel stethoscope with simple beamforming and matched filtering improves heart-sound quality prior to classification (P4). With the classifier fixed, P5 shows that accuracy changes systematically with different SNRs, so the results can be used to approximate how performance would evolve across future measurement-system upgrades. (P5).

6.1.4 Scope and trade-offs

Topic A — Acoustic Fingerprinting P1 intentionally used a single microphone with magnitude-only features and ambient-noise subtraction to minimize hardware and survey effort for room “place recognition”. This choice favored low instrumentation cost and reproducibility over phase-based disambiguation, which single-channel systems naturally lack [26, 45]. The grid-based framing provided a clean supervised task for stability analysis, even though it does not target continuous 2D regression [23, 105]. It is worth noticing that active fingerprinting can strengthen identifiability in ordinary rooms as in [20]. Furthermore, inaudible-echo room recognition from [106] also shows how short, structured probes unlock robust fingerprinting. Based on these, a controlled excitation approach was motivated and used in P6 .

The follow-on **P6** moved to a four-microphone array with an exponential swept-sine (ESS) per cell to increase distinctiveness and training signal control, so that each label couples to a measured channel response with good SNR and low distortion [92]. This approach strengthens the possibility to distinguish the signals compared to passively sampled ambiance [75]. Additionally, field reports from [60] confirm the practicality of sweep-based impulse-response capture in complex venues, which supports using ESS for reliable training data.

In both works, the fingerprinting perspective complements geometric DoA by embracing room-specific signatures rather than attempting geometry-only inference [25, 44].

Topic B — Array-Based DoA Estimation In P2, three planar geometries (linear, square, quarter-circle) and two microphone types (omni, cardioid) were chosen to

give clear, repeatable contrasts for RQ2.1. Other layouts (e.g., taller or denser arrays, larger apertures, other patterns) were outside of the scope. The TDE based approach follow standard generalized cross-correlation practice for wideband TDOA estimation [37]. Designing under room constraints aligns with aeroacoustic and acoustic-imaging workflows that emphasize interpretable comparisons across arrays [2].

P3 focused on a six-channel uniform circular array at one radius to enable a circular-harmonic focusing matrix. Alternative radius of UCA were extrapolated to be working but were not evaluated experimentally. The pipeline was designed around classical subspace and beamforming methods, so that results can be compared directly with well-established references like MUSIC and MVDR [5,38]. Computationally efficient wideband DOA designs on simple arrays from [107] further support this choice to keep the array compact and computationally light. The result in P3 confirms the use of a UCA enabled lower complexity while preserving estimation quality within the tested bandwidth [67, 108], making it suitable for embedded deployments [27]. Recent steering-vector reconfiguration methods in [109] further illustrate how UCAs can trade flexibility and complexity at low channel counts, while learning in the circular-harmonic domain provides a complementary path that validates the utility of UCA-structured features [42]. However, alternative TOPS refinements show additional robustness routes P3 did not exhaustively benchmark [110].

Topic C — Biomedical Application P4 implemented a four-channel electronic stethoscope with DSB and matched-filter segmentation to enhance S1/S2 heart-sound segments before downstream analysis [54]. The study explicitly linked the sensor and recording design to the classifier, instead of treating signal acquisition and learning as separate stages [111]. The array design and protocol in P4 were inspired by the aspects from the reviews of modern stethoscope systems from [50]. The hardware and protocol emphasized low-noise amplification and clear placement guidance. However, aspects including hygiene, placement consistency, and staff workload were not evaluated [46,112].

P5 varied signal-to-noise ratio and device response while keeping a pretrained CNN fixed, extending earlier analyses of cardiopulmonary sensor arrays in [113]. The study followed the thesis stance that metrology and model choice should be planned together, so that part of the accuracy gain can come from better sensors and acquisition setups instead of only from repeated retraining [49]. This perspective is consistent with observations that many of the clearest gains in acoustic localization and classification arise from improved measurements and well-posed processing chains [2]. In a complementary direction, industrial sensing work recommends targeted SNR improvements as a supplement to data augmentation when labelled data are limited, as discussed in [114]. Taken together, these external results and the findings of P5 reinforce the view that architecture and data quality jointly determine performance, as argued in [57]. However, the study did not cover distortions (e.g., clipping, non-linear sensor behaviour or motion) or a broad range of transducers, so conclusions apply to the tested settings and model. In addition, interactions between training-time augmentation and test-time degradations were not mapped in detail.

6.1.5 Technical Limitations

Residual gain or phase mismatch, clock drift and ADC headroom in low-cost arrays can bias beamforming and subspace estimates and fingerprint stability [3, 26, 75]. In **P1**, a single microphone cannot distinguish mirror-symmetric source configurations unless one imposes additional constraints (e.g., known source side, asymmetric placement, or motion cues) are imposed [67]. System-level studies in [115] further stress that synchronization and platform effects materially influence accuracy. Authors in [27] further highlight that even small timing errors accumulate into direction errors for delay-based methods (**P2**, **P3**).

Reviews of acoustic localization document that multi-talker mixtures, moving sources, and transients remain challenging compared with controlled settings [44]. Surveyed datasets often derive DOA labels from geometry rather than motion capture, limiting absolute-error explainability [116]. Recent overviews call for richer annotation and tracking to strengthen conclusions from room experiments [25]. Thus, controlled office or lab recordings as the measurements taken in **P4** has the restriction including the range of interferes, motion, and reverberation compared with public or industrial spaces [23].

The coherent wideband uniform circular array (UCA) method in **P3** and the fingerprinting in **P6** were validated offline. As TinyML and embedded-audio surveys note, what matters in deployment is worst-case latency, memory, and contention under co-running tasks [117] [40]. However, those platform-specific profiles were not measured as they fall out of the research scope.

6.1.6 Practical Limitations

Acoustic fingerprints **P1**, **P6** are space-specific which will differ with environment condition changes, the practical concern regarding this include measurement portability and sensitivity [67]. Optimal geometry and frequency-dependent aperture choices also shift trade-offs across deployments [41].

DoA performance depends on array configurations [22, 75, 118]. Linear-array analyses in [58] similarly report sensitivities to spacing and sensor mismatch that affect cross-device transfer. Fortunately measurement in both **P2** and **P3** are not having such problem. However, results in **P3** transfer most directly to planar UCAs rather than arbitrary apertures.

Practical use for arrays requires clear aperture, stable mounting, reliable time synchronization, and periodic calibration, understood here as routine checks and small gain/phase/delay corrections to maintain coherence over time as mentioned in the book *Microphone Arrays* from [26]. However, there are additional concerns regarding practical deployment in the application under clinical scope as in **P4**. For example, placement and hygiene keep-up add workflow demand relative to single-sensor stethoscopes [111]. In addition, reports on clinical use from [112] emphasis patient comfort and staff burden as determinants of utility beyond raw signal quality.

