Acoustic source localization in 3D complex urban environments

Bumsuk Choi

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute
and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Ricardo A. Burdisso (Chair)
Marin E. Johnson
Michael J. Roan
Jeffery H. Reed
Richard M. Buehrer

April 30, 2012

Blacksburg, Virginia

Keywords: Acoustic Localization, Matched-Field Processing, Adaptive Beamforming, 3D, Uncertainty, Multi-array processing, Fingerprinting, Outdoor Propagation Modeling

Acoustic source localization in 3D complex urban environments

Bumsuk Choi

(ABSTRACT)

The detection and localization of important acoustic events in a complex urban environment, such as gunfire and explosions, is critical to providing effective surveillance of military and civilian areas and installations. In a complex environment, obstacles such as terrain or buildings introduce multipath propagations, reflections, and diffractions which make source localization challenging. This dissertation focuses on the problem of source localization in three-dimensional (3D) realistic urban environments. Two different localization techniques are developed to solve this problem: a) Beamforming using a few microphone phased arrays in conjunction with a high fidelity model and b) Fingerprinting using many dispersed microphones in conjunction with a low fidelity model of the environment.

For an effective source localization technique using microphone phased arrays, several candidate beamformers are investigated using 2D and corresponding 3D numerical models. Among them, the most promising beamformers are chosen for further investigation using 3D large models. For realistic validation, localization error of the beamformers is analyzed for different levels of uncorrelated noise in the environment. Multiple-array processing is also considered to improve the overall localization performance. The sensitivity of the beamformers to uncertainties that cannot be easily accounted for (e.g. temperature gradient and unmodeled object) is then investigated. It is observed that evaluation in 3D models is critical to assess correctly the potential of the localization technique. The enhanced minimum variance distortionless response (EMVDR) is identified to be the only beamformer that has super-directivity property (i.e. accurate localization capability) and still robust to uncorrelated noise in the environment.

It is also demonstrated that the detrimental effect of uncertainties in the modeling of the environment can be alleviated by incoherent multiple arrays.

For efficient source localization technique using dispersed microphones in the environment, acoustic fingerprinting in conjunction with a diffused-based energy model is developed as an alternative to the beamforming technique. This approach is much simpler requiring only microphones rather than arrays. Moreover, it does not require an accurate modeling of the acoustic environment. The approach is validated using the 3D large models. The relationship between the localization accuracy and the number of dispersed microphones is investigated. The effect of the accuracy of the model is also addressed. The results show a progressive improvement in the source localization capabilities as the number of microphones increases. Moreover, it is shown that the fingerprints do not need to be very accurate for successful localization if enough microphones are dispersed in the environment.

Acknowledgements

First, I would like to thank Dr. Ricardo Burdisso and Dr. Marty Johnson for providing me with the opportunity to work on the interesting research project. Especially, I am deeply grateful to my advisor, Dr. Burdisso for the constant guidance in my work, and for the invaluable and critical feedback on my research and presentation skills. It has been a true privilege to be his student. I would also like to extend my appreciation to Dr. Michael Roan, Dr. Jeffery Reed, and Dr. Michael Buehrer for serving on my advisory committee and for their constructive comments throughout the course of this work.

Next, I would like to thank my fellow researchers at VAL, especially Stephanie Pasareanu, Marcel Remillieux and Joe Corcoran. They made my life at school more enjoyable and less stressful. It has been a pleasure working with them and I wish them all great success in their future academic and professional careers. I also wish to thank Gail Coe for her assistance in many administrative tasks.

I also thank my Korean fellows at Tech, Kyou Woong Kim, Kye Hun Lee, Hae Soo Kim, Jang Hoon Oh, Joo Hong Lee, Jeong heon Lee, Hyungjoon Cho and Jiduck Choi. Moreover, I would like to thank all the members at KBCB for their support and prayers.

The final word of acknowledgement is reserved for my parents, my brother and my wife. I owe everything I have achieved in my life to their love, patience and support.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background

The detection, characterization and localization of important events in complex urban environments, such as moving vehicles, gunfire and explosions, is critical to providing effective surveillance of military and civilian areas and installations. In urban environments, obstacles such as terrain or buildings introduce multipath propagations which make source localization challenging. To this end, multiple inexpensive nodes of simple acoustic sensors forming a distributed sensor network have been the preferred methodology to localize acoustic events (Brooks et al., 2004). Sensor networks potentially facilitate efficient surveillance systems since a large number of sensors enable the redundancy of the observations and close proximity to acoustic sources (Meesookho and Mitra, 2008). Acoustic sensors are also attractive due to long range capability, high-fidelity, no-line-of-sight to target requirement, and passive nature.

Current applications of this sensor network include surveillance, traffic control, industrial and manufacturing automation, and environment and health monitoring (Chong et al., 2003). Specifically, networked surveillance has potential in modern warfare to detect, identify, and locate events on the battlefield accurately which will enhance situational awareness, resulting in decreased casualties (Akyildiz et al., 2002). In addition, the sensor network can aid residential security by reporting gunshot information to the police (Mazerolle et al., 2000). For traffic control, cheap sensors with embedded networking capability can be deployed at every road intersection to detect and count vehicle traffic and estimate its speed (Estrin et al., 2001). Sensor networks can be also used for structural monitoring of buildings by identifying and even localizing damage within the structure (Caffrey et al., 2004). Underwater acoustic arrays have been used to track and study the habits of marine animals, such that human interference into animal habitats can be minimized (Szewczyk et al., 2004; Tiemann et al., 2006).

There have been many proposed acoustic localization techniques. Many techniques are based on time difference of arrival (Drake, et al., 2001), interaural level difference (Birtchfield, et al., 2005), and triangulations (Kaplan, et al., 2001) which passively exploit phase and/or amplitude differences between sensors. These methods are formulated assuming free field wave propagation and break down in more complex environments such as urban areas. The major difference between these environments and a free field is the presence of terrain and/or buildings that act as obstacles to acoustic wave propagation. These obstacles obstruct Line-of-Sight (LOS) and introduce multiple propagation paths due to scattering, reflection and diffraction of acoustic waves.

The problem of multipath propagation is well known and many statistical approaches have been proposed to mitigate the effect of clutter because the complexity made an exact analytical description of sound propagation unrealistic in the past (Carevica, 2001; Broadhead et al., 2000). However, recent advances in computational power enable numerical models of sufficient complexity to be calculated. In a known environment, a powerful method for acoustic source localization is time reversal refocusing (Fink 1992, 1997; Fizell et al., 1985). Due to the reciprocity of the wave equation, inverted signals emitted from the sensor locations converge on the original source location. Investigations of this method have been conducted using two dimensional (2D) numerical models of urban environments (Albert et al., 2005, 2010; Liu et al., 2007). Sensor-measured signals are back-propagated in a virtual environment using a finite-difference time-domain (FDTD) model that incorporates the geometry and propagation physics. In practice, however, a common point in time corresponding to the end of the received source signal must be designated and any network time desynchronization degrades the overall focusing ability. In addition, the acoustic model has to be estimated very accurately.

Beamforming using Matched Field Processing (MFP) was also investigated as a method to locate acoustic sources in urban environments using multiple microphone arrays. Matched Field Processing is a parameter estimation technique for localizing the range, depth, and bearing of a point source from the signal field propagating in an acoustic waveguide (Baggeroer et al., 1993).

It involves sophisticated modeling of the propagation and was first used in underwater acoustics, i.e. ocean acoustics (Tolstoy, 1993). In contrast to the typical underwater application, there is no closed form solution for the array manifold vectors in scattering environments due to the complex nature of the propagation. To this end, the array manifold vectors are obtained by measurement (Fialkowski et al., 2000; Johnson et al., 2006) or numerical modeling such as FDTD (Mennitt et al., 2010). The widespread availability of urban geographic information makes the numerical modeling attractive. Regardless of the employed approach, model mismatch is a concern. Naturally, localization performance is correlated with the quality of a priori information and there is always some uncertainty in the precise location and even presence of scatterers and fluctuation in the atmospheric conditions (temperature, humidity, wind, etc.) are unavoidable.

Some of the effects of model uncertainty have been studied in the context of both time reversal focusing (Liu et al., 2007, Yon et al., 2003) and matched field processing (Mennitt, 2008) in 2D environments. Liu et al. (2007) found that time reversal focusing was generally tolerant of minor changes in scatterer position if a large number of receivers (> 105) are employed. Yon et al. (2003) reported similar robustness to perturbations such as moving persons within a room. Mennitt (2008) investigated several uncertainties such as microphone calibration error, perturbation in temperature, and inter-array coherence.

From the previous brief literature review, there are yet unresolved issues in source localization in urban areas. Firstly, the localization work presented in the context of both time reversal focusing and beamforming has not been investigated in 3D environments. It is possible that 3D environments can provide additional information on the propagation between the source and the receiver, which would significantly affect the localization performance. There was no such an effort to investigate how those localization techniques work in 3D environments. Secondly, the only previous work using phased arrays and beamforming (Mennitt 2008) did not show super directivity characteristic as it was claimed, lacking in accurate localization capability in complex urban environments. Moreover, it was the only beamformer implemented in the study. This shows a limited investigation on beamforming algorithms for this type of application. Thirdly, previous research on model uncertainties is also limited. For example, uncertainty study by Yon et al. (2003) was performed on a very small scale assuming free-field environments. Uncertainty such as temperature or calibration error (Mennitt 2008) is too simple and can be easily detected and compensated. Uncertainty in the scatterer locations (Liu et al., 2007) is practically similar as calibration error. However, there are many other possible sources of model errors that have not been investigated. For example, it is typical that tactical vehicles constantly and randomly move in and out of street blocks for military operations. This results in unexpected addition or elimination of small scatterers in a given model. Wind is another major uncertainty and hard to estimate since wind has the effect of curving the sound rays. This leads potentially to large variation in sound level. Finally and more importantly, both time reversal and beamforming techniques need very accurate models. However, model complexity can increase dramatically to accurately capture all the details of the environment. This can be computationally burdensome and also hinder the potential as a real-time application in the field. Therefore, there is still a significant need for alternative methods that do not require very accurate models.

1.2 Dissertation Objectives

This dissertation addresses the above shortcomings of the current state-of-the-art technologies for source localization in urban environments. The first objective of this dissertation is to develop an effective 3D source localization technique for cluttered environment such as urbanized areas using a set of microphone phased arrays. To this end, the previous approach developed by Mennitt (i.e. beamforming using MFP) is extended from 2D to 3D environments. Several additional beamformers are further investigated since his proposed beamforming method did not show very accurate localization capability (no super directivity characteristics). A key hypothesis of this dissertation is that the localization technique must be investigated in 3D environments. The reasons for working in 3D are related to the sound propagation characteristic, number of propagation paths, and source-sensor configurations.

In 2D, sound propagates as cylindrical waves so the acoustic pressure decays with the inverse of the square root of the distance (cylindrical spreading law). On the other hand, in 3D environments the pressure decays with the inverse of the distance (spherical spreading law). The difference in the decay characteristic described is strictly applicable to free-field propagation but it also occurs in cluttered environments. An additional consequence of the different decay rates is that the microphone signal to noise ratio (SNR) is more favorable in 2D models. In other words, the SNR value will be higher in 2D than in 3D given the same amount of noise in the environment and distance of the array to the source. Hence, it is possible that a source localization technique will appear more effective in 2D than in 3D, in particular over long distances.

Another reason for performing the investigation in 3D environments is related to the number of possible propagation paths the signal takes from the source to the monitoring sensors. Due to the additional dimension in 3D environments, there can be substantially more propagation paths than in 2D environments. Thus, 3D localization approach will benefit from these extra paths, i.e. there is more information useful to the localization algorithm.

Finally, 2D localization is limited in the sense that the source and sensors are implicitly assumed to be at the same height from the ground. For a practical scenario where the source (e.g. shooter) is located at the top of a building, 3D models are needed to more accurately mimic this scenario.

The implementation of a microphone phased array based localization technique requires an accurate knowledge of the environment to account for all acoustic effects. There are, however, always model mismatches between the numerically modeled and the actual environment. Thus, the second objective in this dissertation is to investigate the sensitivity of the localization technique to uncertainties in the environment that cannot be accounted for easily. This study also investigates if multiple arrays can alleviate the negative effects of the errors in the model.

Previous research efforts in the context of time reversal focusing and matched field processing were focused on using accurate models with a few sensors. However, there is a need to seek alternative methods that do not rely heavily on the accuracy of the models. Hence, a source localization method that uses a low fidelity model of the environment in conjunction with large number of sensors is of interest to the community. Thus, the third objective of this dissertation is to develop an efficient source localization technique using dispersed microphones in the environment without relying on its accurate knowledge. The relationship between the localization accuracy and the number of dispersed microphones is also investigated.

1.3 Original Contributions

The key contributions of this work are addressed here.

Firstly, an accurate source localization method in 3D urban environments using a set of microphone arrays was developed in conjunction with a high fidelity model. There was no previous effort to investigate how beamforming using MFP performs in 3D environments. This work clearly showed significant importance of 3D over 2D localization. Several candidate beamformers were investigated and the EMVDR was identified as the only beamformer that had capability to accurately locate the source while still robust to uncorrelated noise in the environment.

Secondly, the effect of model uncertainties which cannot be easily corrected or accounted for and a method to overcome it were investigated using beamforming in conjunction with a high fidelity model. Detrimental effects of temperature gradient and unmodeled objects in the model were identified. Then, multi-array processing was used to overcome their detrimental impact and improved the localization performance.

Thirdly, an efficient source localization method in urban environments using dispersed microphones was developed in conjunction with a low fidelity model. The fingerprinting positioning technique was combined with an efficient diffusion-based energy modeling and applied to acoustic localization in a scattering environment. The investigation showed that the fingerprints do not need to be very accurate for successful localization if enough microphones are dispersed in the environment.

1.4 Outline of Dissertation

The dissertation is organized into seven chapters and five appendices as follows. Chapter 1 briefly discusses the general state-of-the-art on the research topic and the research needs. The objectives of this study are also listed. Chapter 2 discusses previous work in the field of acoustic source localization and the localization methods developed for different situations. Chapter 3 introduces two numerical modeling tools for acoustic propagation: a high-fidelity but computationally expensive FDTD method and a low-fidelity but computationally efficient diffusion-based energy method. These tools are used to provide a representation of the environment. This chapter also describes the models used for all the simulations. Two simple 2D models are discussed followed by the corresponding 3D versions of these models to allow for direct comparison between 2D and 3D results. Two relatively large and highly cluttered 3D models are then presented that are used for more realistic localization study: a simplified urban model consisting of regularly spaced streets forming urban canyons and Army Fort Benning model without any recognizable urban canyons or regular patterns. The accuracy of the energy method is also studied by comparing to the exact solution using the FDTD.

In Chapter 4, beamforming algorithms are presented in conjunction with MFP as effective localization techniques in complex urban environments. General beamforming theory is discussed followed by a review of several different beamforming algorithms. The Enhanced

Minimum Variance Distortionless Response (EMVDR) beamformer is then introduced as the most efficient beamforming method among them. The robustness to uncorrelated noise of the beamformers is also investigated. Finally, the localization error analysis is discussed.

In Chapter 5, the beamformers described in Chapter 4 are evaluated. The numerical investigations are performed using the simple 2D and corresponding 3D models described in Chapter 3 for the initial assessment and down selection of the most promising beamformers. The selected beamformers are used for further evaluation in the more realistic complex 3D urban models. Localization with the more realistic urban scenario by adding uncertainty to the model is then considered. In other words, the influence of such uncertainty on the localization performance is investigated. Such uncertainties include temperature gradient and unmodeled objects.

Chapter 6 investigates source localization using dispersed microphones in conjunction with the low fidelity energy-based acoustic propagation model. Acoustic fingerprinting is proposed because it can be integrated well with the energy-based method to model the environment. Similarly, the localization performance of this proposed method is then evaluated on the two 3D urban environments. The effect of the accuracy of the fingerprints on the source localization performance is also addressed using the approximate (using the energy-based acoustic model) and exact fingerprints (using the FDTD acoustic model). Finally, the accuracy and confidence in the source localization of the proposed method is investigated.

Chapter 7 summarizes the work presented in this dissertation. The research outcomes are briefly discussed. Contributions to a field of acoustic source localization are then addressed. In the end, some comments on future directions are discussed. Supplementary simulation results including some of mathematical derivations are provided in Appendices at the end of this dissertation.

Chapter 2 Literature Review: Acoustic Source Localization

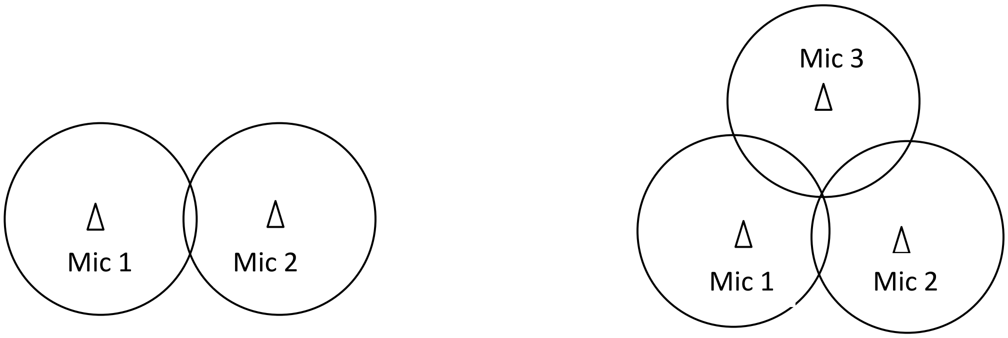
In this chapter, a literature review of acoustic source localization methods is provided. There are two main categories: Free-field methods and Model-based methods. Free-field methods are based on free-field acoustic propagation and have been investigated over the years. On the other hand, model-based methods have gained interest more recently. Some methods are based on statistical propagation model. However, the most recent methods take advantage of precise numerical models of the environments. Two commercially available shooter localization systems are introduced at the end of the chapter as real-market applications.

2.1 Free-field methods

These methods assume free-field acoustic propagation where sound freely propagates and spreads uniformly without reverberation. Hence, the accuracy of these methods is sensitive to clutter. That is, the performance will be degraded due to multipath, reflection and diffraction due to obstacles in the cluttered environment.

2.1.1 Time of Arrival (ToA)

One of the most popular methods for source localization is the Time of Arrival (ToA) (Foy, 1976). The ToA is the most basic positioning technology in time-based systems. The distance between the source and each sensor is calculated by ToA multiplied by the speed of sound (Girod and Estrin, 2001). Ideally, three ToA estimates are enough for localization using a trilateration method. As shown in Figure 2.1a, two ToA case gives two intersection points of the circles. On the other hand, Figure 2.1b shows that three ToA estimates resolves the precise position to a single point. This method is accurate in a free-field environment due to the linear relationship of the distance between the source and each sensor. The challenge of this method is time synchronization between the source and each sensor. In general, synchronization can be accomplished with radio signals among sensors (Girod and Estrin, 2001). It could become a serious problem, however, if there is no prior information about the source because the sensors would not know the exact time when the acoustic event happened. The accuracy of this method is confined by the assumption of a LoS condition (Wang et al., 2003, 2005).



(a) two ToA case (b) three ToA case

Figure 2.1: Schematic diagram of trilateration with ToA measurements.

The performance of the method, however, can be influenced by reverberation in a more complex environment. In fact, there exists much work in the literature on how to deal with reverberation (Stephenne et al., 1999; DiBiase et al, 2000; Julliard et al., 2005; Mak et al., 2008, 2009). Stephenne et al. (1999) proposed a cepstral prefilering technique to attempt to deconvolve the effects of reverberation prior to estimating the position. The experiment performed in his work was, however, limited to colored Gaussian noise as a source due to its sensitivity to the high variability and non-stationarity of speech signals. DiBiase et al. (2000) proposed a filter-and-sum technique which exploits microphone redundancy by combining the microphone signals, rather than combining a multitude of ToA estimates, to enhance the accuracy of location estimation. The obvious drawback is the increase in computational requirements for multiple microphones. Julliard et al. (2005) applied holography to cope with reverberation. Holography was initiated as an interferometric technique for recording the amplitude and phase of a coherent wave. A recording of this interference pattern is called a hologram (Gabor, 1948). Estimation of ToA is obtained based on phase sensitive spectral correlation between holograms. This technique is, however, limited due to large amount of data that needs to be acquired.

Recently, a Non-Line-of-Sight (NLoS) localization technique using multiple microphones in a known environment has been proposed (Mak et al., 2008, 2009). In this technique, ToA is estimated assuming the signal from source firstly arrives at each microphone through its shortest path, rather than multiple paths. NLoS errors are then predicted and corrected recursively by considering the shortest detectable paths between source and the microphones in a map. The main limitation of this technique is that the source has to be known to the receiver and the developed system is only suitable for a small environment due to high path loss.

2.1.2 Time Difference of Arrival (TDoA)

Another time-based localization scheme is time difference of arrival (TDoA) measured with pairs of microphones assuming that the speed of sound in the medium is known (Brandstein, et al., 1995). The TDoA indicates time difference of arrival of an event produced by a source between a pair of microphones, i.e. differences between ToAs. The TDoA method has advantage over the ToA method in a sense that it requires time synchronization between microphones only.

The TDoA-based source localization method consists of a two-step procedure. First, the TDoA between microphones is estimated through the use of time delay estimation (TDE) techniques. Time delay estimation is usually accomplished by taking a snapshot of the synchronized microphone signals. Then, the cross-correlation of pairs of microphone signals is performed. The peak in the cross-correlation gives the TDoA (Omologo et al., 1996; Svaizer et al., 1997). The estimated TDoA's are then transformed into estimated distance difference between the source and microphones as shown in Figure 2.2a. Secondly, this distance difference estimate defines a hyperbola between each pair of microphones on which the source may exist. If this procedure is carried out again with another pair of microphones, another hyperbola is defined.

The intersection of more than two hyperbolas results in the position location estimate of the source as shown in Figure 2.2b. For three dimensional estimates, at least three hyperbolas are required to locate the source accurately (Brandstein and Ward, 1995). The problem of finding the intersection of multiple hyperbolas is solved as a set of nonlinear equations (Chan et al., 2006). A close-form approximate solution to the non-linear problem was also developed, which is well-suited for real-time implementation (Chan et al., 2006). Unfortunately, the point of intersection of two hyperbolas can change significantly based on a slight change in the eccentricity of one of the hyperbolas due to noise and NLoS propagation errors.

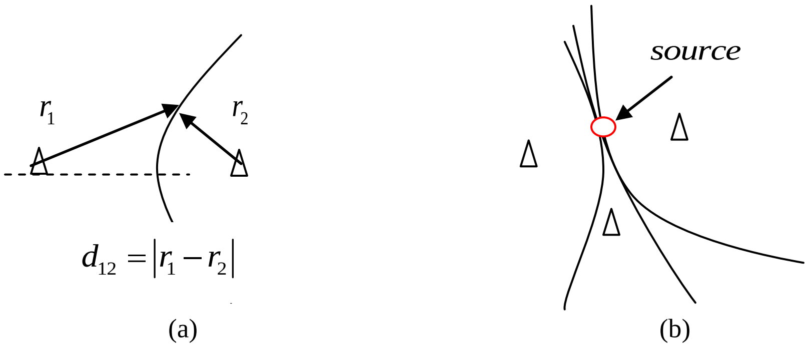


Figure 2.2: Hyperbola drawn based on TDoA between two microphones (left) and location estimate found at the intersection of three hyperbolas (right).

Although the TDoA-based localization scheme is reasonably effective in moderately reverberant environments as compared to the ToA-based scheme (Brandstein and Ward, 1995), the performance is still affected by reverberation. To this end, Sasaki et al. (2006) and Valin et al. (2007) used a beamformer to directionally filter out echoes received by numerous microphones and concurrently locate the sound source in a reverberant environment. Specifically, a conventional delay and sum beamformer is used to calculate the spatial spectrum of the received signal. Then, frequency band selection is performed to filter out a loudest sound from the spatial spectrum until there is no stronger sound remaining after previous filtering. This method localizes multiple sounds from the highest power intensity to the lowest at each step. The main drawback is that they assume there is LoS between the sound source and microphones.

2.2 Model-based methods

This approach enables the inclusion of acoustic propagation models into the localization systems.

2.2.1 Received Signal Strength Indication (RSSI)

Received Signal Strength Indication (RSSI) is a simple but effective source localization approach (Seidel and Rappaport, 1992; Bahl and Padmanabhan, 2000; Ladd and Cooper, 2004; LaMarca et al., 2005; Ferris et al., 2006). Early work on RSSI methods have been applied to radio-frequency (RF)-based localization systems where the sensors are to be localized. Later, this RSSI method has gained interest in the field of acoustic source localization (Li et al., 2003; Sheng et al., 2005; Meesookho et al., 2008). In their work, statistical propagation models rather than physical based approaches are used for acoustic energy attenuation with distance from the source.

The RSSI method uses the signal strength, rather than time delay, information gathered at multiple sensors placed in the environment. Often, the received signal strength (RSS) is equivalently reported as measured signal power. This RSS can be acoustic, RF, or any other signal. In this method, the RSS at each sensor is computed from the time history of the received signal. Then it is transmitted to a fusion center for combination with the RSS from other sensors. Finally, it computes the estimated location by converting the signal propagation loss into a distance with the help of both theoretical and empirical models. Therefore, a predictive propagation model is essential to determine the relationship between signal strength and distance from the source.

For example, a conventional log-distance RF propagation model has been used extensively in the literature. Rappaport (1996) proposed the following empirical model

$$RSS(d) = P_s(d_0) - 10\eta \log_{10}(d / d_0) + X \qquad (dBm) \tag{2.1}$$

where $RSS(d)$ is the RSS value at distance d from the source, $P_s(d_0)$ is the signal strength at a reference distance d_0, η is the path loss exponent, and X is a random attenuation due to clutter which is represented as a Gaussian random variable with zero mean and σ^2 variance. The source power is typically unknown because the source is arbitrary to the sensors. The variance describes the shadowing due to topographical obstructions. With the random variable, the model can describe the random shadowing effects which occur over a large number of measurement locations. The loss parameter η depends on the cluttered environment (buildings, etc.). For instance, in free field, η is equal to 2, and when obstructions are present, η will have a larger value (Rappaport, 1996). In practice, the value of η and σ^2 are computed from measured data. The unit of the RSS is dBm which indicates the power ratio in dB of the measured power referenced to 1mW. For example, 3dBm means 3dB higher than 1mW, which is 2mW. This unit is widely used in RF signals because of its capability to express both very large and very small values in a short form.

Given the measured RSS value and using eq.(2.13), the estimated distance from the source can be computed as

$$d = d_0 \cdot 10^{\left(\frac{P_s(d_0)+X-RSS(d)}{10\eta}\right)} \tag{2.2}$$

Assuming there are at least three sensors, the source location can be then estimated by using trilateration method to determine the position of an object based on simultaneous distance measurements from three stations located at known sites (Eberly, 1996). It can be trivially expressed as the problem of finding the intersection of three spheres (Coope, 2000).

As mentioned, the RSSI method has been focused on RF signals. To this end, a variety of radio propagation models have been developed for both indoor and outdoor applications (Anderson et al, 1995; Neskovic et al., 2000; Iskander and Yun, 2002). These models are empirical in nature so the collection of data has to be sufficiently large to provide enough accuracy of the model. The key problem of this RSSI method, however, stems from the complexity of signal propagation through space, especially in the presence of obstacles such as buildings. For example, it is difficult to estimate the loss parameter η and Gaussian random variable X needed in eq. (2.13). Moreover, the source signal can have arbitrary characteristics which make the source power and spectral content of the signal unknown to the sensors. Without knowing the source power, the distance cannot be estimated in eq. (2.14). To this end, Meesookho et al., (2008) introduced Gaussian random process to deal with arbitrary static source localization. Still, the correlation function was assumed to be known.

2.2.2 Time Reversal Refocusing

In a known environment, a powerful method for acoustic source localization is time reversal refocusing (Fink 1992, 1997; Fizell et al., 1985). Due to the reciprocity of the wave equation, inverted signals emitted from the sensor locations converge on the original source location. Figure 2.3 shows that the inverted signals at two different sensors are emitted back to the environment, which in turn converges on where the actual source is. Time reversal acoustics have been developed recently in many fields such as medical ultrasound (Tanter et al. 2008). This method is often applied in the physical medium itself, for example to focus waves on the tone in lithotripsy (Thomas et al., 1996) or remove reverberation in underwater communication (Kuperman et al., 1998). Recently, investigations of this method have been conducted using 2D numerical models of urban environments (Albert et al., 2005, 2010; Liu et al., 2007). Sensor-measured signals are back-propagated in a virtual environment using finite-difference time-domain (FDTD) model that incorporates the geometry and propagation physics.

Time reversal processing involves the following steps: first, the sound signature produced by a source is recorded at a number of sensor locations after propagation through the complex medium; next, the time series signatures are reversed in time; finally, the reversed time series are emitted from the sensor locations and propagate back through the complex medium. Because of the symmetry of the wave equation, this procedure would refocus acoustic energy at the original source location.

In practice, however, a common point in time corresponding to the end of the received source signal must be designated and any network time desynchronization degrades the overall focusing ability. Moreover, Mennitt (2008) suggested that dissipation in the medium breaks the invariance of the wave equation and, practically, the time reversal operation is achievable only on a limited aperture time reversal mirror that results in decreased focusing quality in diffracting media. Additionally, the effect of uncorrelated noise in the environment can be detrimental, which was ignored in the recent localization work (Albert et al., 2005; Liu et al., 2007).

Another major concern with time reversal is its overall focusing ability related to the number of sensors in the network. Albert et al. (2005) used only eight NLoS sensors for simulation to demonstrate its accurate localization. However, there was no investigation regarding the number of sensors needed to achieve a certain localization performance.

The effects of model uncertainty have been studied in 2D numerical models (Liu et al., 2007). Liu et al. (2007) found that time reversal focusing was generally tolerant of building location errors if a large number of sensors are employed, as with the 105 sensors considered in his work. It has been also shown that time reversal refocusing with only a few NLoS sensors is still possible, though with degraded resolution, if the building position errors are minor.

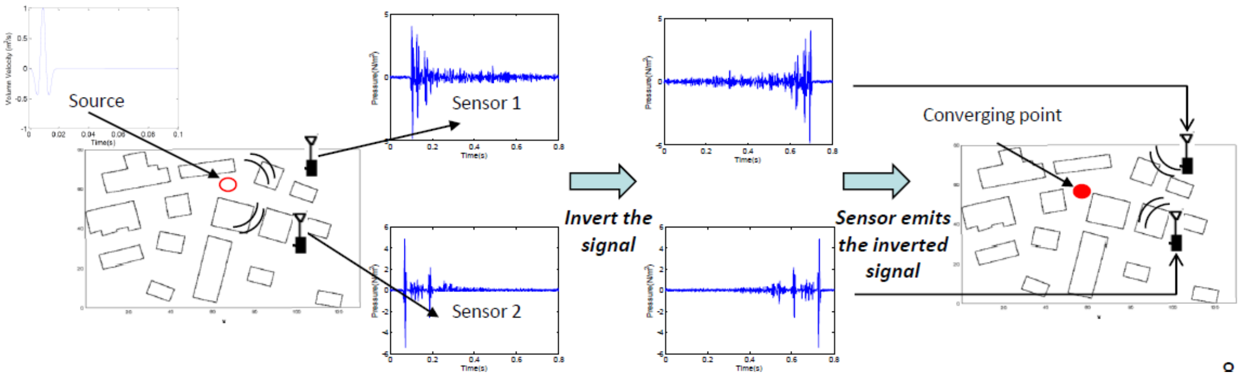


Figure 2.3: Example of time reversal refocusing with two sensors in the environment.

2.2.3 MFP-based Beamforming

Microphone phased arrays in conjunction with MFP have been studied as a method to locate an acoustic source in cluttered environment (Mennitt et al., 2010). The MFP accounts for the unique signature of the reverberant acoustic field at a receiver such that location of a radiating source can be determined anywhere in the known environment. In fact, it adds diffraction plus attenuation to a simple plane wave model for the propagation. Since there is no closed form solution for the array manifold vectors in scattering environments due to the complex nature of propagation, the array manifold vectors are obtained by numerical modeling such as FDTD.

Mennitt and Johnson (2010) recently investigated a phased array localization approach in 2D numerical models using what was basically a conventional delay-and-sum beamformer, though it was referred as minimum variance distortionless response (MVDR) beamformer. In their work, only one beamformer was investigated. Moreover, the proposed beamformer did not show super-directivity characteristic (i.e. poor resolution)

Similarly to the previous time reversal refocusing, the effect of uncertainty has been studied in context of MFP-based beamforming (Mennitt and Johnson, 2010). In particular, uncertainty in array location, temperature, and various parameters affecting inter-array coherence were investigated. It was found that uncertainty in array location leads to a perturbation of the steering vectors, causing the localization performance degradation. Like an uncertainty in location, uncertainty in temperature can be interpreted as a phase uncertainty, having a similar effect. On the other hand, uncertainty affecting inter-array coherence was shown to intensify the effect of other uncertainty rather than disrupt the localization alone.

This investigation on model uncertainties, however, is limited in a sense that uncertainty such as temperature or array location can be easily detected and compensated. Moreover, uncertainty affecting inter-array coherence can be neglected in large urban environments because sound signal that propagates over a long distance through a cluttered environment will eventually become uncorrelated.

2.2.4 Fingerprinting

Fingerprinting positioning methods provide an alternative solution to the location determination problem (LaMarca et al., 2005; Ferris et al., 2006). Fingerprints are unique signatures that are collected a priori. Theoretically, any consistent information can be used as part of the signature. In a static environment, a source generates a unique signature at each location in the environment.

Fingerprinting position systems involve two steps process. Firstly, a set of predefined reference locations (grid points forming a uniform grid to cover the entire area with the desired resolution) are used to collect the signal strength assuming the source is located at each reference location. A database or table of predetermined signal strengths at the grid points is then created. The signal strength vector, which consists of signal strength at each receiver, is normalized so that the Euclidean length of the vector is equal to one. The normalized signal strength vector is referred to as the fingerprint of that point. This normalization enables direct comparison with each fingerprint without any knowledge of source power.

Secondly, a new fingerprint (or pattern) is detected with the sensors at unknown source location. The positioning algorithm tries to find the best match between the currently observed fingerprint and the reference fingerprints in the database. The most common algorithm to estimate the location computes the Euclidean distance between the measured signal strength vector and each reference fingerprint in the database. The coordinates associated with the fingerprint that provides the smallest Euclidean distance is returned as the estimate of the source location. This is known as the Nearest Neighbor algorithm (Gutin et al., 2002). In fact, two versions of Nearest Neighbor can be used referred to as unconstrained and constrained search-space. The unconstrained search-space looks at the entire fingerprint map to find the closest match. Constrained search-space only searches within a given distance from a previously predicted location. The idea is that a moving object can only travel up to a maximum distance from its previous location within the time it takes to collect the signal strength from sensors and searching through the entire map is unnecessary (Navarro and Makinen, 2009). Re-calibration due to a possible dynamic environment can be avoided using constrained search-space.

The maximum accuracy that can be achieved is directly related to the resolution of the fingerprinting database. Obviously, a source cannot be localized in an area for which no fingerprint is available. It is also impractical to construct the entire database for a large model, in particular experimentally. Therefore, the fingerprinting position system, in general, requires dense training coverage and extrapolation to areas not covered during training (Paul et al., 2009). To this end, such as spatial smoothing (LaMarca et al., 2005) and Gaussian processing (Schwaighofer et al., 2003; Ferris et al., 2006) have been proposed to overcome this issue. LaMarca et al. (2005) introduced a hierarchical Bayesian technique for learning local Gaussian likelihood models of signal strength. An important aspect of this method is that the spatial smoothing correlates the signal strengths measured at neighboring locations. Gaussian processing is used to generate an observation model for signal strength measurements from calibration data. Once the observation model is built, the location of the source can be estimated from signal strength measurements. This method is able to interpolate between data points while correctly estimating the resulting uncertainties in predictions. More detail will be discussed in Chapter 6.

2.3 Commercial Acoustic Localization Systems

This section briefly introduces two shooter localization systems as real-market applications. The technology used in these systems is first described (Gourley, 2009). Their limitations in urban areas are then addressed.

2.3.1 Boomerang

The Boomerang system was developed in response to a Department of Defense request for an affordable acoustic sniper detector. This system has been constructed for mounting on vehicles deployed in conflict areas. Such system employs arrays of seven small acoustic sensors where six sensors are distributed over the surface of a sphere and the seventh sensor located in the center of the sphere (see Figure 2.4).