Computer-aided auscultation studies from [49] show SNR improvements yet un-

underscore the need for prospective, and clinical-scale validation is still required to connect acoustic improvements to decision-making. Moreover, multi-sensor cardiopulmonary work from [113] illustrates the benefit and computational cost trade-off when moving from one sensor to arrays in bedside contexts.

6.2 Research Outcomes

6.2.1 Use and Misuse of Outcomes

- **Tracking and privacy:** Research results from **P1, P6** for indoor localization and fingerprinting, such techniques that recognize places could, if misused, support position/location tracking without explicit consent.

Use Guidance: Recommend consent-based deployment, clear user notices, provide quit option, on-device processing where possible, and storage of non-content features (e.g., magnitude/phase summaries) rather than raw audio. In addition, any shared datasets should exclude intelligible speech, or apply strict anonymization.

- **Direct use of the results:** Comparison results for microphone directivity from **P2**, the conclusion of reduced computation with UCA from **P3**, and the performance trend extrapolation of the pre-trained models in **P5** might be directly and blindly copied and used without detailed clarification.

Use Guidance: The re-use or re-claim from the results needs to not only follow, but also re-list explicitly the conditions of validity to maintain rigor and report what enables replication.

- **Biomedical interpretation:** Array-enhanced auscultation and classification (**P4, P5**). Analyzation result might be misused by over-interpret as direct diagnosis.

Use Guidance: Needs to state that the methods are not diagnostic, and use only as decision support with clinician oversight, plus require clinical validation and regulatory review before any medical use.

6.2.2 Originality and Novelty

Across **P1–P6**, the contributions combine simple, buildable hardware with analysis choices that expose what the measurements themselves can and cannot support. In the following paragraphs, for each paper the originality will be first described as what has been achieved, and then the novelty will be highlighted.

P1 investigated the feasibility of passive acoustic fingerprints without beacons. The work demonstrated that reliable region recognition is possible using ambient noise with a single microphone, without synchronized arrays, dense hardware, or complex computation. The novelty lies in showing that such minimal passive measurements already support useful region-level localization.

P2 presents a fair microphone type and geometry comparison for indoor DoA estimation. The result achieved a clear, empirical picture of how sensor directivity and layout shift accuracy. The novelty lies in providing a strong evidence that measurement decisions matter as much as the algorithm.

The proposed modified AMI method with UCA in P3 achieved computational cost reduction without precision degrading. The novelty is not a new estimator per se, but a focusing matrix that makes coherent processing more efficient for uniform circular arrays.

The proposed beamforming approach in P4 applied to a designed stethoscope prototype, shows that basic spatial processing can “clean” heart sounds at the point of capture, improving downstream quality without relying on heavy post-processing. The novelty lies mainly in the system design and segmentation and beamforming pipeline rather than in new theory, its role in the thesis is to provide a controlled way of improving acoustic data quality for heart sounds.

P5 then builds directly on this by quantifying how such quality changes propagate to classification accuracy in pretrained CNNs, linking measurement choices to machine-learning performance. By augmenting recordings with controlled noise, trends from lower to higher quality acquisitions were estimated. This offers a practical way to forecast returns from hardware upgrades or data-collection changes on classifier performance without a full set of re-gathering of dataset and retraining a model for every configuration.

P6 continues the silent object acoustic fingerprinting in P1 with an active known excitation signal. Linear room impulse responses from ESS measurements are treated as an “acoustic CSI” and fed to a compact CNN for grid classification. The novelty lies in showing that a structured array with learned RIR-based fingerprints can automatically separate each grid cells with distance errors comparable to P1, but on a denser grid and without manual matching.

6.2.3 Significance and Impact

By treating microphones, geometries, and acquisition protocols as first-class design variables, the work helps narrow the gap between algorithmic demonstrations and deployed systems in indoor and biomedical settings. The contributions provide:

- (i) A path to scalable and robust acoustic fingerprinting
- (ii) Array configuration comparison for DoA evidence under realistic conditions
- (iii) Demonstrations of how basic array processing improves biomedical signals at the source
- (iv) Quantitative links between measurement quality and DL outcomes

These results can inform standards, test plans, and hardware considerations for systems that rely on Sound Source Localization (SSL) and acoustic classification.

6.3 Risks and Ethical Aspects

The section reflects on risks and ethical aspects associated with the studies summarized in this thesis (P1–P6), both the process (methods and data handling), and the research outcome (use or misuse of results), and outlines mitigation adapted.

6.3.1 Uncertainties and measurement bias

- **Room variability and nonstationarity (P1, P2, P3, P6):** Acoustics can shift with people, furniture, HVAC, or humidity, which may bias results if training and testing conditions differ.
Mitigation: conduct measurements on different days, record basic room conditions, and include sensitivity checks over noise levels and array geometry (P2, P3).
- **Device and array dependence (P2, P3, P4, P5, P6):** Performance may depend on microphone directivity, aperture, and electronic front-end.
Mitigation: fair comparisons of two sensor types and geometries in identical room (P2), explicit reporting of hardware configurations (P3, P4, P6), and controlled noise level of dataset to quantify measurement impact (P5).
- **Model generalization (P5, P6):** Learned models may work well with one type of environment or device but drop in accuracy elsewhere.
Mitigation: using simple, transparent model choices, cross-validation, and tests under planned degradations (e.g SNR variations in P5) to show how accuracy changes.
- **Safety (Acoustic exposure) (P1, P2, P3, P6):** Active measurements used short, band-limited signals at comfortable levels.
Mitigation: Sound pressure level (SPL) checks below occupational exposure limits, minimal duty cycles, and clear signage during measurements.
- **Research integrity (through all research studies in this thesis):** Ambiguity in pre-processing or unreported parameters can hinder reproducibility.
Mitigation: Methodology transparency needs to be reported including fixed sampling rates, windowing, passbands, and algorithmic settings. In addition, feasible and processing steps are described sufficiently for reproduction (e.g., focusing matrix structure in P3, filtering/beamforming in P4). Negative or neutral findings (e.g., geometry trade-offs in P2) are reported rather than omitted.

6.3.2 Ethical aspects in research process

- **Human involvement and privacy concern (P1, P4, P5, P6):** Although the method in P1 targets spectral fingerprints of a room with silent objects (human body), ambient recordings could accidentally capture speech.
Mitigation: measurement sessions in controlled settings. There is no speech recognition or content analysis, and the fingerprints derived from smoothed spectra rather

than raw waveforms. In addition, files are kept only as long as needed and the access is restricted.

- **Sensitive data handling (P4, P5):** Both dataset used, and real-world recordings measured in P4 and P5 involve human participants. Thus, such biomedical audios, when they are used as signals in the research, can be sensitive.