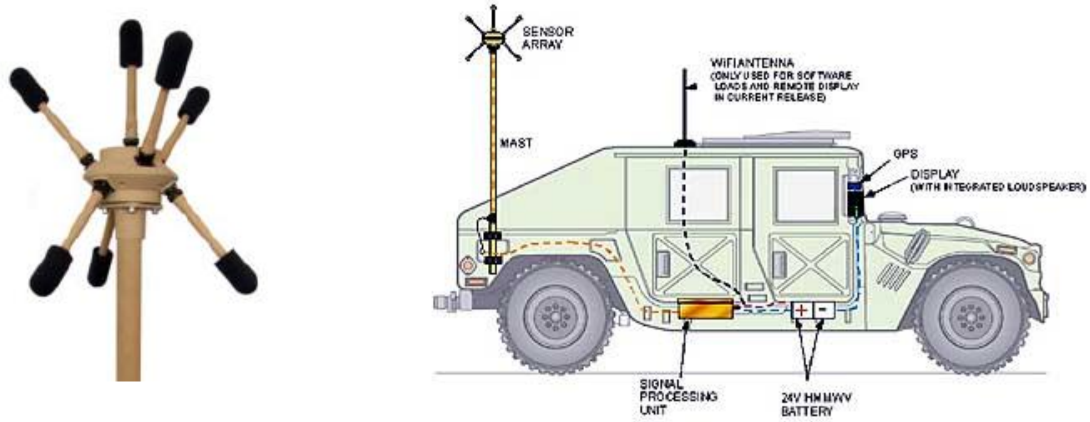


Figure 2.4: Boomerang System (left) and how it is mounted on the vehicle (right) (Hanlon, 2005).

Assuming that bullets travel faster than the speed of sound, a gunshot generates two important acoustic events: a shock wave and a muzzle blast as shown in Figure 2.5. It is noteworthy that the localization techniques described earlier work on the muzzle blast only assuming an arbitrary source signal.

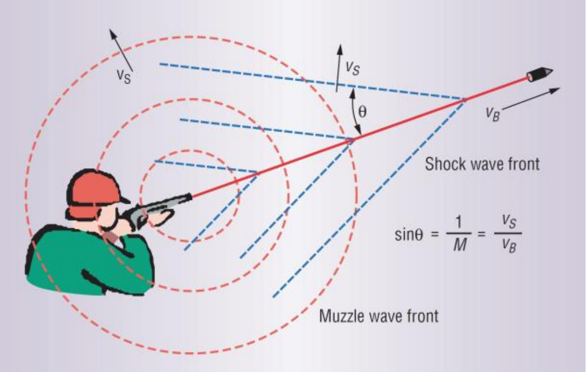


Figure 2.5: Acoustic events generated by a shot. (Maroti et al., 2004).

The supersonic projectile generates a shock wave in every point of the trajectory, producing a cone-shaped wave front assuming the speed of the projectile is constant V_B. The projectile's Mach number M determines the angle of the shockwave cone. On the other hand, the muzzle blast produces a spherical wave front that travels at the speed of sound (V_S) from the muzzle. As a projectile passes the system, its shock wave and muzzle blast will reach different sensors in the array at different times, depending on the trajectory of the projectile. By comparing arrival times of signals received at different sensors of the array similar to TDoA, the trajectory of a projectile may be determined. Through a series of mathematical calculations, the trajectory of the projectile may be extended back to the source of the projectile, revealing the location of the shooter who launched the projectile. The TDoA measurement due to the muzzle blast is not necessary but useful for localization. When the TDoA is taken as the difference between the muzzle blast and shock wave arrival times at each sensor, the resulting localization performance can be independent of time synchronization and be less affected by geometry (Whipps et al., 2009).

Challenges to overcome were the following: 1) to filter out noise from the vehicle on which it could be mounted (such as loud engines and static sounds from the radio), 2) to ignore sounds similar to that of a gunshot (such as fireworks or a car back-firing), and 3) to ignore outgoing fire from friendly troops.

Still, the performance can be significantly degraded due to NLoS error in TDoA measurements for a complex urban environment. In fact, this system was tested in a real environment and verified to localize fire from AK-47s and other small arms at ranges from 50 to 150 *m* but in a very limited urban environment where there is always LoS with low number of buildings in the area (Gourley, 2009). Another limitation comes from 2D-based assumption, i.e. the constant bullet trajectory and speed. In 3D environments, this assumption is no longer realistic and therefore, a degrading bullet speed and the effects on the trajectory due to the gravity should be taken into account to improve the localization performance.

2.3.2 Shoulder-Worn Acoustic Targeting System (SWATS)

A system called Shoulder-Worn Acoustic Targeting System (SWATS) is a warfighter-wearable variation of the previous localization system as shown in Figure 2.5. The device consists of a small sensor box worn on the shoulder, connected by wire to an earpiece and a small visual display which is strapped to the chest. Unlike the previous Boomerang system, four microphones are used, which are contained in a small package measuring 20 cubic inches.

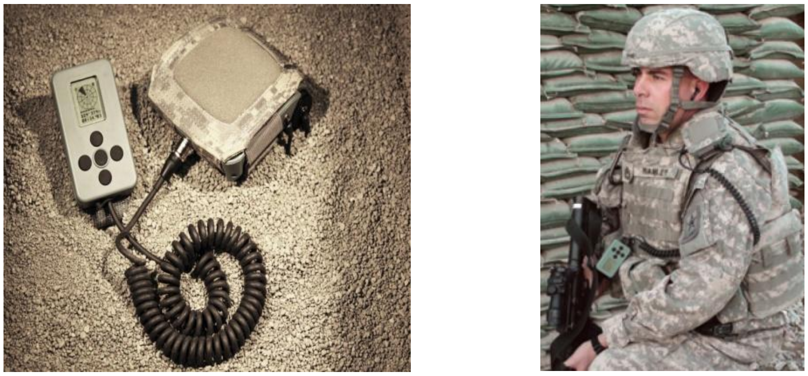


Figure 2.6: SWATS (left) and installation of the device (right) (Gourley, 2009)

This system also works by detecting the shockwave of a bullet in flight and the muzzle blast. It then goes through a large number of algorithms and, after confirming that those two events belong together, derives the approximate bearing and distance of an enemy shooter’s position. The detail algorithms are limited due to the classified information. Relevant technical detail is reported by Barger and Stanley (2008). The specification of the sensor is described in table 2.1.

Table 2.1: Specifications of the SWAT sensor (Barger and Stanley, 2008)

Dimensions	Weight	Operating Temperature	Storage Temperature	Humidity	Power Requirements
7.6cm x 7.6cm x 1.9 cm	0.45 kg	-20°C to 60°C	-20°C to 70°C	5 to 95%	<1W

In urban terrain, the sound of a muzzle blast echoes off buildings and vehicles which makes calculating range challenging. Since this system is located closer to the ground as compared to the Boomerang, the ground reflection of the shock wave with a delay corresponding to the path length difference may cause detrimental effect on the localization performance.

In overall, since these systems are based on free-field acoustic propagation, some performance degradation due to clutter in the environment are expected. Note that the localization techniques which will be proposed later in the dissertation are model-based and therefore intended for complex urban environments where there is not direct LoS to the source.

Chapter 3 Numerical Modeling Methods for Acoustic Propagation

Because of its major importance, noise propagation in urban areas has been studied recently. A history of various sound propagation models can be found by Picaut et al. (1999). Urban propagation is a very complicated problem due to the presence and often irregular structure of many intersecting streets and variously shaped buildings. The complexity of the environment makes an exact description of the sound propagation very challenging and unrealistic. Fortunately, advances in computational power enable numerical methods to be used in the source localization techniques. In this chapter, two acoustic numerical tools are presented: the FDTD and energy method. The FDTD is a high fidelity model used in previous research. Here, the 3D FDTD is presented for more realistic acoustic simulation. This FDTD model will be used in conjunction with the beamforming-based source localization technique described in Chapter 4. On the other hand, the energy method has been developed more recently. It is a low fidelity but computationally efficient model. This energy method will be used as a tool in combination with the acoustic fingerprinting source localization technique discussed in Chapter 6. Several numerical models are then presented in 2D and 3D in section 3.3. These models are used in the investigation of the proposed localization techniques.

3.1 Finite Difference Time Domain (FDTD)

The FDTD is a high fidelity modeling tool and it was used in previous source localization techniques. This FDTD method solves the acoustic wave equation using finite difference, e.g. equation is solved stepwise in time and space. Because the equations are solved at each discretized point and each time step, the FDTD method is flexible with regard to changing parameters spatially or temporally. The FDTD method allows for the creation of high fidelity models which can predict the acoustic field in the presence of complicated phenomena such as scattering obstacles, transient sources, and a moving inhomogeneous medium. Unfortunately, this method suffers from excessive computational time and memory requirements as frequency of interest increases. Regardless, due to the superior flexibility, accuracy, and ease of implementation, the FDTD is chosen for this work.

For more realistic acoustic simulation, the 2D FDTD code was extended to 3D using the three dimensional coupled wave equation of pressure and velocity (Corcoran et al., 2010). For the sake of completeness, a brief description of the formulation of the equations governing the 3D FDTD model is presented here. Atmospheric attenuation is not included in this formulation. The derivation begins with the differential wave equation and Euler's equation in 3D as

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 \cdot p \tag{3.1}$$

and

$$\nabla \cdot p = -\rho \frac{\partial \vec{v}}{\partial t} \tag{3.2}$$

where p is the acoustic pressure; $\vec{v} = v_x \hat{i} + v_y \hat{j} + v_z \hat{k}$ is the particle velocity vector in 3-D; $\hat{i}$, $\hat{j}$, and $\hat{k}$ are the unit normal vectors in the x, y, and z-directions; t is the time variable; $\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$ is the gradient operator; ρ is the density of the medium in which the waves propagate; c is the speed of sound in that medium. Substituting eq. (3.2) into eq. (3.1) and integrating once with respect to time gives

$$\frac{\partial p}{\partial t} = -\rho c^2 \nabla \cdot \vec{v} \tag{3.3}$$

Before using finite difference to solve the partial differential equations, eq. (3.3) is rewritten in terms of velocity components as

$$\begin{cases} \dfrac{\partial v_x}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial x} \\ \dfrac{\partial v_y}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial y} \\ \dfrac{\partial v_z}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial z} \end{cases} \tag{3.4}$$

Equations (3.3) and (3.4) are discretized in time and space as shown in Figure 3.1. Each pressure node is positioned at point $\left(i\Delta x, j\Delta y, k\Delta z\right)$ where *i, j*, and *k* are integer indexes and Δx , Δy , and Δz are the spatial steps or grid spacing in three orthogonal directions. Velocity nodes are offset by half of a spatial step from the corresponding pressure node with the same indexes as shown in Figure 3.1. Therefore, the nodes which contain the x-component of velocity are positioned at $\left(\left(i+\tfrac{1}{2}\right)\Delta x, j\Delta y, k\Delta z\right)$, the y-component at $\left(i\Delta x, \left(j+\tfrac{1}{2}\right)\Delta y, k\Delta z\right)$, and the z-component at $\left(i\Delta x, j\Delta y, \left(k+\tfrac{1}{2}\right)\Delta z\right)$. The temporal index is *n* with temporal spacing or time step Δt. Each pressure is known at times $n\Delta t$, while velocities are again offset and known at times $\left(n+\tfrac{1}{2}\right)\Delta t$.

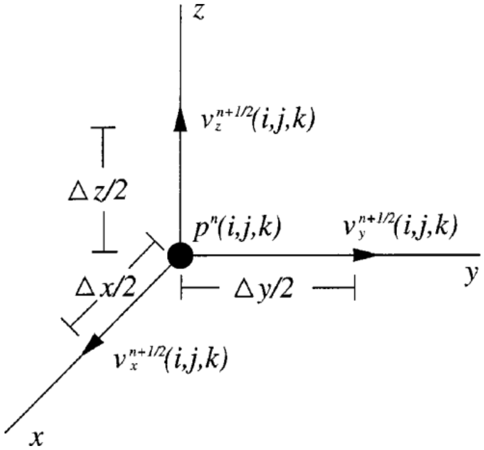


Figure 3.1: Diagram of partial differential equation discretizing scheme to obtain FDTD equations (Tolan and Schneider, 2003).

Applying the above scheme, the discretized forms of eqs. (3.3) and (3.4) for constant grid spacing $\Delta x = \Delta y = \Delta z = \Delta$ are

$$p^{n+1}(i,j,k)=p^{n}(i,j,k)-\Delta t\rho c^{2}\left(\begin{array}{l}\dfrac{\left[v_x^{n+1/2}(i,j,k)-v_x^{n+1/2}(i-1,j,k)\right]}{\Delta}\\+\dfrac{\left[v_y^{n+1/2}(i,j,k)-v_y^{n+1/2}(i,j-1,k)\right]}{\Delta}\\+\dfrac{\left[v_z^{n+1/2}(i,j,k)-v_z^{n+1/2}(i,j,k-1)\right]}{\Delta}\end{array}\right) \tag{3.5}$$

and

$$\begin{aligned}v_x^{n+1/2}(i,j,k)&=v_x^{n-1/2}(i,j,k)-\frac{\Delta t}{\rho}\left(\frac{\left[p^{n}(i+1,j,k)-p^{n}(i,j,k)\right]}{\Delta}\right)\\v_y^{n+1/2}(i,j,k)&=v_y^{n-1/2}(i,j,k)-\frac{\Delta t}{\rho}\left(\frac{\left[p^{n}(i,j+1,k)-p^{n}(i,j,k)\right]}{\Delta}\right)\\v_z^{n+1/2}(i,j,k)&=v_z^{n-1/2}(i,j,k)-\frac{\Delta t}{\rho}\left(\frac{\left[p^{n}(i,j,k+1)-p^{n}(i,j,k)\right]}{\Delta}\right)\end{aligned} \tag{3.6}$$

Note that the superscripts indicate iterations rather than powers in eqs. (3.5) and (3.6). The time offset by half a step allows for the calculation of pressure from velocities at a previous time, and the calculation of velocities from the updated pressures. This is repeated for each successive time step. A number of these iterations are run before pressure data is stored in the computer memory. These nodes are defined by spatial ratios of the number of grid points to the number of observers in the *x*, *y*, and *z*-directions. The code propagates until a desired number of time steps is reached at which point data is saved. The total number of time steps the simulation runs is equal to the number of time steps to iterate before saving data multiplied by the number of time steps to save data.

After selecting a uniform grid point with equal spacing for simplicity, the time step is calculated based on the Courant condition (Tolan and Schneider, 2003), which restricts the time step in order for a 3D FDTD simulation to have a stable solution as shown below,

$$\Delta t \leq \frac{\Delta}{c\sqrt{3}} \tag{3.7}$$

Due to this discretization (either temporal or spatial) inherent in FDTD simulations, there is an upper frequency limit for accurate simulations. In the temporal case, the maximum frequency of the simulation, $f_{\max}$, must be lower than the Nyquist frequency, f_N , given by

$$f_{\max} < f_N = \frac{1}{2\Delta t} \tag{3.8}$$

This limits the application of the 3D FDTD code to relatively low frequencies due to computational time restrictions (Corcoran et al., 2010).

3.2 Energy-based Method

The energy-based method has been developed for room acoustics under the assumption of a diffuse sound field in order to predict the sound pressure levels. Investigation to extend the diffuse sound field concept to the prediction of sound propagation in urban areas has been undertaken by Picaut et al., (1999). This method has been originally introduced as a more computationally efficient method than the FDTD. However, predicting the proper time-delay between an input source and a receiver positioned in the computational domain is not achievable. Therefore, the focus was on the steady-state diffusion equation and a 3D energy code based on the acoustic diffusion equation has been implemented recently by Pasareanu et al. (2011). It has been demonstrated that this propagation model provides fairly accurate predictions of the sound pressure level amongst buildings in urban areas, despite the lack of modeling of edge diffraction at building corners. Detrimental effects of that flaw in the model have been shown significant for receivers close to the source only. This modeling tool is still being investigated and further improvements are expected in the future. The major advantage of the model is that it allows fast computation of the sound field in large and complex urban environments (Choi et al, 2012). For instance, the computational time of running this energy-based method for the model in Figure 3.3 is just a few seconds, which is significantly faster than FDTD.

The diffusion equation without any medium loss is written as (Pasareanu et al., 2010)

$$\frac{\partial e(\vec{r},t)}{\partial t} - D\nabla^2 e(\vec{r},t) = W(t)\delta(\vec{r}-\vec{r}_s) \tag{3.9}$$

where $W(t)$ is the acoustic power of the source, modeled as a point source at position $\vec{r}_s$ in the environment, *e* is the energy density (internal energy per unit volume), and D is the diffusion coefficient. In the case of the steady-state, the diffusion equation simplifies to:

$$-D\nabla^2 e(\vec{r},t) = W\delta(\vec{r}-\vec{r}_s) \tag{3.10}$$

A popular finite difference scheme, the 5-point scheme, is then used to solve the eq. (3.10). In 2D problem, this scheme is based on the calculation of the value at each point using its four neighbors (Leveque, 2007). Extended to 3D, the computation of the solution at each point of a domain is done based on all six neighbors. More details on the solution of the eq. (3.10) can be found in the reference (Pasareanu et al., 2010).

The output data obtained from the solution of the system is the energy density. In turn, the sound pressure level can be obtained from the relationship between the energy density and the root-mean-square (rms) pressure, p_{rms}. That is,

$$p_{rms}^2 = e\rho c^2 \tag{3.11}$$

and

$$SPL(dB) = 10\log_{10}\left(\frac{p_{rms}^2}{p_{ref}^2}\right) \tag{3.12}$$

where p_{ref} is the reference pressure.

The energy method does not account for the constructive/destructive interference effect of the multiple waves present in the sound field, e.g. direct, reflected, diffracted waves. Thus, it is applicable for the estimation of the acoustic energy distribution in frequency bands such as 1/3rd octave bands.

3.3 Numerical Models

Both 2D and 3D models are introduced here to evaluate the localization methods which will be discussed further in Chapters 4 through 6. Firstly, 2D single and three-building models are introduced. Then, the same models are extended to 3D to allow for direct comparison between 2D and 3D results. This comparison will allow establishing the importance of assessing the localization performance in 3D environments. Finally, two large 3D models are built for more realistic localization study. All models assume perfect reflections from the buildings and ground, e.g. no acoustic losses. Perfect absorption was assumed at the boundaries of the domain.

3.3.1 2D Models

First, a single rectangular object (e.g. building) is introduced in a $40m \times 35m$ area as shown in Figure 3.2. The object has a size of $2m \times 9m$ and it will obstruct the direct LoS from the sources to the array. The sound field in the shadow region due to the object is the result of only edge-diffraction effects. This model represents the simplest case with minimum complexity and therefore was used for initial investigation on the beamformers.

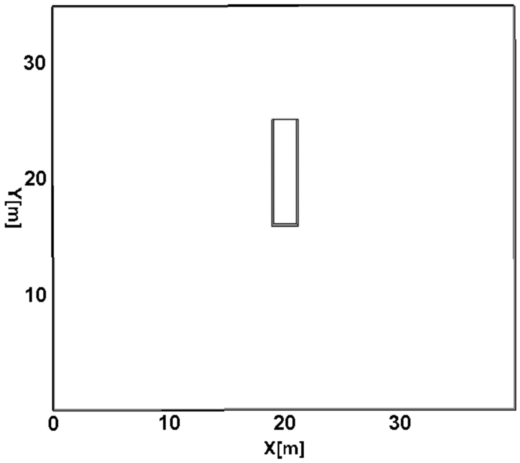


Figure 3.2: 2D model with single obstructing building.

The second model consists of three buildings in a $50m \times 45m$ area as depicted in Figure 3.3. The three buildings introduce multiple reflections in addition to the edge-diffraction effect, thus resulting in many more propagation paths depending on source-receiver configuration. This model was built to increase the complexity of the previous model. This model is also used for initial investigation on the beamformers.

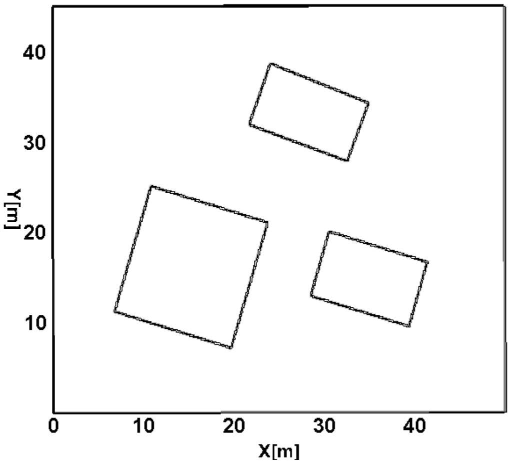


Figure 3.3: 2D model with three obstructing buildings.

3.3.2 3D Models

Several 3D models are introduced here. The previous single and three-building models are extended to 3D. Then, two more realistic urban models are discussed: simplified urban model and Army Fort Benning model.

a. Single building model:

Figure 3.4 shows the 3D version of the previous 2D single building model. The building has the same size with 3*m* height in a 50*m* × 45*m* × 16*m* volume. The main difference of this model as compared to the corresponding 2D in Figure 3.2 is that there is an additional propagation path over the top of the object. Due to its small height, this new propagation path is shorter and likely more dominant that the two side propagation paths.

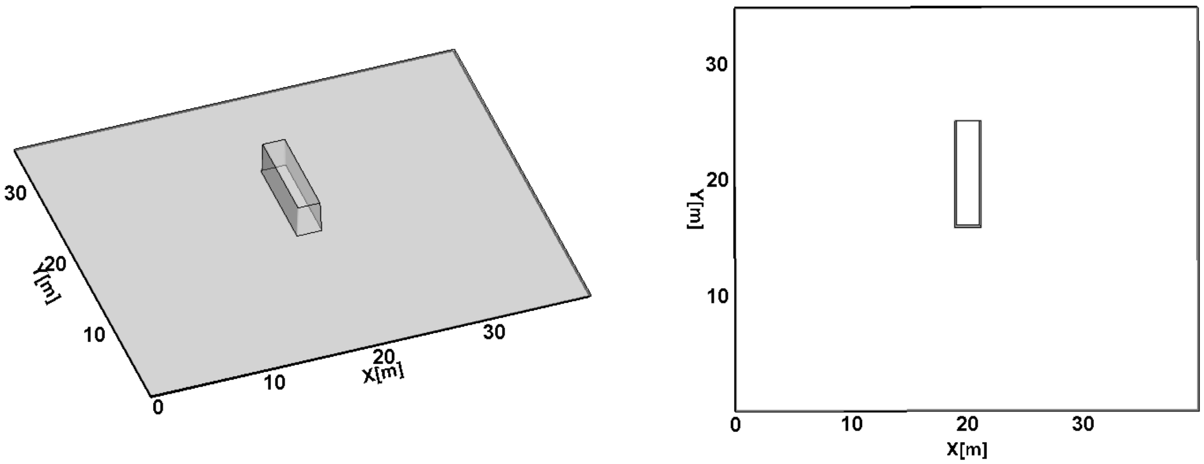


Figure 3.4: 3D model with single obstructing building: 3D view (left) and top view (right).

b. Three-building model:

Similarly, Figure 3.5 shows the 3D version of the 2D three-building model. The same three buildings are introduced with an identical height (3*m*) in a 50*m* × 45*m* × 20*m* volume. In this model, numerous propagation paths exist between the source and the receiver (depending on their locations).

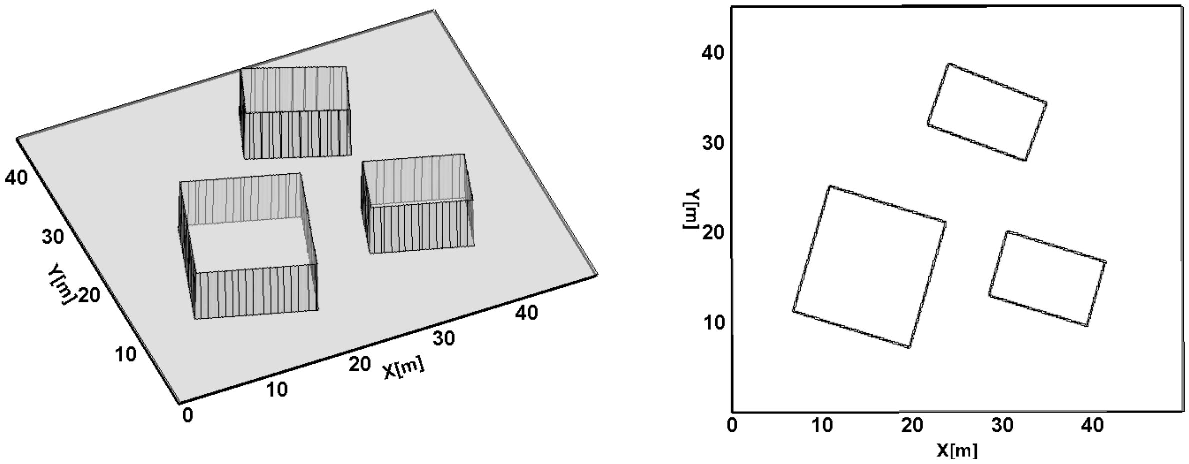


Figure 3.5: 3D model with three obstructing buildings: 3D view (left) and top view (right).

c. Large 3D Models:

Here, two large complex urban models are introduced for more realistic scenarios. The first one resembles the Army Fort Benning training facility (Dale et al., 2002) without any recognizable urban canyons as shown in Figure 3.6. In this environment, there are 15 buildings with different size, shape, and height in a 130*m* × 80*m* × 14*m* volume. This model is considered as the most realistic one.

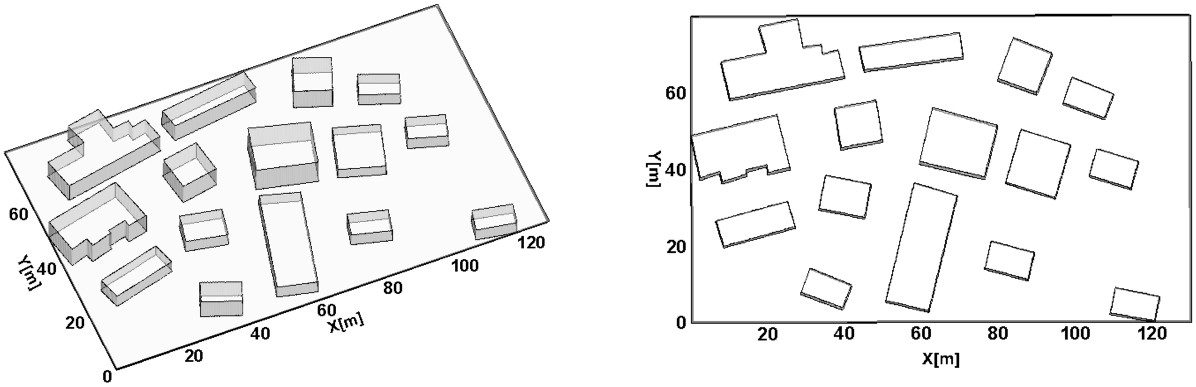


Figure 3.6: 3D Army Fort Benning model: 3D view (left) and top view (right).

The second model is a simplified urban environment consisting of regularly spaced buildings forming urban canyons as shown in Figure 3.7. It consists of 16 buildings with the same size, shape, and height ($30m \times 15m \times 7m$) in a $150m \times 100m \times 10m$ volume. Though the Fort Benning model is a more realistic case, it is not very suitable to perform fundamental localization studies due to its complexity where the localization performance can be biased by irregular clutters in the environment. On the other hand, the simplified urban model allows more efficient localization performance study due to its regularly spaced streets and regular clutter distribution in the environment.

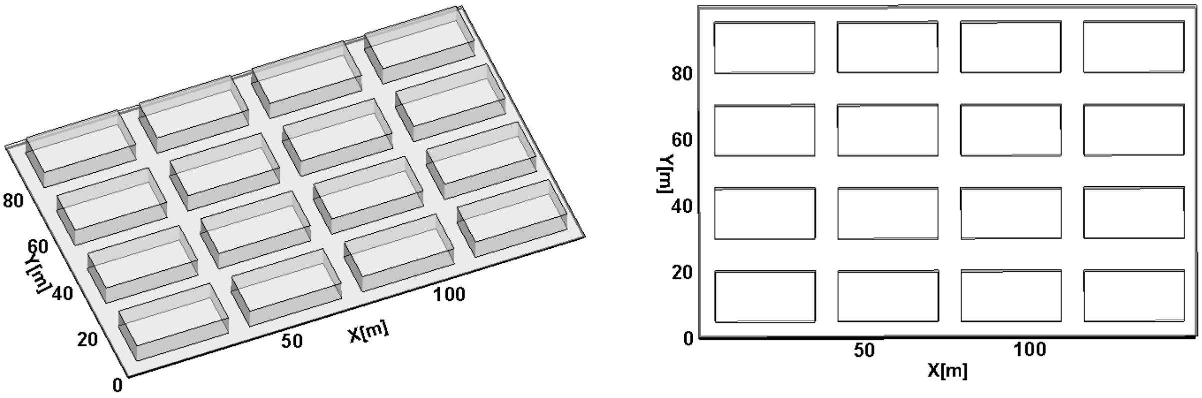


Figure 3.7: 3D simplified urban model: 3D view (left) and top view (right).

3.3.3 Illustrative Results

Here, illustrative results using both the FDTD and the energy method are presented to show the outputs of both modeling tools in a given model. The simulations presented here are limited to the large 3D models.

a. FDTD Simulations:

Results for the 3D FDTD simulation on the Army Fort Benning model are presented in Figure 3.8. The monopole noise source is located at the coordinate (20, 10, 2) m and it radiates a Ricker pulse (Robinson and Treitel, 1980) with a central frequency, f_c ,of 100 Hz. The mathematical formula for a Ricker pulse is

$$s(t) = (1 - 2\pi^2 f_c^2 t^2) e^{-\pi^2 f_c^2 t^2} \quad (3.13)$$

Due to the numerical stability of the FDTD method, the maximum frequency that can be used to obtain accurate results without distortion is 274 Hz for the grid resolution of 0.125 *m* implemented (Choi et al., 2011a). Figure 3.8a shows the Ricker pulse normalized to unit amplitude. It is shown that the waveform decays rapidly in time domain. The dominant part of the spectrum of this waveform concentrates around the center frequency as shown in Figure 3.8b. The source information is mostly found in the 50 to 250 Hz frequency range with virtually no content above 250 Hz which is below the maximum frequency allowed in the simulations. In addition to the limitation imposed by the grid resolution, to ensure numerical stability of the results, the time step in the FDTD calculation must be chosen to avoid temporal aliasing of the signals. Thus, a time step of 0.2 *ms* is used and the simulations are run for a total of 800 *ms* to allow the pulse excitation to fully propagate through the environment.

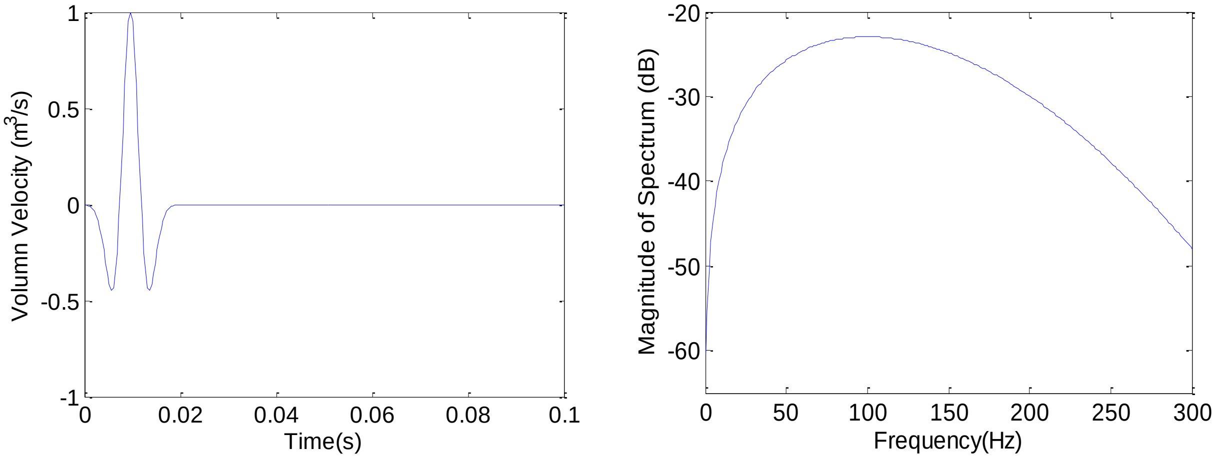


Figure 3.8: (a) Time history and (b) magnitude of spectrum of the Ricker waveform with center frequency of 100Hz.

Snapshots of the instantaneous pressure at 0.0378, 0.0498, and 0.0698 seconds into the simulation are shown in Figure 3.9. The figure shows the top view of the pressure map at a height of 2 m. A color scheme is used to indicate the instantaneous acoustic pressure, i.e. from -3 (blue) to 3 (red) *Pa*. It can be seen that the pressure generated by the source propagates through the environment as time goes by revealing reflections and diffractions from buildings and absorption from the boundaries of the model.

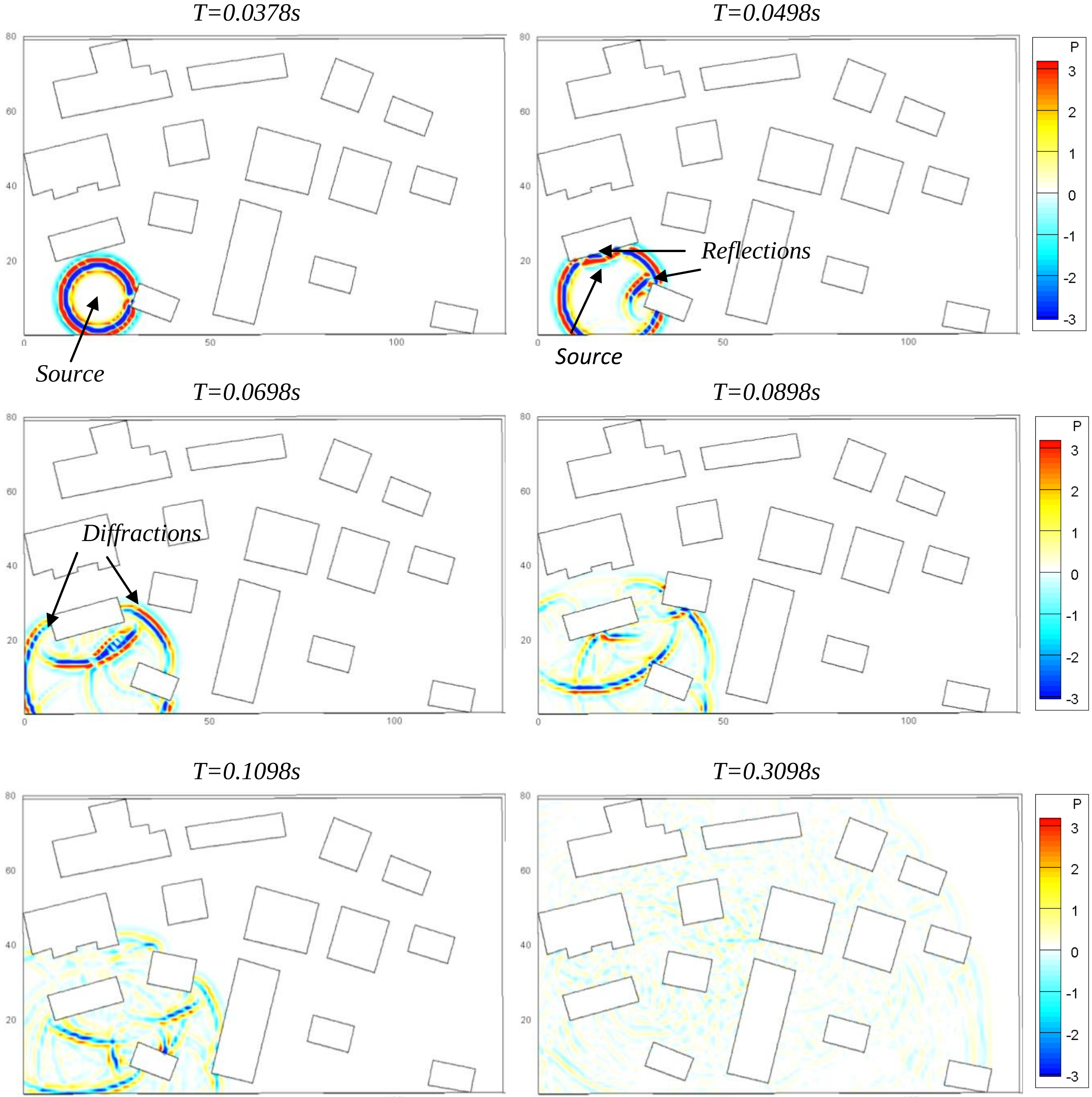


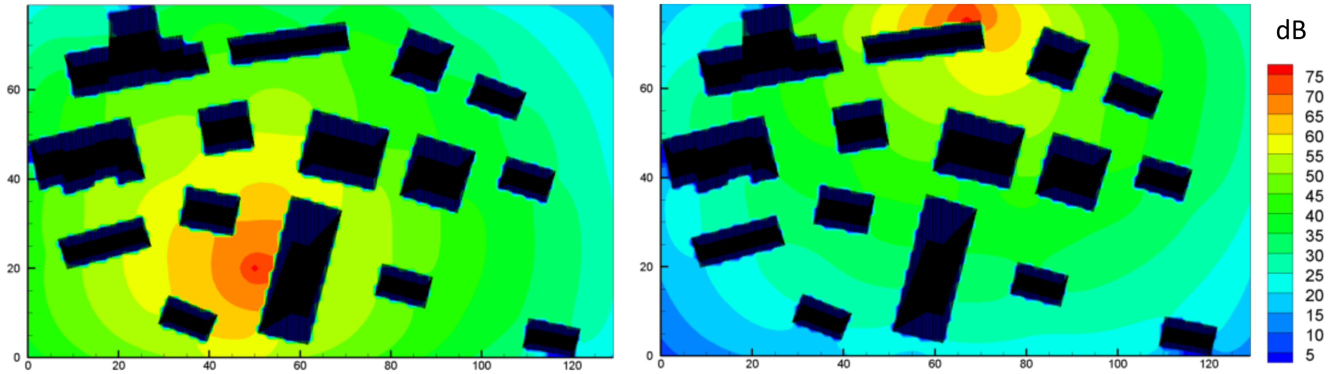
Figure 3.9: Instantaneous acoustic pressure maps (top-view) by 3D FDTD method for Army Fort Benning model. Source is at (20, 10, 2) *m*.

b. Energy method

Similarly, the 3D energy code was run on the Army Fort Benning training facility model shown in Figure 3.7. Similar to the simulations before, the buildings and the ground are assumed to be perfectly reflective. Perfect absorption was assumed at the boundaries of the domain. For simulation results, the diffusion coefficient was chosen constant through the environment. It was calculated based on the average of the streets' width. In the numerical simulations, the source sound power was assumed to be a unit step input. Figure 3.10 shows the steady-state sound pressure level (SPL) maps (2m above the ground) produced by a source at different locations in the Army Fort Benning model. Note that the results here are valid for any frequency band since no parameter of the energy-based method depends on the frequency. The results are presented in decibels on a scale from 5 to 75 dB.