Mitigation: The dataset used in P4 and P5 already includes ethics approval and informed consent. No personally identifying data are stored. In addition, analyses are limited to cardiac sound quality and classification, not to broader health inferences. Moreover, the workflow records only what is needed, and has low-frequency band-limiting for just auscultation, removes metadata, and follows standard data-protection regulations.

Chapter 7

Conclusions

This chapter summarizes what was learned from the measurement-driven studies in acoustic localization and classification (P1–P6), reflects on methodological choices, and outlines directions for continued work. The conclusions are organized into (i) work conclusions, structured by the three research topics and their research questions as defined in Chapter 1, and (ii) future work that extends the measurement-focused perspective of the thesis.

7.1 Work Conclusions

7.1.1 Chronological Network of the Study

Figure 7.1 presents the thesis as a time-ordered network of papers, laid out from left to right by publication year and arranged top to bottom by paper ID for clarity. Curved links trace the main threads that run through the work, such as silent object fingerprinting (Topic A), array-based DoA estimation (Topic B), and biomedical applications (Topic C). The color, shape coding, and edge styles are defined in the embedded legend.

7.1.2 Summary by Problem Area

P1, P2, P3, and P6 form a spatial and environmental sensing thread that spans Topics A and B. P1 and P6 explore how different fingerprinting strategies and microphone configurations affect indoor localization performance when the object of interest remains silent. P2 and P3 investigate how compact array layouts, microphone directivity, and coherent wideband processing influence DoA estimation under realistic room conditions. Together, these studies map out how much of the localization problem can be addressed by careful measurement design with relatively standard algorithms, and where theoretical advantages are eroded by practical constraints

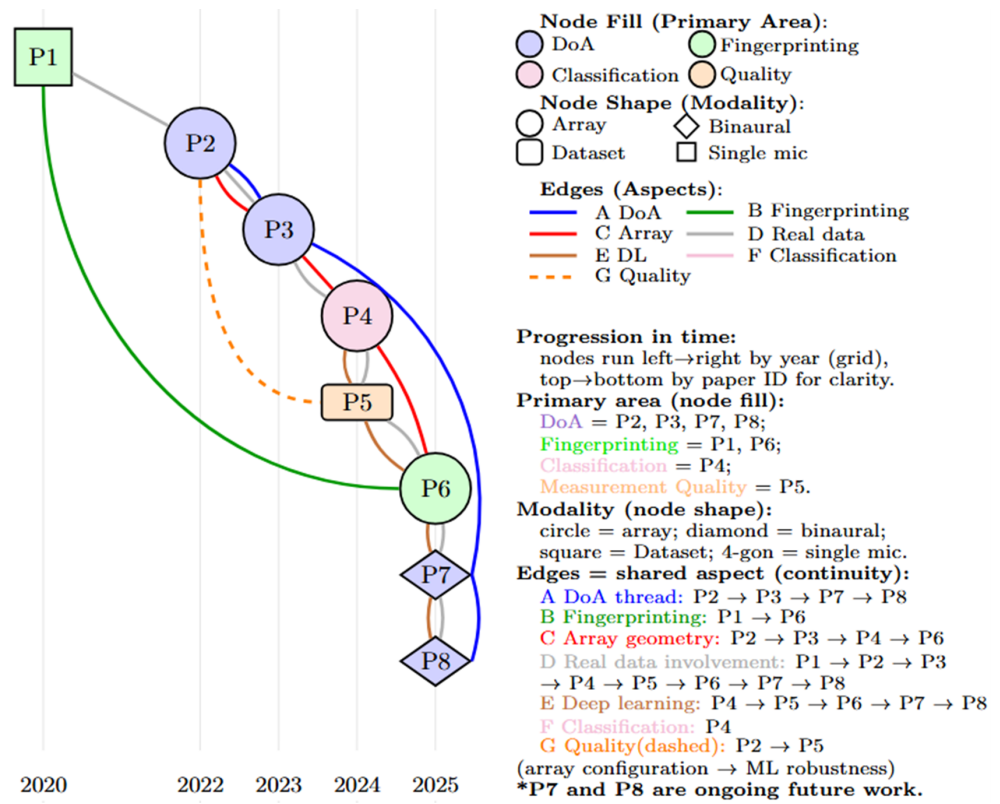


Figure 7.1: Illustration of the progression of the studies and connections between all papers.

such as array aperture, reverberation, and device variability.

P4 and P5 form a biomedical sensing thread in Topic C, centered on heart sound classification. P4 introduces and evaluates a four-channel digital stethoscope with beamforming for matched filtering, showing how a concrete multi-channel measurement setup affects heart-sound segmentation and murmur classification. P5 then steps back from sensor design and focuses on measurement quality for learning-based models by fixing a pretrained CNN and systematically varying SNR and device response. In doing so, P5 provides a useful perspective on a common dilemma in data-driven biomedical applications: when existing recordings were collected under earlier or suboptimal measurement conditions, should one re-collect an entirely new dataset for a proven architecture, or if it is possible to continue work within the current model and measurement constraints? The results suggest that, at least for the tested settings, substantial changes in performance can be achieved through controlled measurement upgrades and noise management, without necessarily requiring a complete cycle of model redesign, retraining, and fine-tuning. This question is particularly pressing in domains where data are privacy-sensitive, rare, or expensive to acquire over long periods.

7.1.3 Key Contributions of the Thesis

The main contributions of the thesis can be summarized as follows:

- It demonstrates that passive, minimal-hardware acoustic fingerprinting using ambient noise is feasible for silent object localization on indoor grids, and that array-aided, deep-learning-based methods can extend this to more robust and fine-grained fingerprints (P1, P6).
- It provides experimental evidence on how microphone directivity and compact array geometry affect indoor DoA accuracy, offering measured comparisons rather than only algorithmic benchmarks (P2).
- It introduces a coherent wideband UCA pipeline that leverages circular-harmonic structure to reduce computational complexity while maintaining high-resolution DoA estimation, making embedded implementations more practical (P3).
- It shows that simple multi-channel array processing with beamforming for matched filtering can improve the input quality of heart sound recordings for downstream classification in cardiac auscultation (P4).
- It quantifies how systematic changes in SNR bound classifier reliability and calibration for pretrained CNNs, and illustrates how an evaluation pipeline can be used to anticipate the effect of future measurement upgrades without retraining from scratch (P5).

7.1.4 Knowledge Gained

Across Topics A–C, several recurring insights emerge:

- Array geometry and microphone directivity matter measurably in small apertures. Gains from cardioid elements are both layout- and method-dependent, and their benefits depend on how processing handles reverberation (P2).
- Wideband coherence on compact UCAs can be achieved efficiently by reusing modal structure, which maintains accuracy while lowering computational load and thus supports low-power or embedded deployments (P3).
- Passive acoustic fingerprints can localize without beacons when ambient conditions are stable, whereas array diversity combined with learned features improves robustness and supports finer spatial resolution for similar survey effort in more demanding scenarios (P1, P6).
- Input SNR and device consistency set practical ceilings for classifier performance in biomedical audio. Front-end improvement through arrays and beamforming increases the usefulness of each measurement, and controlled SNR studies reveal how far existing models can be pushed by measurement upgrades alone (P4, P5).