Source is at (50,20,2) m

Source is at (67,75,2) m



Source is at (85,10,2) m

Source is at (5,72,2) m

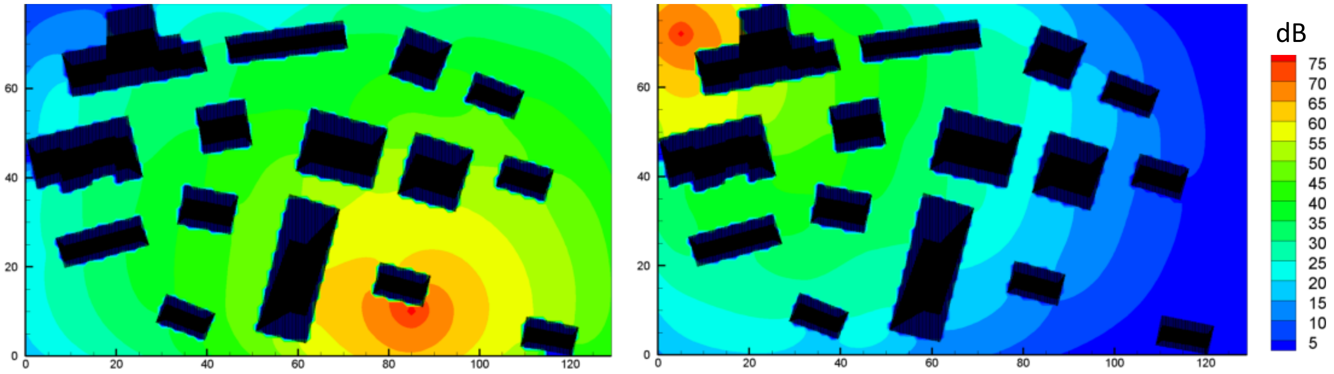


Figure 3.10: SPL (dB) map for the Army Fort Benning model due to different source locations using energy method for 1/3 octave band with center frequency 63Hz .

3.4 Accuracy of the Energy-based Method

The development of an energy-based approach for modeling urban environment is beyond the scope of this dissertation. The energy model is used here as a tool in combination with a source localization technique described in chapter 6. It is assumed that as this tool is further improved, the localization approach will also yield better results. In this section, results are presented for the only purpose of illustrating the current level of accuracy of the energy method.

The energy-based method was originally developed for the modeling of stationary sound propagation with diffusely reflecting boundaries for medium and high frequency bands. The accuracy of the energy-based method was studied by comparing to the exact solution using FDTD (Pasareanu et al., 2010). It should be noted that the FDTD model was used at low frequency with buildings producing specular reflections. To this end, the transient response from the FDTD code had to be converted to steady-state responses to be consistent with the energy code output. This conversion is not directly relevant to this dissertation but it can be found in the work by Pasareanu. To begin with, the results from the both methods are obtained assuming a constant source acoustic power at all frequency bands. For the FDTD results, diffraction effects are frequency dependant and thus the sound distribution in the environment differs for each frequency band. On the other hand, in the energy method the results are the same for all bands.

For comparison, Figure 3.11 presents the SPL maps (2m above the ground) using both methods for the Army Fort Benning model. The source is located at (20, 10, 2) *m*. The results obtained from the energy method and the FDTD approach (data in 1/3 octave band) for the 63Hz, 100Hz and/or 200Hz bands for the Fort Benning model are shown on the left and right column, respectively. Results have shown that the energy method models the steady-state sound distribution in a relatively similar pattern as the FDTD. The agreement between both methods is good in the vicinity of the source and a bit further away at higher frequency, e.g. differences are below ~6 dB up to a distance of 60 m from the source. This can be explained by the increased clutter in the environment, meaning higher reverberation of the sound field better matches the assumption of diffuse field. However, it is also noticed that the errors tend to increases as the observer moves far away from the source. This is because the levels tend to decrease faster in the energy method as the sound move through the clutter.

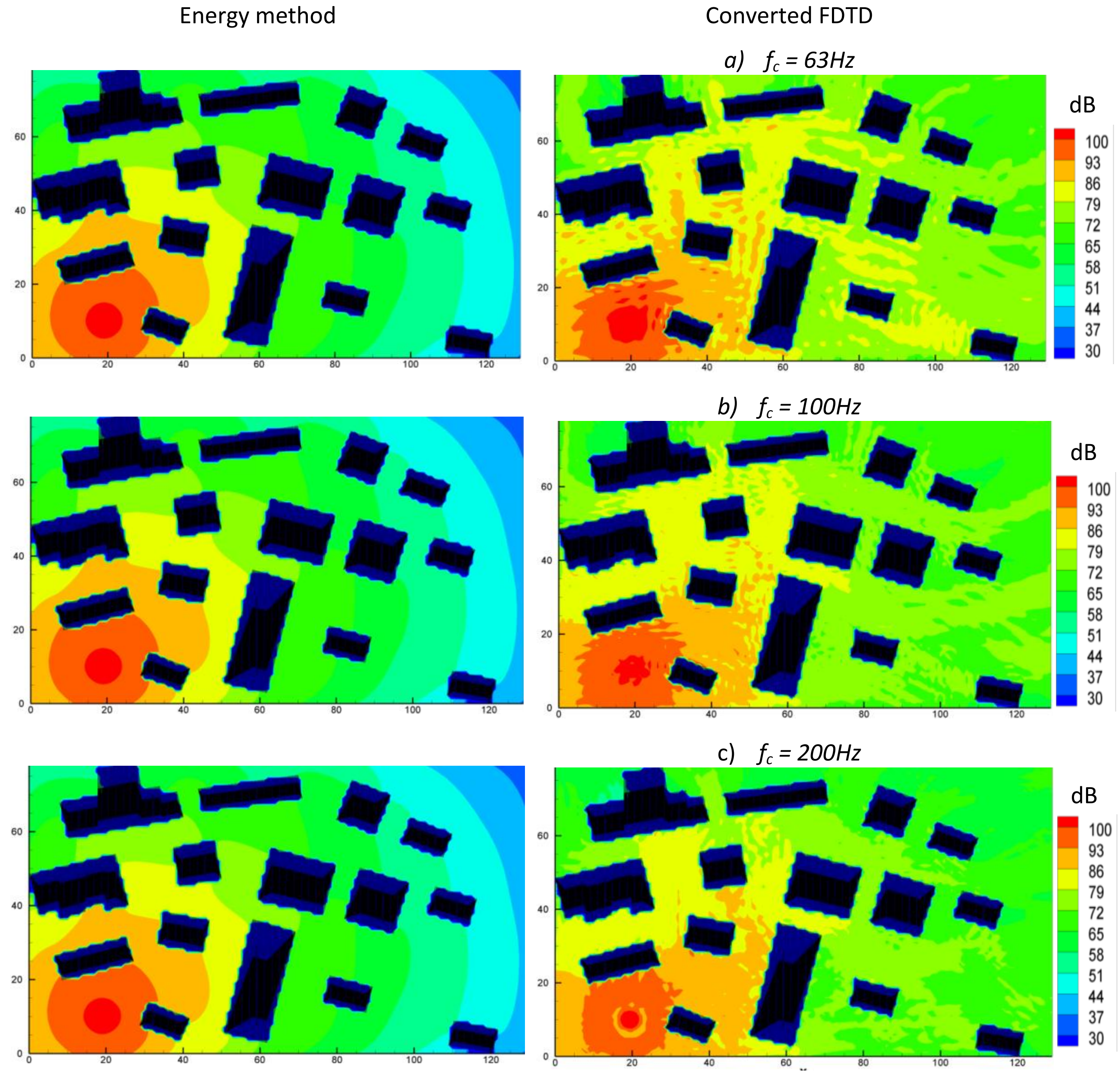
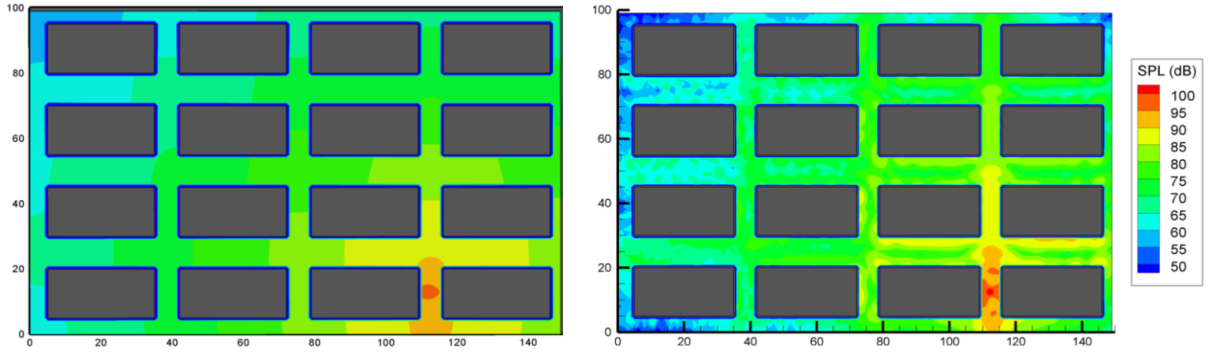


Figure 3.11: SPL (dB) distribution 2 m above the ground for the Fort Benning model using energy method and FDTD method. Comparisons for 1/3 octave bands with center frequency a) 63Hz, b) 100Hz and d) 200Hz. Source is at (20, 10, 2) *m*.

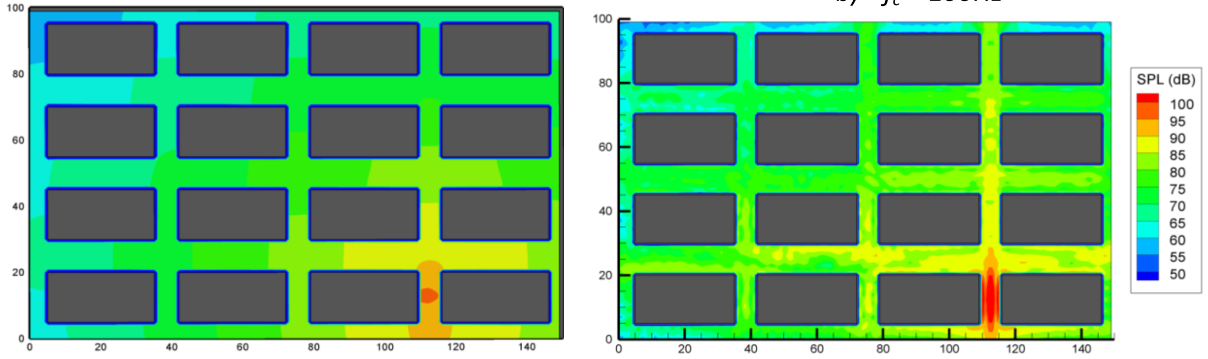
The same type of results for the simplified urban model is presented in Figure 3.12. The source is at (112.5, 10, 2) *m*. The results have shown that the energy method models the steady-state sound distribution in more similar pattern as the FDTD for the simplified urban model. It is because the simplified urban model consists of regularly spaced buildings and streets, which is more suitable for the energy method. However, it is noticed that in the vicinity of the source the levels computed by the energy method are a bit lower than the FDTD results at all frequency bands.

Energy method Converted FDTD

a) $f_c = 63Hz$



b) $f_c = 100Hz$



c) $f_c = 200Hz$

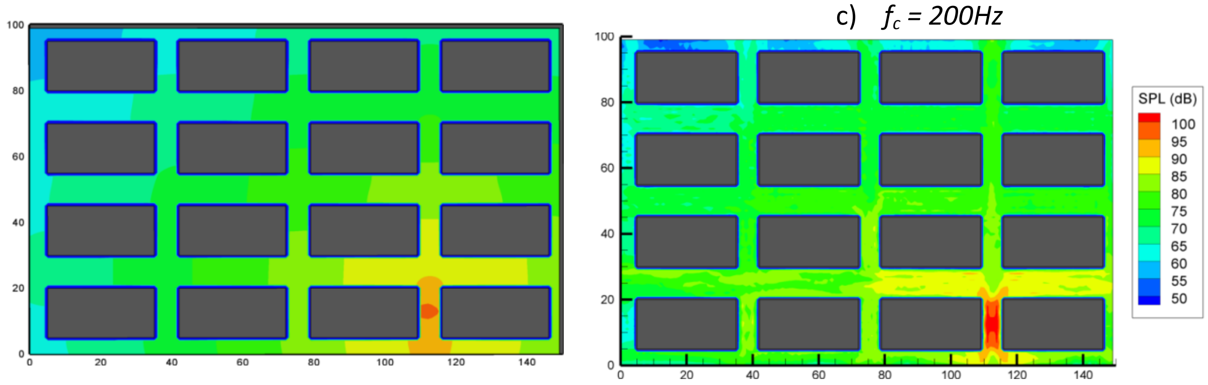


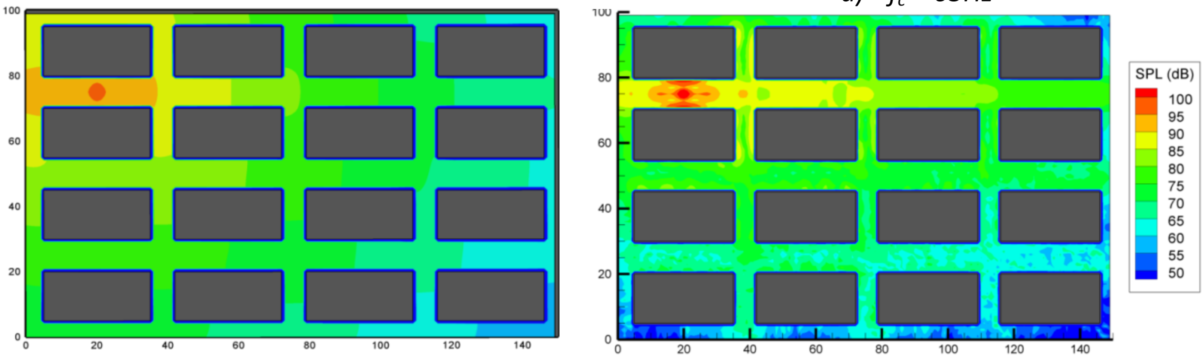
Figure 3.12: SPL (dB) distribution 2 m above the ground for the simplified urban model using energy method and FDTD method. Comparisons for 1/3 octave bands with center frequency a) 63Hz, b) 100Hz and d) 200Hz. Source is at (112.5, 10, 2) *m*.

Similar results are observed for different source position i.e. the source is at (20, 75, 2) *m* in the simplified urban environment as shown in Figure 3.13. The relatively similar pattern between both methods can be found, as expected. The result for 200Hz band, for which there is more frequency lines, shows the best match among them.