These findings translate into reference points for choosing microphones, array layouts, excitation schemes, preprocessing, and data collection protocols. The overarching lesson is that many of the most interpretable gains arise when measurement design and algorithm choice are considered together rather than in isolation.

7.1.5 Overall Conclusion

The overall aim of this dissertation was to investigate acoustic sensing under real-world noise, with a focus on how measurement quality in microphone-based systems shapes what can be reliably inferred from sound in indoor localization and heart sound classification. Taken together, P1–P6 indicate that this aim has been addressed in a concrete, measurement-driven way. Across all three research topics, performance differences that might otherwise be attributed to “better algorithms” can often be traced back to how microphones, arrays, excitations, and recording protocols are chosen and implemented.

Overall, the thesis concludes that measurement design is not just an implementation detail but a central part of the scientific problem in acoustic sensing. By treating microphones, arrays, and data quality as first-class design variables, the work provides a structured view of when existing methods are sufficient under real-world noise, where they break down, and how targeted changes in measurement quality can extend their useful range in both indoor localization and heart sound analysis.

7.2 Future Work

The work in this thesis opens several directions for continued research. Each direction extends the measurement-focused perspective developed here and aims to

make acoustic sensing more robust under real-world noise.

7.2.1 Ongoing and Related Projects

Beyond P1–P6, ongoing projects explore measurement-aware binaural localization and wearable sensing. Later work (P7, P8 in the timeline from Figure 7.1) investigates high-precision localization with head-related responses and larger, more realistic datasets of binaural recordings. These directions extend Topic B into human-like sensing and create new benchmarks where microphones are mounted on listeners rather than on fixed arrays. They continue the idea that understanding the measurement configuration is a prerequisite for interpreting model performance.

7.2.2 Possible Directions

Beyond these ongoing activities, two possible directions appear from the results in P1–P6: For acoustic fingerprinting, future work could address larger and more dynamic environments. P1 and P6 studied relatively small grids under controlled movement of a single silent object. In realistic settings, furniture, occupancy, and ventilation can change over time. This motivates:

- Extending acoustic CSI-style fingerprints to multi-room or corridor layouts, where propagation paths are more complex and measurement cost per location must remain low.

This direction is relevant to the thesis because it tests whether the measurement-driven design used in P1 and P6 can scale when background conditions vary more strongly and when the cost of re-measurement becomes a major constraint.

In biomedical sound sensing, P4 and P5 highlight the benefits that better acquisition and SNR strongly influence classifier behavior in heart sound analysis. Future work can build on this to extend P5 into a more generalized pipeline:

- Integrating measurement-quality estimates (for example, online SNR or quality scores) into model selection and decision thresholds, and use controlled degradation studies to extrapolate how future measurement setup or model upgrades are likely to perform.

The aim is to fully exploit existing, possibly suboptimal datasets rather than discarding them whenever hardware or protocols change. This direction is particularly relevant for data that are sensitive, heterogeneous, and expensive to recollect, where reusing of older data is essential.

7.3 Closing Remark

Across the studies in this thesis, prioritizing measurement design consistently provided reliable gains and clearer insight than focusing on algorithms alone. From indoor fingerprinting to compact-array DoA and heart sound classification, the results show that the quality and structure of acoustic measurements largely determine what can be inferred under real-world noise. Continued progress is likely to come from designs that treat microphones, geometry, excitation, and recording protocols as central scientific variables, and that integrate these choices tightly with the selected inference methods.

Bibliography

- [1] C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.
- [2] P. Chiariotti, M. Martarelli, and P. Castellini, "Acoustic beamforming for noise source localization—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 120, pp. 422–448, Apr. 2019, num Pages: 27 Place: London Publisher: Academic Press Ltd- Elsevier Science Ltd Web of Science ID: WOS:000456641400027. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/ffaf4b5a-3759-4275-863b-bc6901443a95-d551a98e/times-cited-descending/1>
- [3] J. Benesty, J. Chen, and I. Cohen, *Design of Circular Differential Microphone Arrays*. Springer Cham, 2015, vol. 12. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-319-14842-7>
- [4] L. Marchegiani and P. Newman, "Listening for Sirens: Locating and Classifying Acoustic Alarms in City Scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17 087–17 096, Oct. 2022, conference Name: IEEE Transactions on Intelligent Transportation Systems. [Online]. Available: <https://ieeexplore.ieee.org/document/9737390>
- [5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986. [Online]. Available: <https://ieeexplore.ieee.org/document/1143830>
- [6] M. U. Liaquat, H. S. Munawar, A. Rahman, Z. Qadir, A. Z. Kouzani, and M. A. P. Mahmud, "Localization of Sound Sources: A Systematic Review," *Energies*, vol. 14, no. 13, p. 3910, Jan. 2021, number: 13 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/1996-1073/14/13/3910>
- [7] A. Politis, A. Mesaros, S. Adavanne, T. Heittola, and T. Virtanen, "Overview and Evaluation of Sound Event Localization and Detection in DCASE 2019," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 684–698, 2021, conference Name: IEEE/ACM

- Transactions on Audio, Speech, and Language Processing. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9306885>
- [8] T. Kim Geok, K. Zar Aung, M. Sandar Aung, M. Thu Soe, A. Abdaziz, C. Pao Liew, F. Hossain, C. P. Tso, and W. H. Yong, "Review of Indoor Positioning: Radio Wave Technology," *Applied Sciences-Basel*, vol. 11, no. 1, p. 279, Jan. 2021, num Pages: 44 Place: Basel Publisher: MDPI Web of Science ID: WOS:000605913700001. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/842efe28-15d1-4d69-a0e9-752ef481fb36-d5497a06/relevance/1>
- [9] H. W. Löllmann, C. Evers, A. Schmidt, H. Mellmann, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA Challenge Data Corpus for Acoustic Source Localization and Tracking," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2018, pp. 410–414, iSSN: 2151-870X. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8448644>
- [10] C. Evers, H. W. Loellmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA Challenge: Acoustic Source Localization and Tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1620–1643, 2020, num Pages: 24 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:000542977800002. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/ffaf4b5a-3759-4275-863b-bc6901443a95-d551a98e/times-cited-descending/1>
- [11] N. Aloui, K. Raoof, A. Bouallegue, S. Letourneur, and S. Zaibi, "Performance evaluation of an acoustic indoor localization system based on a fingerprinting technique," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 13, Jan. 2014. [Online]. Available: <https://doi.org/10.1186/1687-6180-2014-13>
- [12] A. Mesaros, T. Heittola, E. Benetos, P. Foster, M. Lagrange, T. Virtanen, and M. D. Plumbley, "Detection and Classification of Acoustic Scenes and Events: Outcome of the DCASE 2016 Challenge," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 379–393, Feb. 2018, conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8123864>
- [13] ELSVIER, "Credit autor statement," accessed: 2025-11-25. [Online]. Available: <https://www.elsevier.com/researcher/author/policies-and-guidelines/credit-author-statement>
- [14] J. Aparicio, F. J. Álvarez, Hernández, and S. Holm, "A Survey on Acoustic Positioning Systems for Location-Based Services," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–36, 2022, conference Name: IEEE Transactions on Instrumentation and Measurement. [Online]. Available: <https://ieeexplore.ieee.org/document/9906111/references#references>