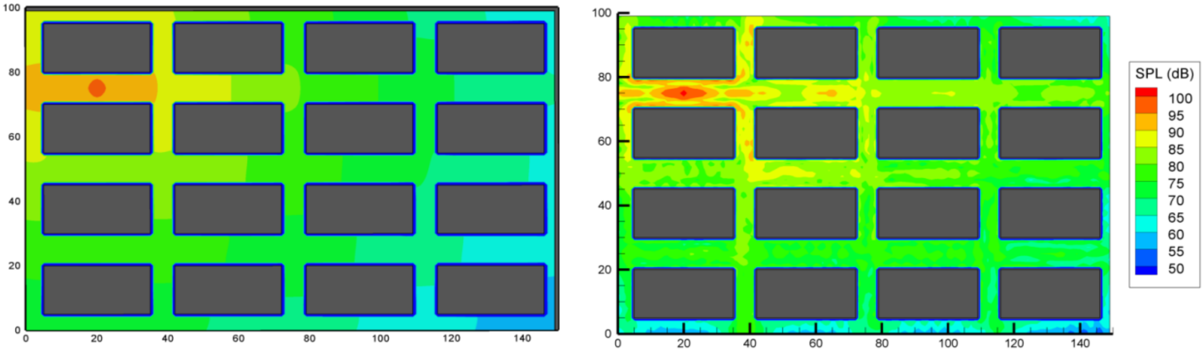
Energy method

Converted FDTD

a) f_c = 63Hz



b) f_c = 100Hz



c) f_c = 200Hz

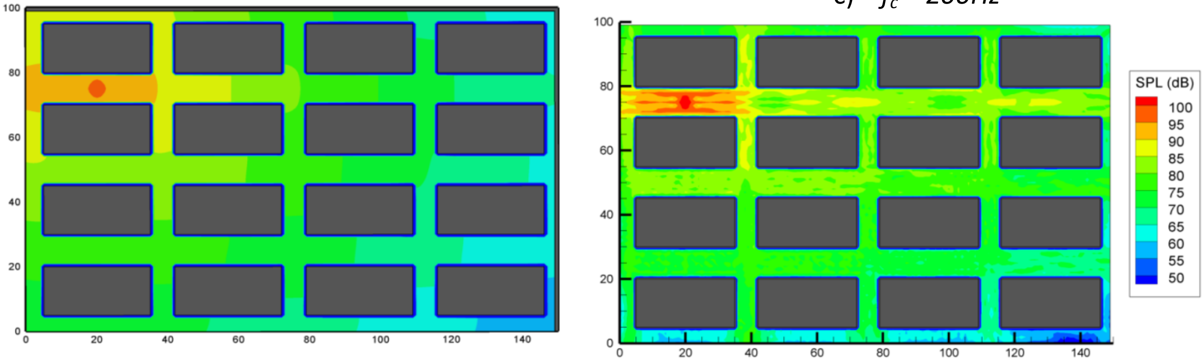


Figure 3.13: SPL (dB) distribution 2 m above the ground for the simplified urban model using energy method and FDTD method. Comparisons for 1/3 octave bands with center frequency a) 63Hz, b) 100Hz and d) 200Hz. Source is at (20, 75, 2) *m*.

3.5 Computational Time

The FDTD method is a simple high-fidelity modeling tool but computationally inefficient. On the other hand, energy method allows fast computation of the sound field, providing a low-fidelity model. To compare the computational time of both methods, Figure 3.14 shows the run time required to simulate the sound field created by a single noise source for the numerical model considered earlier. Note that the computational time of the energy method is per frequency band and here only 3D models are considered. As it can be observed and expected, the energy code runs much faster than the FDTD code. For instance, the computational time for the Army Fort Benning model using the energy-based method is only 30 seconds for a single noise source while the FDTD code takes approximately 24 hours. In overall, the energy method runs around 1000 times faster than the FDTD code using the same machine under the same environment.

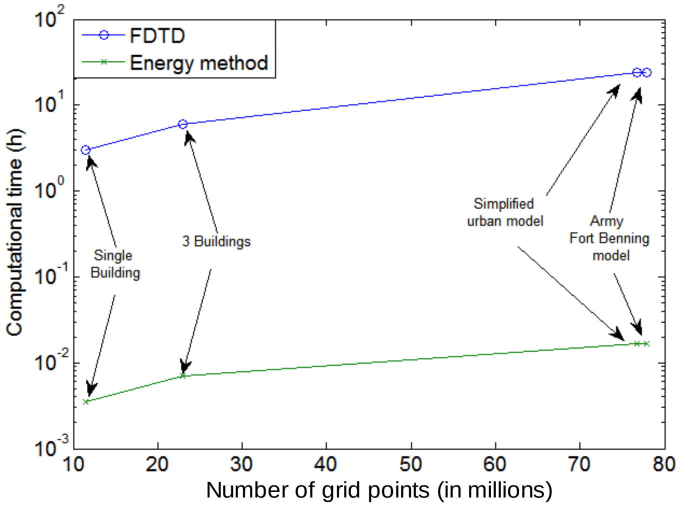


Figure 3.14: Computational time for the sound field created by a single noise source in the numerical models using both FDTD method and Energy method

Finally, it can be concluded that the energy-based method has significant potential for efficiently predicting the SPLs in urban environments, in particular compared to the FDTD method. Firstly, the energy-based method is relatively accurate to model the sound field in urban environments, though it still needs improvements. Secondly, the energy-based method allows much faster computation of the sound field, an important requirement for source localization applications. Therefore, the energy-based method is used in conjunction with the fingerprinting for source localization, which will be discussed in more detail in Chapter 6.

Chapter 4 Localization technique: Beamforming

In this Chapter, the source localization technique using a microphone phased array is introduced. The source localization system for an impulsive sound source in complex urban environments is described in section 4.2. Beamforming in conjunction with MFP is then presented as an effective localization techniques using mathematical formulation. Several beamformers are investigated as candidates. They are the conventional delay-and-sum (DS), conventional Minimum Variance Distortionless Response (MVDR), the beamformer proposed by Mennitt and Johnson (2010) referred here as Mennitt-MVDR, and enhanced MVDR (EMVDR) beamformers. The robustness of the beamformers to uncorrelated noise is also investigated. In the end, an approach to estimate the source location is described and used for localization performance analysis.

4.1 Beamforming Theory

Beamforming is a signal processing technique used in conjunction with a microphone array to provide a versatile form of spatial filtering (Veen et al., 1988). It enhances the signal from the desired spatial direction while reducing the signal from other directions. A beamformer combines the spatially distributed microphone array data linearly with the beamforming weight to achieve spatial filtering. The main idea behind the beamforming weight is to electronically steer a beamformer in all possible locations and look for the maximum beamforming output.

Typically, the simplest beamformer assumes a “free-field” model of sound propagation to steer the array to a particular location in the environment referred to as a grid point. In other words, it ignores the actual environment. In the presence of clutter, however, the free-field assumption breaks down, as the propagating waves are affected by objects in the environment, e.g. reflections, scattering, etc. To overcome this problem, an accurate model of the actual environment must be used. Note that beamforming in a cluttered environment is challenging but also provides an opportunity. The clutter adds unique characteristics into the sound field due to diffraction effects, multi-path propagation and so forth. These characteristics can be used to distinguish multiple sources located at the same bearing relative to the array.

Beamformers that used an estimate of the actual environment to steer the array are known as MFP algorithms (Baggeroer et al., 1993). This MFP has been largely developed in ocean acoustic for source localization and tomography (Fialkowski et al., 2000). In contrast to the typical underwater application, there is no closed form solution for the array manifold in scattering environments due to the complex nature of propagation. The array manifold can be obtained by measurement or numerical modeling. The manifold is relatively easy to calculate if an accurate model of the environment is used.

4.2 Source Localization System

Figure 4.1 depicts the proposed localization system for an impulsive sound source in a complex urban environment using a microphone phased array. Figure 4.1a represents an example of a real scenario where an acoustic event occurs at unknown position in the environment (solid circle). The sound is measured by a microphone phased array deployed in the environment (dotted circle) such that it does not have a direct LoS to the source. In the work here, the phased array is assumed to have six microphones (M=6) arranged on an open sphere as shown in Figure 4.1b. For MFP beamforming, the transfer function between each candidate source location in the environment and the microphones in the array must be either measured or estimated (Figure 4.1d). It is, however, practically impossible to measure these transfer functions for a large environment. To this end, numerical techniques are used to model the acoustic environment and provide the transfer functions (Figure 4.1c). In MFP, these transfer functions are the steering vectors in the beamforming process (Figure 4.1e). From the microphone signal in the array the cross spectral matrix (CSM) is calculated for all frequencies within a selected frequency band (Figure 4.1f). Finally, the beamforming output is computed at each spectral line, i.e. narrowband processing, and added as uncorrelated signals to yield the localization map (Figure 4.1g). The source is then localized using this map.

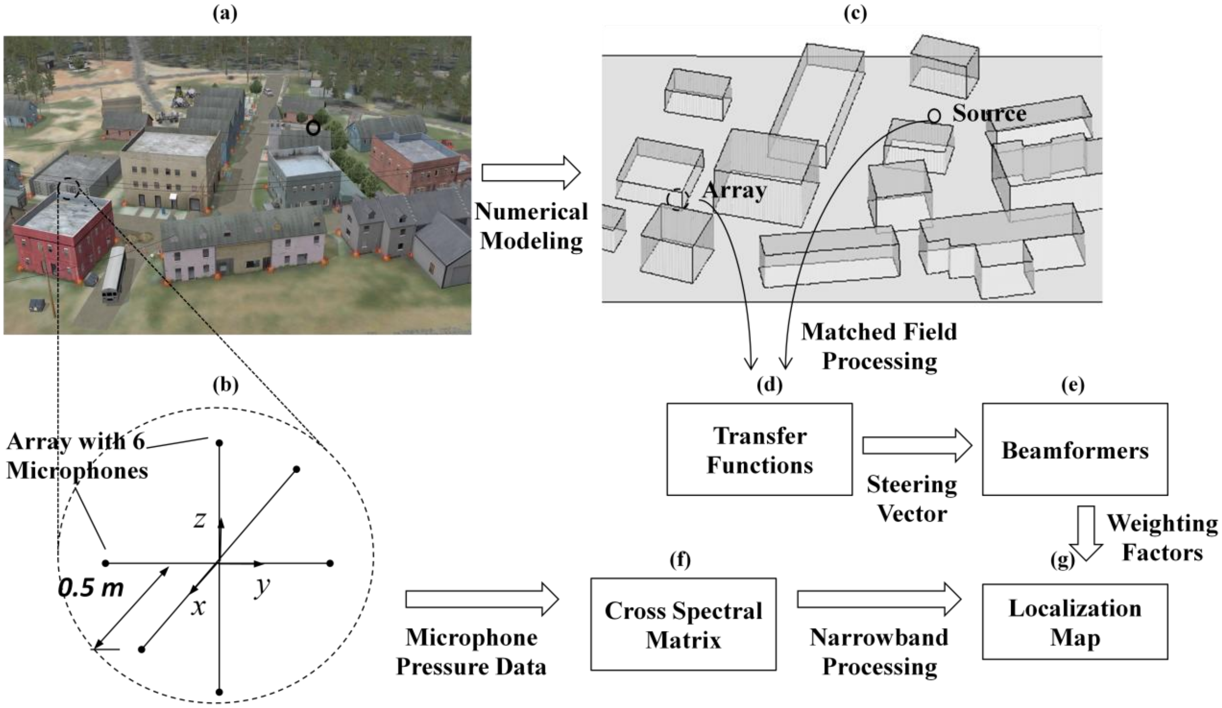


Figure 4.1: Schematic of localization system using beamforming.

4.3 Mathematical Formulation

This section presents how the proposed source localization is formulated in general. Specifics of the beamformers are addressed in the following section. This section also presents how a localization map is obtained. A description of the approach taken for multi-array processing is presented. In this work, the FDTD code is used to generate the microphone signals due to the source. The pressure signal recorded by the m^{th} microphone is,

$$r_m(t) = p_m(t) + v_m(t) \qquad m = 1, 2, \cdots, M \tag{4.1}$$

where

$p_m(t)$ is the pressure at the microphone due to the source from the FDTD code output,

$v_m(t)$ is contaminating noise in the microphone signal assumed to be spatially uncorrelated white noise, and,

M is the number of microphones in the array, e.g. 6.

In general, there are several sources for noise in the microphones. Typical sources are microphone self-noise and wind induced noise. The wind noise is probably the one of most concern and particularly dominant at low frequency.

Since the uncorrelated white noise is uniform in the environment, signal-to-noise ratio (SNR) for each microphone is potentially different. Due to the close proximity of the microphones in an array, the signals from the microphones have, however, approximately the same SNR. The SNR for different arrays will vary more significantly. The SNR of the microphone signal is typically defined as the ratio of the average mean-square-value of the signal to the noise. That is

$$SNR = 10Log_{10}\left(\frac{\frac{1}{W}\sum_{w=1}^{W}\left(p_m(t_w) - \frac{1}{W}\sum_{w=1}^{W} p_m(t_w) \right)^2}{\frac{1}{W}\sum_{w=1}^{W}\left(v_m(t_w) - \frac{1}{W}\sum_{w=1}^{W} v_m(t_w) \right)^2} \right) \tag{4.2}$$

where $T = W\Delta t$ is the time window length, Δt is the sampling interval, and W is the number of samples in the time window. Note that for an impulsive source, the SNR as defined in eq. (4.2) will depend on the time window length T used in the simulations. To illustrate this point, Figure 4.2 shows the time history of the impulsive signal recorded by the microphone and the white noise signal added. It is straightforward that lower SNR will be obtained as the length of the time window increases. To resolve this issue, an unambiguous SNR can be defined using the ratio of the peak value of the impulsive signal (p_{peak}) to the peak value of the noise (v_{peak}) as illustrated in Figure 4.2. That is,

$$SNR_p = 20Log_{10}\left(\frac{p_{peak}}{v_{peak}}\right) \tag{4.3}$$

While the peak value for the impulsive signal is obtained from the simulations, the peak value for the white noise is computed using the crest factor ≅ 4, e.g. v_{peak} is approximately 4 times the root-mean-square value of the noise (Ott, 1976). Due to the stochastic characteristics, data sets of the white noise are generated multiple times for the simulations. As an example, for the case in Figure 4.2, SNR and SNR_p are 10 and 21dB, respectively. In the work here, the SNR in terms of the peak values are considered to be more representative and used in the numerical simulations.

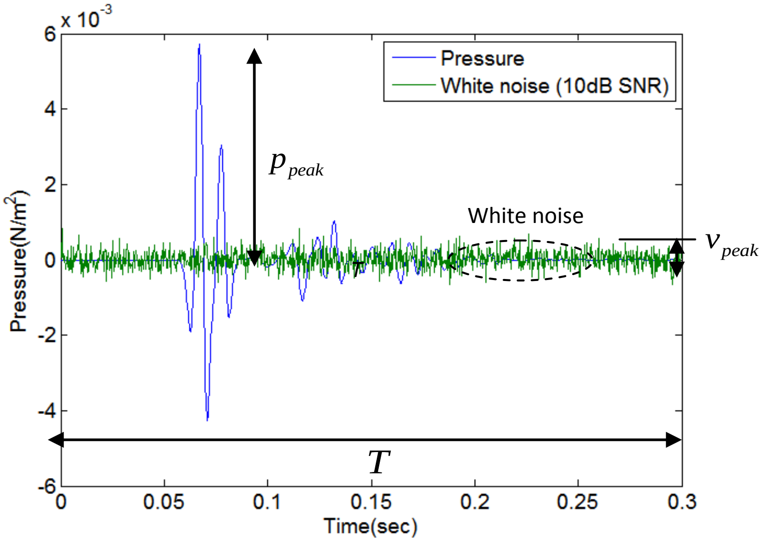


Figure 4.2: Example of the pressure data and the additive white noise (10dB SNR).

Since all the beamformers here are implemented in the frequency domain, the Fourier Transform (FT) of the microphone signals are computed resulting in

$$r_m(\omega) = p_m(\omega) + v_m(\omega) \qquad m = 1, 2, \cdots, M \tag{4.4}$$

where ω is the angular frequency.

As mentioned, the phased arrays are assumed to have six microphones (M=6) on an open sphere of 0.5m radius. For convenience, pairs of microphones were aligned with the Cartesian coordinate system, i.e. two for each x, y, and z coordinate (one for positive direction, another for negative direction) as shown in Figure 4.3. It is noteworthy that the microphones are located exactly at a grid point in the FDTD code so that the ratio of the size of array to the grid spacing should be always an integer. Similarly, the source location should match the grid of the FDTD code.

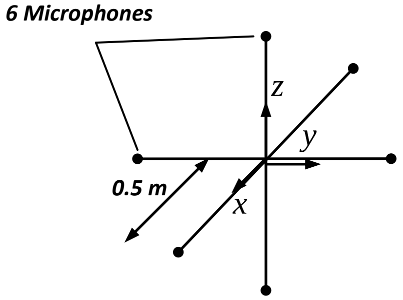


Figure 4.3: Phased array used in the simulations.

To implement the beamformers, a grid of potential source locations must be used, i.e. $j = 1, 2, \cdots, J$. All beamformer compute a weighted average of the microphone signals for each grid point. By definition, the beamformer output is the mean-square-value (msv) of this weighted average. The beamformer output, b , at frequency ω for the j^{th} grid point is then commonly expressed in matrix form as

$$b_j(\omega) = \boldsymbol{w}_j^H \boldsymbol{\Phi}_{RR} \boldsymbol{w}_j \tag{4.5}$$

where $\boldsymbol{w}_j$ is the weighting vector for the j^{th} grid point

$$\boldsymbol{w}_j = \left\{ w_{1j}(\omega), w_{2j}(\omega), \cdots, w_{mj}(\omega), \cdots, w_{Mj}(\omega) \right\}^T \tag{4.6}$$

and $\boldsymbol{\Phi}_{RR}$ is the cross spectral matrix (CSM) of the microphone signals, i.e. diagonal terms are the auto spectra while the off-diagonal are the cross-spectra, given as

$$\boldsymbol{\Phi}_{RR}(\omega) = \boldsymbol{r}\,\boldsymbol{r}^H \tag{4.7}$$

where

$$\boldsymbol{r} = \left\{ r_1(\omega), r_2(\omega), \cdots, r_m(\omega), \cdots, r_M(\omega) \right\}^T \tag{4.8}$$

is the vector of the FT of the microphone signals. Note that in the above equations the superscripts *H* and *T* indicate the Hermitian and transpose, respectively.

It is important to note that beamforming assumes the signals are stationary and that multiple time windows (or snapshots) are taken and the auto and cross spectrum of the microphone signals averaged. However, in the current application a single time window is taken due to the impulsive nature of the source signal considered in this research. Thus, the cross-spectral matrix is computed from a single snapshot (no averages).

The MFP beamformer implies that the steering vectors (manifold) used in the beamformer must represent the actual acoustic environment. That is, they are

$$d_{mj}(\omega) = A_{mj} e^{i\omega\varphi_{mj}} \qquad m = 1, 2, \cdots, M\,;\; j = 1, 2, \cdots, J \tag{4.9}$$

where A_{mj} and φ_{mj} are the magnitude and phase of the transfer function between the m^{th} microphone and the source at the j^{th} grid point computed using the FDTD acoustic propagation model. The transfer function is obtained by taking the FT of the input source and received signals at each microphone in the array assuming a source at the j^{th} grid point. Thus, the normalized j^{th} steering vector, g_j, becomes

$$\boldsymbol{g}_j(\omega) = \frac{\boldsymbol{d}_j(\omega)}{\left|\boldsymbol{d}_j(\omega)\right|} \tag{4.10}$$

with

$$\boldsymbol{d}_j = \left\{ d_{1j}(\omega), d_{2j}(\omega), \cdots, d_{mj}(\omega), \cdots, d_{Mj}(\omega) \right\}^T \tag{4.11}$$

This normalized steering vector simply defines the relative phase and magnitude between the microphones to the source. Note that beamforming is performed in the frequency domain and thus the beamformer output at multiple spectral lines should be combined to generate a single map. To this end, the beamforming output is computed at each k^{th} spectral line as illustrated in Figure 4.4a, i.e. narrowband beamforming. In fact, the processing of the data in narrow bands within the considered frequency band resulted in the best results among other methods investigated such as using a single band or several sub-bands for all scenarios and beamformers (Choi et al., 2010). The reason is that the urban acoustic environment in the presence of clutter is highly dependent on the frequency, thus precluding the possibility of collapsing spectral lines into wider frequency bands.