- [15] C. Evers and P. A. Naylor, "Acoustic SLAM," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 9, pp. 1484–1498, Sep. 2018, num Pages: 15 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:000433371500002. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/ffaf4b5a-3759-4275-863b-bc6901443a95-d551a98e/times-cited-descending/1>
- [16] N. Chu, Y. Ning, L. Yu, Q. Liu, Q. Huang, D. Wu, and P. Hou, "Acoustic source localization in a reverberant environment based on sound field morphological component analysis and alternating direction method of multipliers," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [17] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.
- [18] 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, Jul. 2012, conference Name: IEEE Transactions on Audio, Speech, and Language Processing. [Online]. Available: <https://ieeexplore.ieee.org/document/6112205>
- [19] S. Dupont, M. Sanalattii, M. Melon, O. Robin, A. Berry, and J.-C. L. Roux, "Characterization of acoustic materials at arbitrary incidence angle using sound field synthesis," *Acta Acustica*, vol. 6, p. 61, 2022, publisher: EDP Sciences. [Online]. Available: <https://acta-acustica.edpsciences.org/articles/aacus/abs/2022/01/aacus220067/aacus220067.html>
- [20] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster, "RoomSense: An indoor positioning system for smartphones using active sound probing," in *Adjunct Proc. of the 2013 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp Adjunct)*. Stuttgart, Germany: Association for Computing Machinery, New York, NY, United States, 2013, pp. 89–95, smartphone-based active acoustic room fingerprinting.
- [21] T.-K. Hon, L. Wang, J. D. Reiss, and A. Cavallaro, "Audio Fingerprinting for Multi-Device Self-Localization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1623–1636, Oct. 2015, conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing. [Online]. Available: <https://ieeexplore.ieee.org/document/7118681>
- [22] S. He and S.-H. G. Chan, "Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.
- [23] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

- [24] X. Tong, Y. Wan, Q. Li, X. Tian, and X. Wang, "CSI Fingerprinting Localization With Low Human Efforts," *IEEE/ACM Transactions on Networking*, vol. 29, no. 1, pp. 372–385, Feb. 2021, conference Name: IEEE/ACM Transactions on Networking. [Online]. Available: <https://ieeexplore.ieee.org/document/9259006>
- [25] K. A. Kordi, M. Roslee, M. Y. Alias, A. Alhammadi, A. Waseem, and A. F. Osman, "Survey of indoor localization based on deep learning," *Computers, Materials and Continua*, vol. 79, no. 2, pp. 3261–3298, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1546221824002753>
- [26] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer Science & Business Media, 2001.
- [27] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, *Robust Localization in Reverberant Rooms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 157–180. [Online]. Available: https://doi.org/10.1007/978-3-662-04619-7_8
- [28] D. Desai and N. Mehendale, "A Review on Sound Source Localization Systems," *Archives of Computational Methods in Engineering*, vol. 29, pp. 4631–4642, 2021. [Online]. Available: <https://consensus.app/papers/review-sound-source-localization-systems-desai/75c4098bd72a55da99e5e4e9dac4d7d5/>
- [29] F. Grondin and F. Michaud, "Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations," *Robotic and Autonomous Systems*, vol. 113, pp. 63–80, Mar. 2019, num Pages: 18 Place: Amsterdam Publisher: Elsevier Web of Science ID: WOS:000459358000006. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/ffaf4b5a-3759-4275-863b-bc6901443a95-d551a98e/times-cited-descending/1>
- [30] D. Davis, S. Gaye, L. Mandjoupa, J. An, W. Mahmoud, L. Wang, and M. Denis, "Locating impulsive sound sources in microscale urban spaces," *The Journal of the Acoustical Society of America*, 2022. [Online]. Available: <https://consensus.app/papers/locating-sound-sources-spaces-davis/415b98c071675a998bd71eaf985aa2d8/>
- [31] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985.
- [32] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553–1560, 1988.
- [33] M. Swartling, N. Grbic, and I. Claesson, "Direction of arrival estimation for multiple speakers using time-frequency orthogonal signal separation," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 4. IEEE, 2006, p. IV–IV.

- [34] W. Zhang and B. D. Rao, "Two microphone based direction of arrival estimation for multiple speech sources using spectral properties of speech," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, p. 2193–2196.
- [35] J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2510–2526, 2007.
- [36] E. Grinstein, E. Tengan, B. Çakmak, T. Dietzen, L. Nunes, T. van Waterschoot, M. Brookes, and P. A. Naylor, "Steered response power for sound source localization: A tutorial review," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 59, 2024. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11557718/>
- [37] L. Chen, Y. Liu, F. Kong, and N. He, "Acoustic source localization based on generalized cross-correlation time-delay estimation," *Procedia Engineering*, vol. 15, pp. 4912–4919, 2011, cEIS 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877705811024167>
- [38] B. G. Ferguson, "Minimum variance distortionless response beamforming of acoustic array data," *The Journal of the Acoustical Society of America*, vol. 104, no. 2, pp. 947–954, Aug. 1998. [Online]. Available: <https://doi.org/10.1121/1.423311>
- [39] A. Dehghan Firoozabadi, P. Irarrazaval, P. Adasme, D. Zabala-Blanco, P. P. Játiva, and C. Azurdia-Meza, "3d multiple sound source localization by proposed t-shaped circular distributed microphone arrays in combination with gev and adaptive gcc-phat/ml algorithms," *Sensors*, vol. 22, no. 3, January 2022.
- [40] B. Liao, Z. Zhang, and S. C. Chan, "Chapter 5 - subspace tracking for time-varying direction-of-arrival estimation with sensor arrays," in *IoT and Spacecraft Informatics*, ser. Aerospace Engineering, K. Yung, A. W. Ip, F. Xhafa, and K. Tseng, Eds. Elsevier, 2022, pp. 129–155. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128210512000118>
- [41] S. Yan, "Optimal design of modal beamformers for circular arrays," *Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. 2140–2151, Oct. 2015, num Pages: 12 Place: Melville Publisher: Acoustical Soc Amer Amer Inst Physics Web of Science ID: WOS:000368186600030. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/ffaf4b5a-3759-4275-863b-bc6901443a95-d551a98e/times-cited-descending/1>
- [42] K. SongGong, W. Wang, and H. Chen, "Acoustic source localization in the circular harmonic domain using deep learning architecture," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2475–2491, 2022.