The beamforming outputs for all frequencies (from the lower frequency ω_ℓ to the upper frequency ω_u) are then simply added as (Figure 4.4b),

$$b_j = \sum_{k=\ell}^{k=u} \boldsymbol{w}_j^H \left(\omega_k\right) \boldsymbol{\Phi}_{RR} \left(\omega_k\right) \boldsymbol{w}_j \left(\omega_k\right) \tag{4.12}$$

where ω_k is the angular frequency of the k^{th} spectral line.

Finally, the resulting beamforming output is normalized to the maximum level in the environment, i.e. $0 \le b_j \le 1$ (Figure 4.4c). The source location estimation will be in the region of maximum beamforming output as illustrated in Figure 4.4.

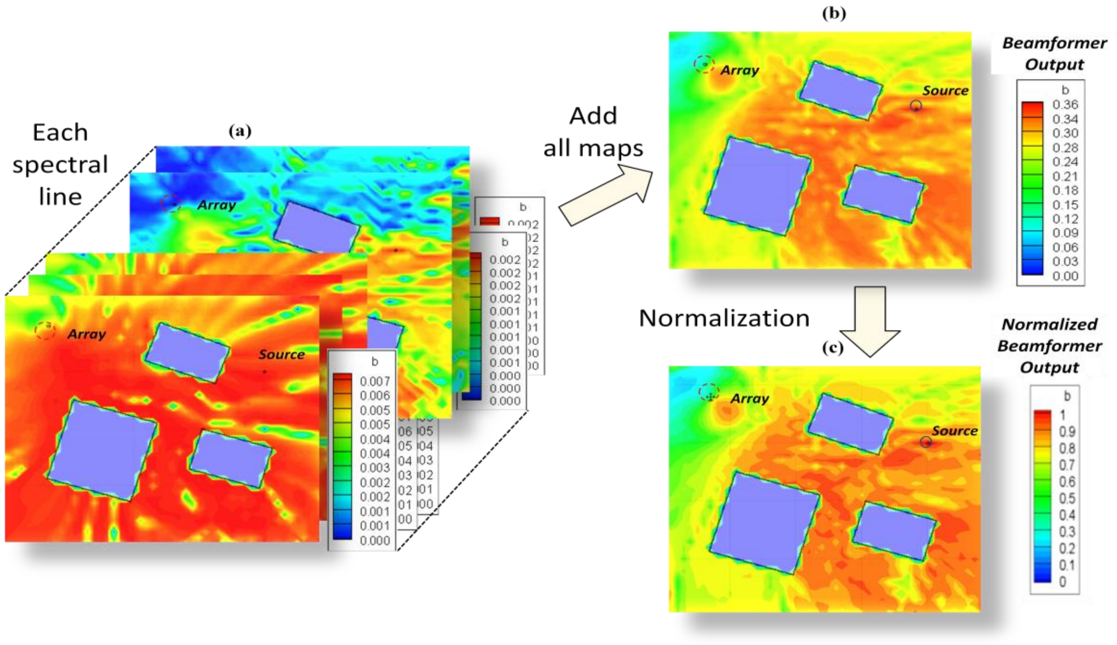


Figure 4.4: Example of narrowband processing to obtain a single map: (a) beamforming output at each spectral line ω_k (b) combined beamforming output for all frequencies from the lower frequency ω_ℓ to the upper frequency ω_u and (c) normalization of the combined beamforming outputs to obtain the localization map.

For the case where multiple microphone arrays are deployed in the environment, the processing described above is applied to each array independently. Then, the array outputs are combined assuming they are uncorrelated prior to normalization, i.e. no phase information is shared. The main reason for this processing is that the noise signals propagating over a long distance through a cluttered environment will eventually become uncorrelated. . Assuming there are N_A arrays, the array outputs are combined as

$$b_j = \frac{1}{N_A} \sum_{q=1}^{N_A} b_{qj} \tag{4.13}$$

where b_{qj} is the q^{th} array beamforming output (un-normalized) for the j^{th} grid point. The beamforming map is once again normalized to the maximum level in the environment. Figure 4.5 shows an example of two-array processing.

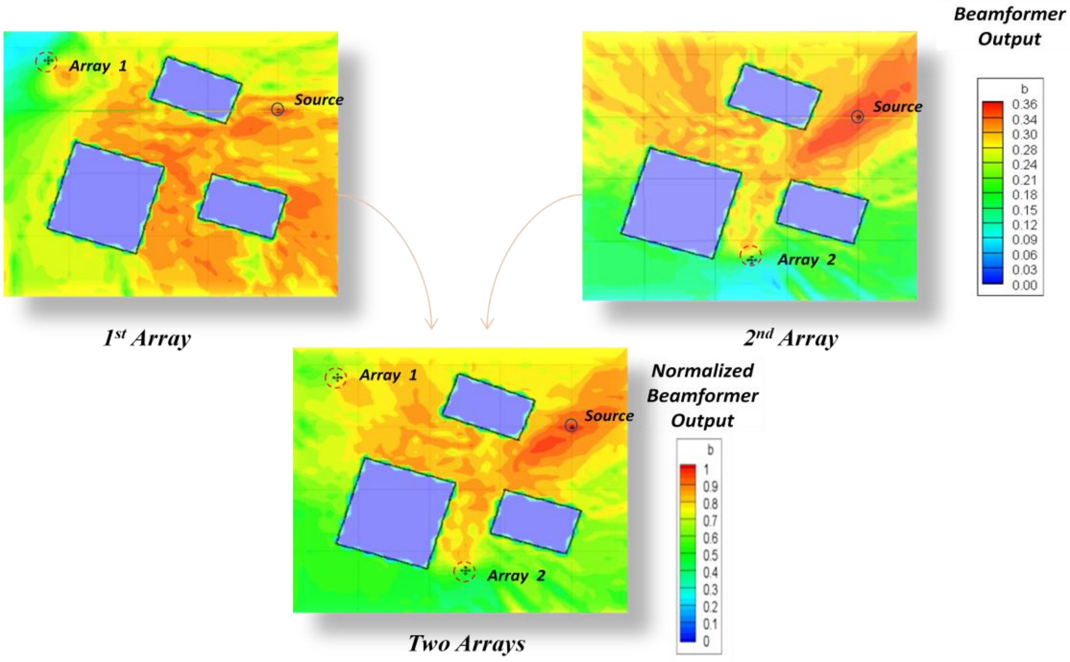


Figure 4.5: Example of two-array processing to obtain a single map The source is at (40, 30, 2) and the array 1 and 2 are at (7, 42, 2) and (20, 10, 2), respectively.

4.4 Beamformers

The beamformers investigated in this work are described here. They are the conventional delay and sum (DS), conventional Minimum Variance Distortionless Response (MVDR) (Nuttall et al., 1969), Mennitt-MVDR (Mennitt, 2008) and the enhanced MVDR (Abraham and Owsley, 1990). These beamformers are investigated to identify the most efficient beamforming method to locate sources in noise urban environments.

Due to its importance, the white noise gain (*WNG*) is used as a measure for assessing robustness of the beamformers to uncorrelated noise. The *WNG* is defined as the ability of the microphone array to suppress spatially distributed noise, e.g., sensor noise of the microphones (Brandstein and Ward, 1995). The *WNG* (in decibels) indicates the improvement in the SNR due to the beamformer, i.e. the higher the *WNG* the better capabilities of the beamformer to suppress uncorrelated noise. It is important to note that the *WNG* as conventionally defined assumes that the signals are stationary. This implies that signals are averaged over multiple time windows and the cross spectral matrix of the noise is diagonal with diagonal terms equal to the noise power. However, in this work impulsive sound sources are assumed and a single time window is used for all spectral calculations, i.e. cross spectral matrix of the noise is not diagonal. However, the conventional *WNG* is still a good estimate of the capability of the beamformer to handle uncorrelated noise. As a reference, the best *WNG* possible to achieve is $10\log_{10} M$ where M is the number of microphone in the array. The derivation of the *WNG* is presented in Appendix A.1. Note that WNG is computed at each spectral line and the dependence on the frequency is omitted for the sake of brevity.

a. Conventional Delay-and-Sum (DS) Beamformer

The simplest approach to beamforming is conventional DS beamforming. In the conventional DS beamformer, the weighting vector is the normalized steering vector. That is,

$$\boldsymbol{w} = \boldsymbol{g} = \frac{\boldsymbol{d}}{|\boldsymbol{d}|} \tag{4.14}$$

It is robust and provides the best rejection to uncorrelated noise. This is measured by the *WNG* as given as,

$$WNG = 10\log_{10}\frac{\left|\boldsymbol{w}^H\boldsymbol{d}\right|^2}{\boldsymbol{w}^H\boldsymbol{w}} = 10\log_{10}\frac{\left|\boldsymbol{g}^H\boldsymbol{d}\right|^2}{\boldsymbol{g}^H\boldsymbol{g}} = 10\log_{10}\frac{\left|\boldsymbol{d}^H\boldsymbol{d}\right|^2}{\boldsymbol{d}^H\boldsymbol{d}} = 10\log_{10} M \tag{4.15}$$

The DS beamformer is an optimal beamformer to suppress spatially uncorrelated noise.

b. Minimum-Variance-Distortionless-Response (MVDR) Beamformer

In the MVDR beamformer, the output signal is minimized subject to a distortionless constraint that sets the gain in the steering direction to unity. The effect is to steer nulls into the direction of interference while preserving the response in the look direction (Nuttall et al., 1969). In general, the MVDR beamformer gives higher resolution compared to the conventional DS beamformer. In this beamformer, the weighting vector is given by,

$$\boldsymbol{w} = \frac{\boldsymbol{\Phi}_{RR}{}^{-1}\boldsymbol{g}}{\boldsymbol{g}^{H}\boldsymbol{\Phi}_{RR}{}^{-1}\boldsymbol{g}} \tag{4.16}$$

Computing the weight vector in MVDR involves finding the inverse of the CSM matrix, $\boldsymbol{\Phi}_{RR}{}^{-1}$. The straight computation of this inverse could be difficult to obtain because the matrix might be poorly conditioned or singular, i.e. the rank of $\boldsymbol{\Phi}_{RR}$ is smaller than M. In fact, for the single source investigated here the rank of the CSM is unity and the inverse doesn't exit. The traditional remedy for this singularity issue is diagonal loading, i.e. the addition of a small term to the diagonal to make the CSM invertible (Gilbert and Morgan, 1955). Note that all the simulation results in this dissertation using the MVDR beamformer are obtained by the diagonal loading.

Another main issue is that the MVDR beamformer is extremely sensitive to even small errors in the look direction. For instance, when the direction of the signal source does not exactly coincide with any of the steering angles, the output signal tends to get suppressed. In addition, the MVDR beamformer is sensitive to uncorrelated noise (i.e. the *WNG* is small), especially at low frequencies (Gilbert and Morgan, 1955). The *WNG* of the MVDR beamformer is given as

$$WNG = 10\log_{10}\frac{\left|\boldsymbol{w}^{H}\boldsymbol{d}\right|^{2}}{\boldsymbol{w}^{H}\boldsymbol{w}} = 10\log_{10}\frac{\left|\boldsymbol{d}\right|^{2}\left|\boldsymbol{w}^{H}\boldsymbol{g}\right|^{2}}{\boldsymbol{w}^{H}\boldsymbol{w}} = 10\log_{10}\frac{M}{\boldsymbol{w}^{H}\boldsymbol{w}} \tag{4.17}$$

where $\left|\boldsymbol{w}^{H}\boldsymbol{g}\right|^{2} = 1$ because of the MVDR unity response constraint.

c. Mennitt-MVDR Beamformer

Mennitt (2008) recently proposed a beamformer where an artificially generated CSM $\boldsymbol{\Phi}_{VV}$ is used in the MVDR formulation instead of the microphone cross spectral matrix $\boldsymbol{\Phi}_{RR}$. In this beamformer, the weighting vector is given by

$$\boldsymbol{w} = \frac{\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}{\sqrt{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}} \tag{4.18}$$

where

$$\boldsymbol{\Phi}_{VV}(w) = (1-\varepsilon)\boldsymbol{D}\boldsymbol{D}^{H} + \varepsilon\frac{\left\|\boldsymbol{D}\boldsymbol{D}^{H}\right\|}{M}\boldsymbol{I} \tag{4.19}$$

$$\boldsymbol{D}\boldsymbol{D}^{H} = \sum_{j=1}^{J}\boldsymbol{d}_{j}\boldsymbol{d}_{j}^{H} \tag{4.20}$$

and $\varepsilon < 1$ ($\varepsilon = 0.01$ selected for simulations).

As described by Mennitt et al. (2010), the CSM $\boldsymbol{\Phi}_{VV}$ is generated using all the steering vectors from the region been scanned as in eq. (4.19). In other words, this method assumes that at each grid point there is an artificial noise source. He mentioned that this design was to reject correlated noise, resulting in uncorrelated noise amplified at low frequencies. To overcome this, he artificially added a small scalar to the main diagonal of the matrix, which also conditions the matrix for inversion as in eq. (4.18). The factor ε can range from 0 to 1, which is claimed to result in the unconstrained superdirective ($\varepsilon = 0$) or conventional DS beamformer ($\varepsilon = 1$), respectively. Since the noise field is unknown a priori, he assumed that $\varepsilon = 0.01$ can keep conservative balance between sensitivity (i.e. super-directivity characteristics) and the *WNG*.

However, a problem with this approach is that the beamformer results depend on the grid selected, i.e. the results are not independent of the scanned domain. Moreover, this can be computationally burdensome for a large complex environment. The calculation of this matrix should be made prior to the implementation and stored for real-time source localization. Unlike the MVDR beamformer, it is noteworthy that the inverse of $\boldsymbol{\Phi}_{VV}$ exists as long as $J \geq M$ in eq. (4.19) and hence the diagonal loading suggested by Mennitt (2008) is not required as described in eq. (4.19).

As it will be shown in simulations later, another issue is that the Mennitt-MVDR beamformer shows insufficient superdirective properties. Rather, it performed similar to the conventional DS beamformer (see results in Chapter 5) while showing robustness to uncorrelated noise.

In this beamformer, the *WNG* is given as (see Appendix A.2 for mathematical derivation in detail)

$$WNG = 10\log_{10}\frac{\left|\boldsymbol{w}^H\boldsymbol{d}\right|^2}{\boldsymbol{w}^H\boldsymbol{w}} = 10\log_{10}\frac{\left(\boldsymbol{d}^H\left(\boldsymbol{\Phi}_{VV}{}^{-1}\right)^H\boldsymbol{d}\right)^2}{\boldsymbol{d}^H\left(\boldsymbol{\Phi}_{VV}{}^{-1}\right)^H\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}} \tag{4.21}$$

Unfortunately, the *WNG* cannot be explicitly written in terms of the number of microphones *M* because of the unconventional form of the weighting vector selected in eq. (4.18). Thus, direct comparison to the other beamformers through analytical expressions is not feasible.

d. Enhanced MVDR (EMVDR) Beamformer

This beamformer is based on the conventional MVDR but only the dominant eigen-property of the CSM is used (Abraham and Owsley, 1990). This beamformer was examined using a bottom-mounted horizontal line array in a shallow-water environment by Messerschmitt and Gramann (1997). Bao (2002) showed the robustness of this beamformer for passive sonar by resolving the bearings of acoustic sources using experimental data.

The weighting vector in the EMVDR beamformer is given as (see Appendix B.1 for mathematical derivation in detail)

$$\boldsymbol{w} = \frac{\boldsymbol{g} - \beta\,\boldsymbol{m}_1\boldsymbol{m}_1^H\boldsymbol{g}}{1-\beta\,\left|\boldsymbol{m}_1^H\boldsymbol{g}\right|^2} \tag{4.22}$$

with

$$\beta = \frac{e\psi_1}{\sigma^2 + e\psi_1} \tag{4.23}$$

where λ_1 and $\boldsymbol{m}_1$ are the dominant eigenvalue and eigenvector of the CSM, respectively; e is a scalar enhancement factor controlled by $\psi_1 = \lambda_1 - \sigma^2$; σ^2 is the known uncorrelated noise variance (if it is unknown, it should estimated). The parameter β can be adjusted by changing *e* which effectively changes the resolution of the beamformer. The maximum resolution is achieved by setting $\beta = 1$, or equivalently $e \to \infty$. At the other extreme, a coefficient e=0 results in the conventional DS beamformer.

The *WNG* for this beamformer is derived in detail in Appendix B.2 for a general value of β. For the two extreme cases, it is given as

$$WNG = 10\log_{10}\frac{\left|\boldsymbol{w}^H\boldsymbol{d}\right|^2}{\boldsymbol{w}^H\boldsymbol{w}} = \begin{cases} 10\log_{10} M & (\beta = 0) \\ 10\log_{10} M\left(1-\left|\boldsymbol{m}_1^H\boldsymbol{g}\right|^2\right) & (\beta = 1) \end{cases} \tag{4.24}$$

This shows the EMVDR beamformer provides the possibility of trade-off between rejection to the uncorrelated noise and super-directivity characteristics depending on the adjustment factor β. For example, when $\beta = 0$, it becomes optimal to reject noise (as the conventional DS beamformer) but at the expense of poor resolution. It will be shown later that this beamformer has the superdirective properties while maintaining sufficient *WNG*. This is one of the key reasons this beamformer is superior to the other beamformers investigated.

e. Comparison of the WNGs

As mentioned, the *WNG* predicted in the above equations cannot be directly quantified except for the conventional DS. Here, the *WNG* for all beamformers is numerically computed for the simple 2D single building model (Figure 3.4). Firstly, a FDTD simulation is run for this model to obtain the steering vector for the source position, the CSM needed to compute the eigenvector in the EMVDR beamformer and the artificially generated matrix $\boldsymbol{\Phi}_{VV}$ in the Mennitt-MVDR. For this beamformer, $\boldsymbol{\Phi}_{VV}$ is computed based on all the possible grid points in the model with $\varepsilon = 0.01$. For the EMVDR, two adjustment factors ($\beta = 0.3$ and 0.7) are considered. The weighting vector for each beamformer is then computed at each frequency and inserted into the *WNG* equations for each beamformer.