- [43] M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, A. Ali, M. Roch, S. Gannot, and C.-A. Deledalle, "Machine learning in acoustics: Theory and applications," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3590–3628, 2019. [Online]. Available: <https://pubs.aip.org/asa/jasa/article/146/5/3590/994832/Machine-learning-in-acoustics-Theory-and>
- [44] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, "A survey of sound source localization with deep learning methods," *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 107–151, 2022. [Online]. Available: <https://pubs.aip.org/asa/jasa/article/152/1/107/2838290/A-survey-of-sound-source-localization-with-deep>
- [45] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.
- [46] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark, "Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016," in *2016 Computing in cardiology conference (CinC)*, 2016, pp. 609–612.
- [47] T.-H. Tan, Y.-T. Lin, Y.-L. Chang, and M. Alkhaleefah, "Sound source localization using a convolutional neural network and regression model," *Sensors*, vol. 21, no. 23, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/23/8031>
- [48] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 34–48, 2019.
- [49] M. A. Reyna, Y. Kiarashi, A. Elola, J. Oliveira, F. Renna, A. Gu, E. A. P. Alday, N. Sadr, A. Sharma, S. Mattos *et al.*, "Heart murmur detection from phonocardiogram recordings: The george b. moody physionet challenge 2022," in *2022 Computing in Cardiology (CinC)*, vol. 498, 2022, pp. 1–4.
- [50] J. J. Seah, J. Zhao, D. Y. Wang, and H. P. Lee, "Review on the advancements of stethoscope types in chest auscultation," *Diagnostics*, vol. 13, no. 9, 2023.
- [51] H. Gray, *Anatomy of the human body*. Lea & Febiger, 1878, vol. 8.
- [52] Vinne2, "Heart sounds auscultation areas," Wikimedia Commons, 2011, derivative work based on Ickle's image and Gray's Anatomy plate; released by the copyright holder into the public domain (PD-self). [Online]. Available: <https://commons.wikimedia.org/wiki/File:Heart.sounds.auscultation.are.as.svg>
- [53] L. Servier, "Cardiovascular system — heart 3," Wikimedia Commons, 2016, from SMART-Servier Medical Art. Licensed under CC BY-SA

- 3.0. [Online]. Available: https://commons.wikimedia.org/wiki/File:Cardiovascular_system_-_Heart_3_-_Smart-Servier.png
- [54] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation," *IEEE transactions on bio-medical engineering*, vol. 63, no. 4, pp. 822–832, 2016.
- [55] A. M. McKee and R. A. Goubran, "Chest sound pick-up using a multisensor array," in *Sensors, 2005 IEEE*, 2005, pp. 4 pp.–.
- [56] I. McLane, D. Emmanouilidou, J. E. West, and M. Elhilali, "Design and comparative performance of a robust lung auscultation system for noisy clinical settings," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2583–2594, 2021.
- [57] D. Salvati, C. Drioli, and G. L. Foresti, "Exploiting CNNs for Improving Acoustic Source Localization in Noisy and Reverberant Conditions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 103–116, Apr. 2018, num Pages: 14 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:000679685300003. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/ffaf4b5a-3759-4275-863b-bc6901443a95-d551a98e/times-cited-descending/1>
- [58] A. AlShehhi, M. L. Hammadih, M. S. Zitouni, S. AlKindi, N. Ali, and L. Weruaga, "Linear and Circular Microphone Array for Remote Surveillance: Simulated Performance Analysis," Mar. 2017, arXiv:1703.02318 [cs]. [Online]. Available: <http://arxiv.org/abs/1703.02318>
- [59] O. Robin, A. Berry, C. Kafui Amédin, N. Atalla, O. Doutres, and F. Sgard, "Laboratory and in situ sound absorption measurement under a synthesized diffuse acoustic field," *Building Acoustics*, vol. 26, no. 4, pp. 223–242, Dec. 2019, publisher: SAGE Publications Ltd STM. [Online]. Available: <https://doi.org/10.1177/1351010X19870307>
- [60] P. Guidorzi, L. Barbaresi, D. D’Orazio, and M. Garai, "Impulse responses measured with mls or swept-sine signals applied to architectural acoustics: An in-depth analysis of the two methods and some case studies of measurements inside theaters," *Energy Procedia*, vol. 78, pp. 1611–1616, 2015, 6th International Building Physics Conference, IBPC 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1876610215019682>
- [61] M. Slaney, R. F. Lyon, R. Garcia, B. Kemler, C. Gnegy, K. Wilson, D. Kanevsky, S. Savla, and V. G. Cerf, "Auditory Measures for the Next Billion Users," *Ear and Hearing*, vol. 41, p. 131S, Dec. 2020. [Online]. Available: https://journals.lww.com/ear-hearing/fulltext/2020/11001/auditory_measures_for_the_next_billion_users.14.aspx
- [62] D. Sanchez-Rodriguez, P. Hernandez-Morera, J. M. Quinteiro, and I. Alonso-Gonzalez, "A Low Complexity System Based on Multiple Weighted Decision Trees for Indoor Localization," *Sensors*, vol. 15, no. 6, pp.

- 14809–14829, Jun. 2015, num Pages: 21 Place: Basel Publisher: MDPI Web of Science ID: WOS:000357869200135. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/842efe28-15d1-4d69-a0e9-752ef481fb36-d5497a06/relevance/1>
- [63] P. Botsinis, D. Alanis, S. Feng, Z. Babar, V. N. Hung, D. Chandra, S. X. Ng, R. Zhang, and L. Hanzo, “Quantum-Assisted Indoor Localization for Uplink mm-Wave and Downlink Visible Light Communication Systems,” *IEEE ACCESS*, vol. 5, pp. 23327–23351, 2017, num Pages: 25 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:000415170700031. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/842efe28-15d1-4d69-a0e9-752ef481fb36-d5497a06/relevance/1>
- [64] Z. Cheng, D. Zhao, W. Guo, and L. Li, “A channel state information and geomagnetic fused fingerprint localisation algorithm based on multi-input convolutional neural network,” *IET Wireless Sensor Systems*, vol. 14, no. 1-2, pp. 33–46, 2024, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1049/wss2.12075>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1049/wss2.12075>
- [65] X. Song, M. Wang, H. Qiu, K. Li, and C. Ang, “Auditory Scene Analysis-Based Feature Extraction for Indoor Subarea Localization Using Smartphones,” *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6309–6316, Aug. 2019, conference Name: IEEE Sensors Journal. [Online]. Available: <https://ieeexplore.ieee.org/document/8610149>
- [66] E. Latif and R. Parasuraman, “Instantaneous Wireless Robotic Node Localization Using Collaborative Direction of Arrival,” *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 2783–2795, Jan. 2024, num Pages: 13 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:001153911600062. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/842efe28-15d1-4d69-a0e9-752ef481fb36-d5497a06/relevance/1>
- [67] S. Parmar, X. Wang, C. Yang, and S. Mao, “Voice fingerprinting for indoor localization with a single microphone array and deep learning,” in *Proceedings of the 2022 ACM Workshop on Wireless Security and Machine Learning*, ser. WiseML ’22. New York, NY, USA: Association for Computing Machinery, 2022, p. 21–26. [Online]. Available: <https://doi.org/10.1145/3522783.3529528>
- [68] S. Chen, R. Tan, Z. Wang, X. Tong, and K. Li, “Voicemap: Autonomous mapping of microphone array for voice localization,” *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 2909–2923, 2024.
- [69] S. Wang, P. Yang, and H. Sun, “Fingerprinting Acoustic Localization Indoor Based on Cluster Analysis and Iterative Interpolation,” *Applied Sciences-Basel*, vol. 8, no. 10, p. 1862, Oct. 2018, num Pages: 13 Place:

- Basel Publisher: MDPI Web of Science ID: WOS:000448653700157. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/727ec073-8703-4101-9837-835e42984670-d5d7839b/times-cited-descending/1>
- [70] S. Wielandt and L. D. Strycker, "Indoor multipath assisted angle of arrival localization," *Sensors*, vol. 17, no. 11, p. 2522, Nov. 2017, number: 11 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/1424-8220/17/11/2522>
- [71] M. Jiang, J. Lundgren, S. Pasha, M. Carratù, C. Liguori, and G. Thungström, "Indoor Silent Object Localization using Ambient Acoustic Noise Fingerprinting," in *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2020, pp. 1–6, iSSN: 2642-2077. [Online]. Available: <https://ieeexplore.ieee.org/document/9129086>
- [72] T. Pham and B. M. Sadler, "Adaptive wideband aeroacoustic array processing," in *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, Jun. 1996, pp. 295–298. [Online]. Available: <https://ieeexplore.ieee.org/document/534875>
- [73] M. Doran, E. Doron, and A. Weiss, "Coherent wide-band processing for arbitrary array geometry," *IEEE Transactions on Signal Processing*, vol. 41, no. 1, pp. 414–, Jan. 1993, conference Name: IEEE Transactions on Signal Processing. [Online]. Available: <https://ieeexplore.ieee.org/document/193167>
- [74] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.
- [75] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.
- [76] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2019, vol. 16.
- [77] J. Tiete, F. Domínguez, B. D. Silva, L. Segers, K. Steenhaut, and A. Touhafi, "SoundCompass: A Distributed MEMS Microphone Array-Based Sensor for Sound Source Localization," *Sensors*, vol. 14, no. 2, pp. 1918–1949, Feb. 2014, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/1424-8220/14/2/1918>
- [78] H. Wang, X. Wang, X. Lan, T. Su, and L. Wan, "BSBL-Based Auxiliary Vehicle Position Analysis in Smart City Using Distributed MEC and UAV-Deployed IoT," *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 975–986, Jan. 2023, num Pages: 12 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:001011036600002. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/842efe28-15d1-4d69-a0e9-752ef481fb36-d5497a06/relevance/1>
- [79] M. Jiang, N. Chibuzo, J. Lundgren, M. Sjöström, G. Thungström, and S. Gao, "Performance Comparison of Omni and Cardioid Directional

- Microphones for Indoor Angle of Arrival Sound Source Localization | IEEE Conference Publication | IEEE Xplore," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. Ottawa, ON, Canada: IEEE, Jun. 2022, pp. 1–6, iSSN: 2642-2077. [Online]. Available: <https://ieeexplore.ieee.org/document/9806559>
- [80] M. Jiang, C. Nnonyelu, J. Lundgren, G. Thungström, and M. Sjöström, "A Coherent Wideband Acoustic Source Localization Using a Uniform Circular Array," *Sensors*, vol. 23, no. 11, p. 5061, May 2023, this article belongs to the Topic Advances in Array Signal Processing with Errors: Models, Algorithms and Applications). [Online]. Available: <https://www.mdpi.com/1424-8220/23/11/5061>
- [81] M. Adamopoulou, M. Jiang, C. J. Nnonyelu, M. Carratù, C. Liguori, and J. Lundgren, "Improving cardiac auscultation signal quality by using 4-channel stethoscope array," in *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2024, pp. 1–6.
- [82] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 763–776, 2016. [Online]. Available: <https://consensus.app/papers/csibased-fingerprinting-indoor-localization-deep-wang/f80d70033ded5df69784ac7d2fdcc066/>
- [83] X. Wang, L. Gao, and S. Mao, "CSI Phase Fingerprinting for Indoor Localization With a Deep Learning Approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016, num Pages: 11 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:000393048500022. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/0a43e046-5745-4ad0-9538-39ec90745db0-d5d82069/times-cited-descending/1>
- [84] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *MobiSys '11: Proceedings of the 9th International Conference on Mobile Systems, Applications and Services (MobiSys)*, August 2011, p. 155–168.
- [85] H. L. Van Trees, *Optimum Array Processing*, ser. Detection, Estimation, and Modulation Theory. New York: Wiley, 2002.
- [86] A. Foliadis, M. H. C. García, R. Stirling-Gallacher, and R. Thomä, "CSI-Based Localization with CNNs Exploiting Phase Information," *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2021. [Online]. Available: <https://consensus.app/papers/csibased-localization-cnns-exploiting-phase-information-foliadis/bde4b8c9d1af572abb8999ffcf0d17f6/>
- [87] S. Xia, J. Nie, and X. Jiang, "CSafe: An Intelligent Audio Wearable Platform for Improving Construction Worker Safety in Urban Environments," in *Proceedings of the 20th International Conference on Information Processing in Sensor*

- Networks (co-located with CPS-IoT Week 2021)*, ser. IPSN '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 207–221. [Online]. Available: <https://dl.acm.org/doi/10.1145/3412382.3458267>
- [88] J. Lundgren, M. Jiang, V. Laino, V. Gallo, M. Carratù, and C. J. Nnonyelu, “Accuracy impact of increased measurement quality when using pretrained networks for classification,” in *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2024, pp. 1–5.
- [89] K. Liu, X. Liu, and X. Li, “Guoguo: Enabling Fine-Grained Smartphone Localization via Acoustic Anchors,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 5, pp. 1144–1156, May 2016, num Pages: 13 Place: Los Alamitos Publisher: IEEE Computer Soc Web of Science ID: WOS:000374175800008. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/e54eba46-54b2-4606-8e84-e353d322f817-d5d7b80c/times-cited-descending/1>
- [90] I. Constandache, S. Agarwal, I. Tashev, and R. R. Choudhury, “Daredevil: indoor location using sound,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 2, p. 9–19, June 2014.
- [91] B. J. J. DeCourcy and Y.-T. Lin, “Spatial and temporal variation of three-dimensional ship noise coherence in a submarine canyon,” *Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 1042–1051, Feb. 2023, num Pages: 10 Place: Melville Publisher: Acoustical Soc Amer Amer Inst Physics Web of Science ID: WOS:000935611800002. [Online]. Available: <https://www.webofscience.com/wos/woscc/summary/e54eba46-54b2-4606-8e84-e353d322f817-d5d7b80c/times-cited-descending/1>
- [92] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” *Journal of The Audio Engineering Society*, 2000. [Online]. Available: <https://api.semanticscholar.org/CorpusID:9614437>
- [93] A. Farina *et al.*, “Advancements in impulse response measurements by sine sweeps,” in *Audio engineering society convention*, vol. 122, no. 5, 2007.
- [94] Galak76, “Polar pattern omnidirectional,” Wikimedia Commons, 2007, licensed under CC BY-SA (multi-license: 3.0/2.5/2.0/1.0). Accessed: 2025-11-19. [Online]. Available: <https://commons.wikimedia.org/wiki/File:Polar.pattern.omnidirectional.svg>
- [95] Nicoguardo, “Polar pattern cardioid,” Wikimedia Commons, 2016, licensed under CC BY 4.0. Accessed: 2025-11-19. [Online]. Available: <https://commons.wikimedia.org/wiki/File:Polar.pattern.cardioid.svg>
- [96] W. Dong-xia, Q. Chang, Z. Cheng-xu, and N. Fang-lin, “Microphone array optimization design for two-dimensional doa estimation,” *Journal of Dalian University of Technology*, vol. 55, no. 1, p. 103–109, 2015.

- [97] R. Mudhar, "Piezo contact microphone hi-z amplifier - low noise version," <https://www.richardmudhar.com/blog/piezo-contact-microphone-hi-z-amplifier-low-noise-version/>, accessed: 01.09.2022.
- [98] W. Boyes, *Instrumentation Reference Book*. Butterworth-Heinemann, Nov. 2009, google-Books-ID: ZvscLzOlkNgC.
- [99] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor, "The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results," <http://www.peterjbentley.com/heartchallenge/index.html>, accessed: 2025-10-14.
- [100] Q. Lin, Z. An, and L. Yang, "Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices," in *The 25th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: <https://doi.org/10.1145/3300061.3300139>
- [101] N. H. Fletcher, "Shock waves and the sound of a hand-clap—a simple model," *Acoustics Australia*, vol. 41, no. 2, p. 165–168, 2013.
- [102] "The physionet/computing in cardiology challenge 2016," <https://physionet.org/challenge/2016/>, 2016, accessed: 2024.
- [103] Y.-C. Tung and K. G. Shin, "Echotag: Accurate infrastructure-free indoor location tagging with smartphones," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 525–536. [Online]. Available: <https://doi.org/10.1145/2789168.2790102>
- [104] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. USA: Simon & Schuster, Inc., 1992.
- [105] R. Jia, M. Jin, Z. Chen, and C. J. Spanos, "Soundloc: Accurate room-level indoor localization using acoustic signatures," in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, 2015, pp. 186–193.
- [106] Q. Song, C. Gu, and R. Tan, "Deep room recognition using inaudible echos," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 135:1–135:28, 2018.
- [107] B. Suksiri and M. Fukumoto, "A computationally efficient wideband direction-of-arrival estimation method for l-shaped microphone arrays," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5.
- [108] Y.-S. Yoon, L. Kaplan, and J. McClellan, "Tops: new doa estimator for wide-band signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1977–1989, 2006.
- [109] D. Zhao, Z. Deng, and W. Tan, "Low-complexity doa estimation for uniform circular arrays with directional sensors using reconfigurable steering vectors," *Circuits, Systems, and Signal Processing*, vol. 42, no. 3, pp. 1685–1706, 2023.

- [110] H. Hayashi and T. Ohtsuki, "Doa estimation for wideband signals based on weighted squared tops," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–12, 2016.
- [111] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *Biomedical engineering online*, vol. 14, no. 66, pp. 1–37, 2015.
- [112] L. J. Nowak and K. M. Nowak, "Sound differences between electronic and acoustic stethoscopes," *Biomedical engineering online*, vol. 17, no. 104, pp. 1–11, 2018.
- [113] S. Pasha, J. Lundgren, and C. Ritz, "Multi-channel electronic stethoscope for enhanced cardiac auscultation using beamforming and equalisation techniques," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1289–1293.
- [114] M. Hamadache and D. Lee, "Improving signal-to-noise ratio (snr) for inchoate fault detection based on principal component analysis (pca)," in *2014 14th International Conference on Control, Automation and Systems (ICCAS 2014)*, Gyeonggi-do, Korea (South), 2014, p. 561–566.
- [115] W. Wang, J. Li, Y. He, and Y. Liu, "Localizing multiple acoustic sources with a single microphone array," *IEEE Transactions on Mobile Computing*, vol. 22, no. 10, pp. 5963–5977, 2023.
- [116] M. Liu, L. Cheng, K. Qian, J. Wang, J. Wang, and Y. Liu, "Indoor acoustic localization: A survey," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, p. 2, 2020.
- [117] T. Flores, M. Silva, M. Azevedo, T. Medeiros, M. Medeiros, I. Silva, M. M. Dias Santos, and D. G. Costa, "Tinyml for safe driving: The use of embedded machine learning for detecting driver distraction," in *2023 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, Modena, Italy, 2023, p. 62–66.
- [118] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, p. 13–32, March 2009.

Biography

Meng Jiang was born on the 6th of December 1990 in Shanghai, China. She received a Bachelor's degree in Electronics from Mid Sweden University in 2016, and a degree of Master of Science with a major in Electronics from Master by Research Program from Mid Sweden University in 2019. In February 2020, she was admitted to the Licentiate program for sound source localization as a PhD student and later in 2023 admitted to full PhD degree. Meng Jiang in her five-year PhD journey works on sound source localization and classification, focusing on real-world experimentation measurement and application. Her field of interest is designing tailored mixed measurement-first methodologies to study "what is sounding where and when" in real-world environment and improving sound source localization and classification accuracy in critical tasks using array signal processing and deep learning.