Figure 4.6 shows the *WNG* for all the beamformers as a function of frequency. Firstly, it can be observed that the conventional DS beamformer suppresses uncorrelated noise equally at all frequencies, i.e. $10\log_{10} M = 6dB$ with M=4. The SNR in the DS beamformer is improved by 6 dB. On the other hand, the *WNG* for the MVDR is very poor, in particular at low frequencies because the uncorrelated noise is amplified. Note that the SNR for the MVDR at 60 Hz is degraded by almost 15 dB. The Mennitt-MVDR beamformer shows nearly the same WNG as the conventional DS beamformer fluctuating from 2.5 to 6 dB depending on the frequency. The results for the EMVDR beamformers show very good robustness to uncorrelated noise in particular for small β as expected. For $\beta = 0.3$, the *WNG* is nearly 6 dB at all frequencies. For $\beta = 0.7$, the *WNG* is reduced to similar values as the Mennitt-MVDR beamformer. However, the EMVDR beamformer has super-directivity characteristics lacking in the Mennitt-MVDR beamformer as it will be shown later in Chapter 5.

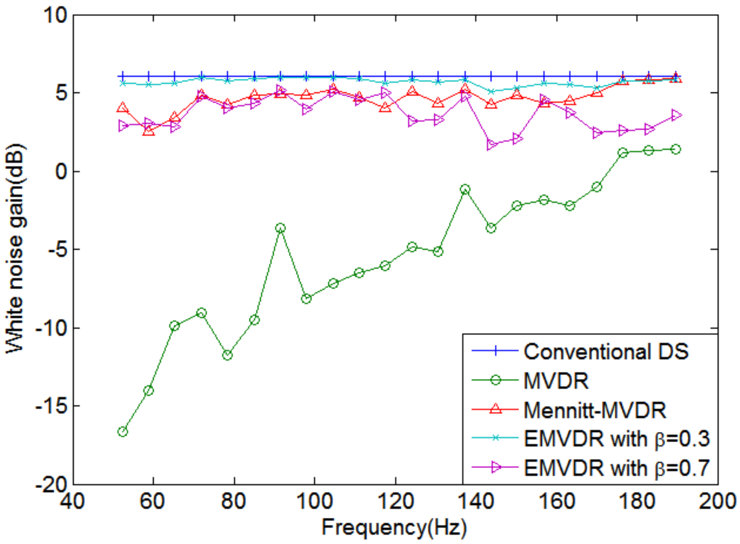


Figure 4.6: White noise gain (*WNG*) for all the beamformers investigated in the 2D single building model.

4.5 Source Localization Estimation

Localization performance using beamforming is model-specific as well as highly dependent on relative distance (or propagation length) between the source and the array. Therefore, the accuracy and confidence in the source localization should be investigated for multiple source-array configurations.

For the localization error analysis, the source location is estimated using a beamformer output cut-off threshold $\hat{b}_{cut-off}$, i.e. grid points with a normalized beamforming output lower than the cut-off threshold are ignored. That is, the source coordinate is estimated as

$$\hat{\boldsymbol{l}}_s = \frac{\sum_{j=1}^{M} \hat{b}_j \boldsymbol{l}_j}{\sum_{j=1}^{M} \hat{b}_j} \tag{4.25}$$

where M is the number of grid points whose beamforming outputs are greater than cut-off threshold, $\boldsymbol{l}_j = \{x_j, y_j, z_j\}$ is the coordinate vector for the j^{th} grid point, and $\hat{b}_j$ is the corresponding beamforming output included only if $\hat{b}_j \geq \hat{b}_{cut-off}$.

In an actual practical scenario, it is very unlikely that a grid point would be situated right where the source is located. Since the simulations are limited to the case where the source is located on a grid point, it is reasonable to choose a weighted average of the coordinates of the grid points rather than pinpoint the single location with largest beamforming output.

Figure 4.7 shows an example of the estimation of the source location using the beamforming cut-off threshold. Once the localization map is obtained (Figure 4.7a), the beamforming outputs that are lower than the cut-off threshold, $\hat{b}_{cut-off} = 0.6$, are discarded as shown in Figure 4.7b. Then, the source location estimate is obtained by the weighted average of the coordinates of the grid points whose beamforming outputs remain on the map. The actual and estimated source locations are indicated in the localization map using a circle (○) diamond shaped symbol (◊), respectively.

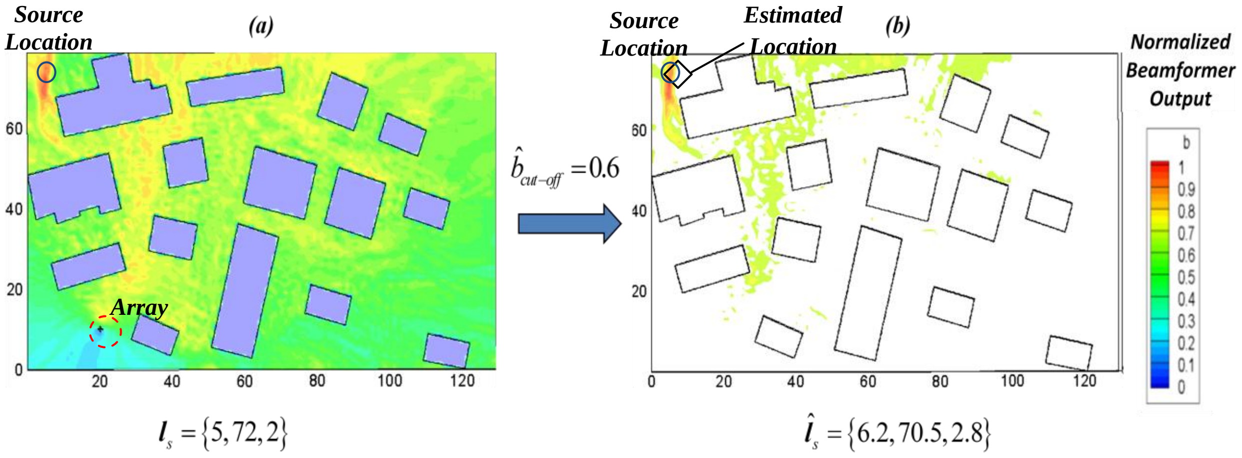


Figure 4.7: Example of localization error analysis using the beamforming output cut-off threshold in the Army Fort Benning model.

In an actual scenario, it is very unlikely that a grid point would be situated right where the source is located. Since the simulations are limited to the case where the source is located on a grid point, it is reasonable to exclude the beamforming output at the actual source position for a more realistic assessment.

The source localization error can then be computed as the Euclidean distance between the true and estimated source location as,

$$e = \left| \hat{\boldsymbol{l}}_s - \boldsymbol{l}_s \right| \tag{4.26}$$

where $\boldsymbol{l}_s$ is the true location vector of the source.

Chapter 5 Numerical Investigation using Beamforming

The beamformers described in the previous chapter are initially investigated using simple 2D and 3D models. These studies allowed for the initial assessment and down selection of the most promising beamformers. Based on this initial investigation, the EMVDR beamformer and the conventional DS beamformers are chosen for further research in the more realistic large 3D urban environments, which are discussed in the final evaluation section. All these simulations assume the perfect knowledge of the acoustic environment (i.e. transfer functions). However, there are always some uncertainties between the numerical model and the actual environment. Therefore, the sensitivity of the beamformers to the uncertainties in the model is discussed using numerical examples. In this study, two uncertainties are investigated: temperature gradient and unmodeled object. A way to alleviate the effects of uncertainty is also investigated.

5.1 Preliminary Evaluation

Illustrative results of the beamformers in the single and three-building models (Figure 3.2 through 3.7) are presented here. For all the cases, the beamforming grid resolution is 1 *m*. The narrowband beamforming processing uses all 64 spectral lines between 50 and 200 Hz and the 64 beamforming outputs are combined as explained in section 4.3. Then, the localization map is obtained by normalization as described earlier. For each localization map, a color scale is used to represent the numerical values: red represents a value of 1 (maximum) while the dark blue is 0 (minimum). The topology, source and array(s) are superimposed on the localization maps. Microphone self-noise is added for more realistic results and quantified in terms of the peak SNR. To this end, an artificial white noise is uniformly added to the environment. Finally, localization error is computed using three different cut-off thresholds of $\hat{b}_{cut-off} = 0.4$, 0.6, and 0.8, respectively. The estimated source location is shown in each localization map using the highest cut-off threshold, $\hat{b}_{cut-off} = 0.8$. Results are presented for (a) conventional DS, (b) MVDR, (c) Mennitt MVDR, and (d) EMVDR beamformers.

5.1.1 2D Models

Here, results for 2D models are presented for particular source-array configurations.

a. 2D Single-building model

In this model, diffraction of the sound waves around the two ends of the single building is the only propagation mechanism present in the environment. The array is located at (12, 18) *m* while the source is at (30, 10) *m*. The LoS from the source to the array is completely blocked by the object.

The localization maps for the beamformers without and with noise are shown in Figure 5.1 and 5.2, respectively. For the case with noise, the peak SNR is 5dB. Figure 5.1 shows that, in the absence of noise, the MVDR and EMVDR beamformers are very accurate showing the expected super-directivity characteristics. On the other hand, the DS and Mennitt-MVDR beamformers show significant ambiguity in the source location. Moreover, they produce almost identical results supporting the argument that the Mennitt-MVDR is essentially the same as the DS beamformer.

The results in Figure 5.2 for a SNR of 5 dB show that all the beamformers other than the MVDR are robust to uncorrelated noise, i.e. maps for DS, Mennitt-MVDR, and EMVDR beamformers are very similar in Figures 5.1 and 5.2. The MVDR is, however, extremely sensitive to noise as shown by Figure 5.2c where a number of fictitious source candidate locations are now observed. As a reminder and mentioned in section 4.4, a constant was artificially added to the diagonal of the CSM to compute its inverse because the CSM is rank deficient.

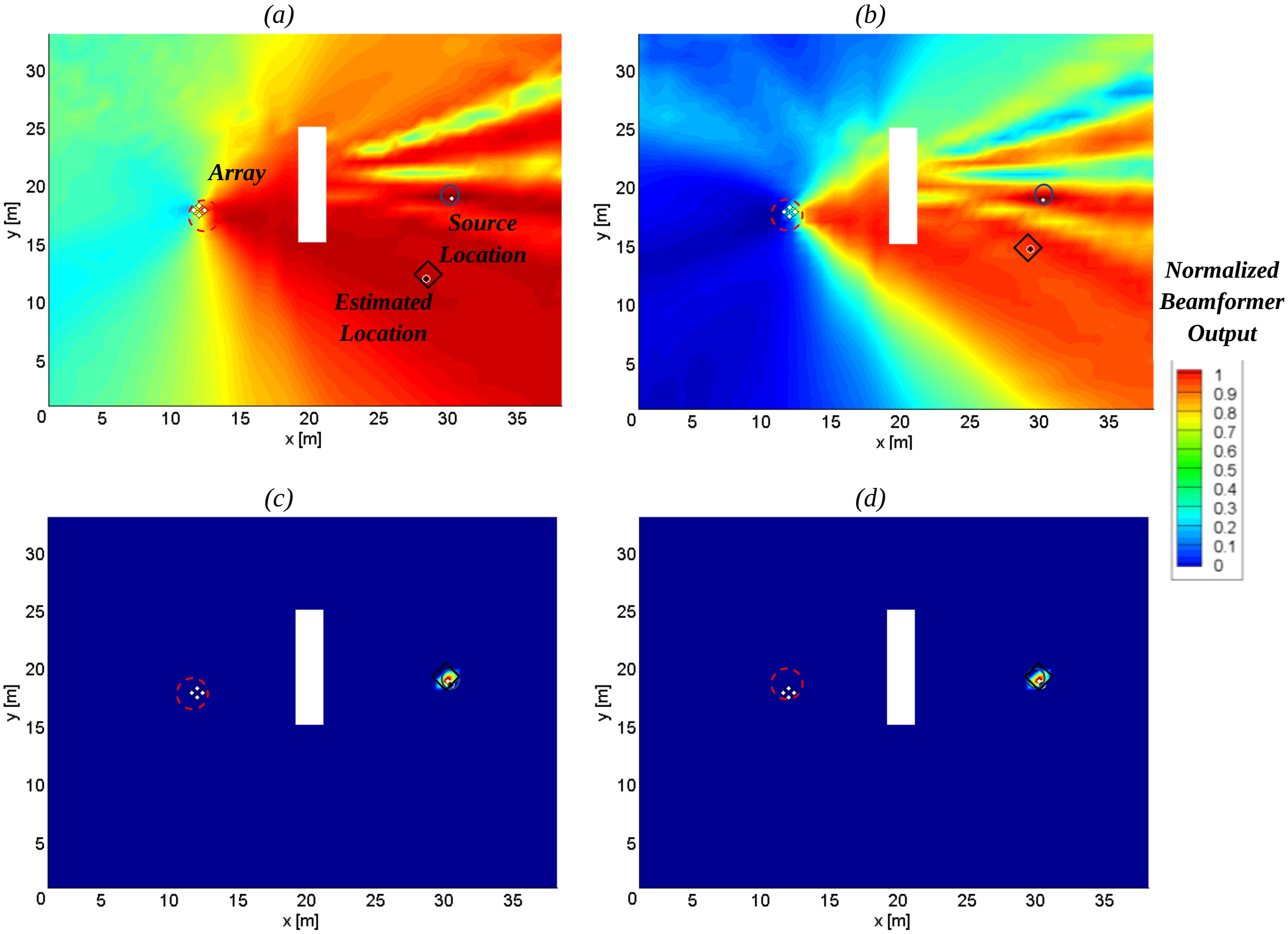


Figure 5.1: Localization maps for 2D single-building model without noise: (a) conventional DS (b) Mennitt-MVDR (c) MVDR and (d) EMVDR beamformers.

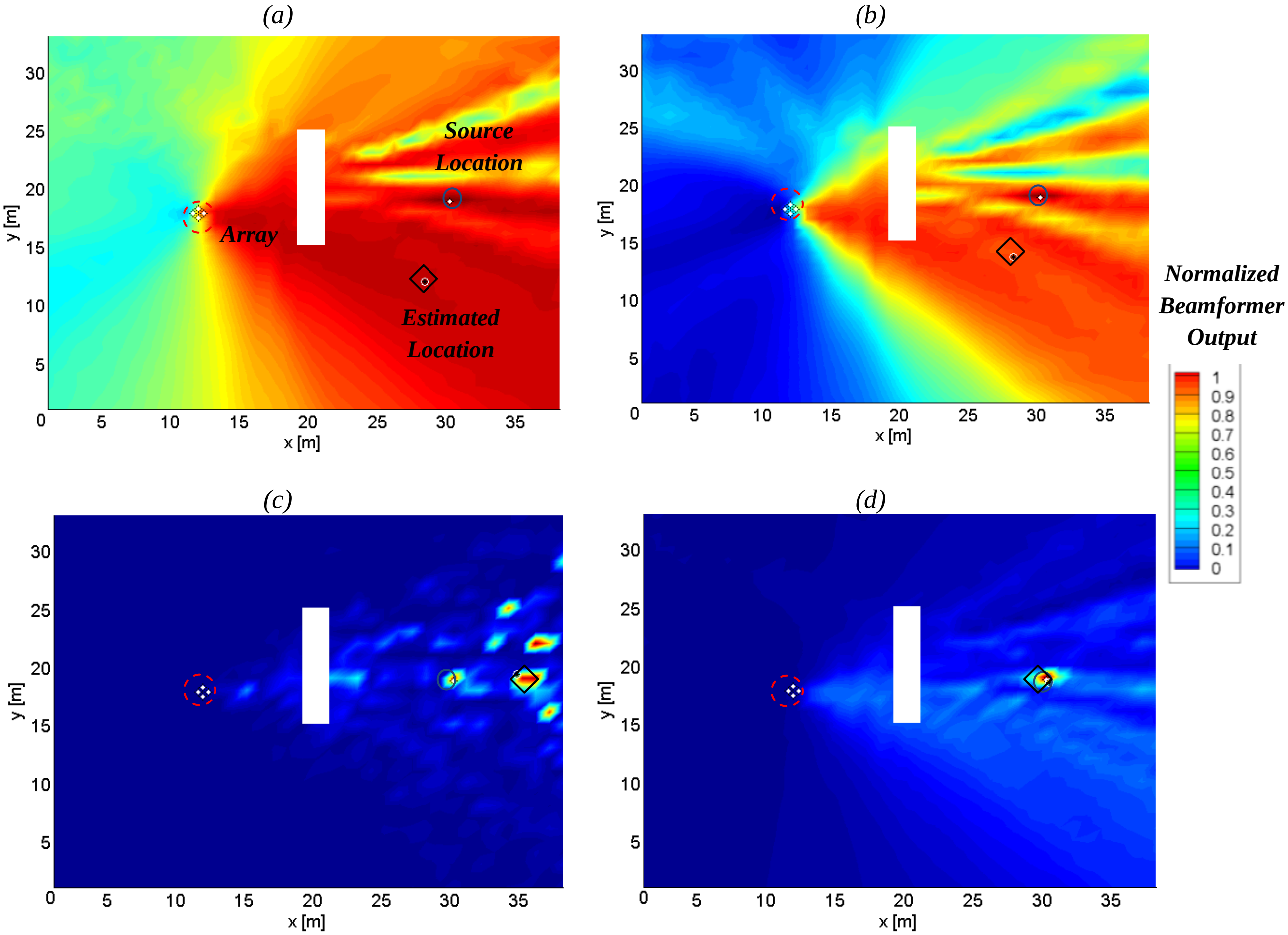


Figure 5.2: Localization maps for 2D single-building model with 5dB SNR noise: (a) conventional DS (b) Mennitt-MVDR (c) MVDR and (d) EMVDR beamformers.

To illustrate the adjustable resolution of the EMVDR beamformer, Figure 5.3 shows the results for different adjustment factors from the lowest ($\beta = 0$) to the highest resolution ($\beta = 1$). For the case of $\beta = 0$, the EMVDR reduces to the DS beamformer as mentioned previously in Chapter 4. Thus, Figures 5.2a and 5.3a are identical. As the adjustment factor is increased, the resolution is improved and a more accurate estimate of the source location is achieved. For a $\beta = 0.3$, the estimated source location is less than 3 *m* from the actual source placement.

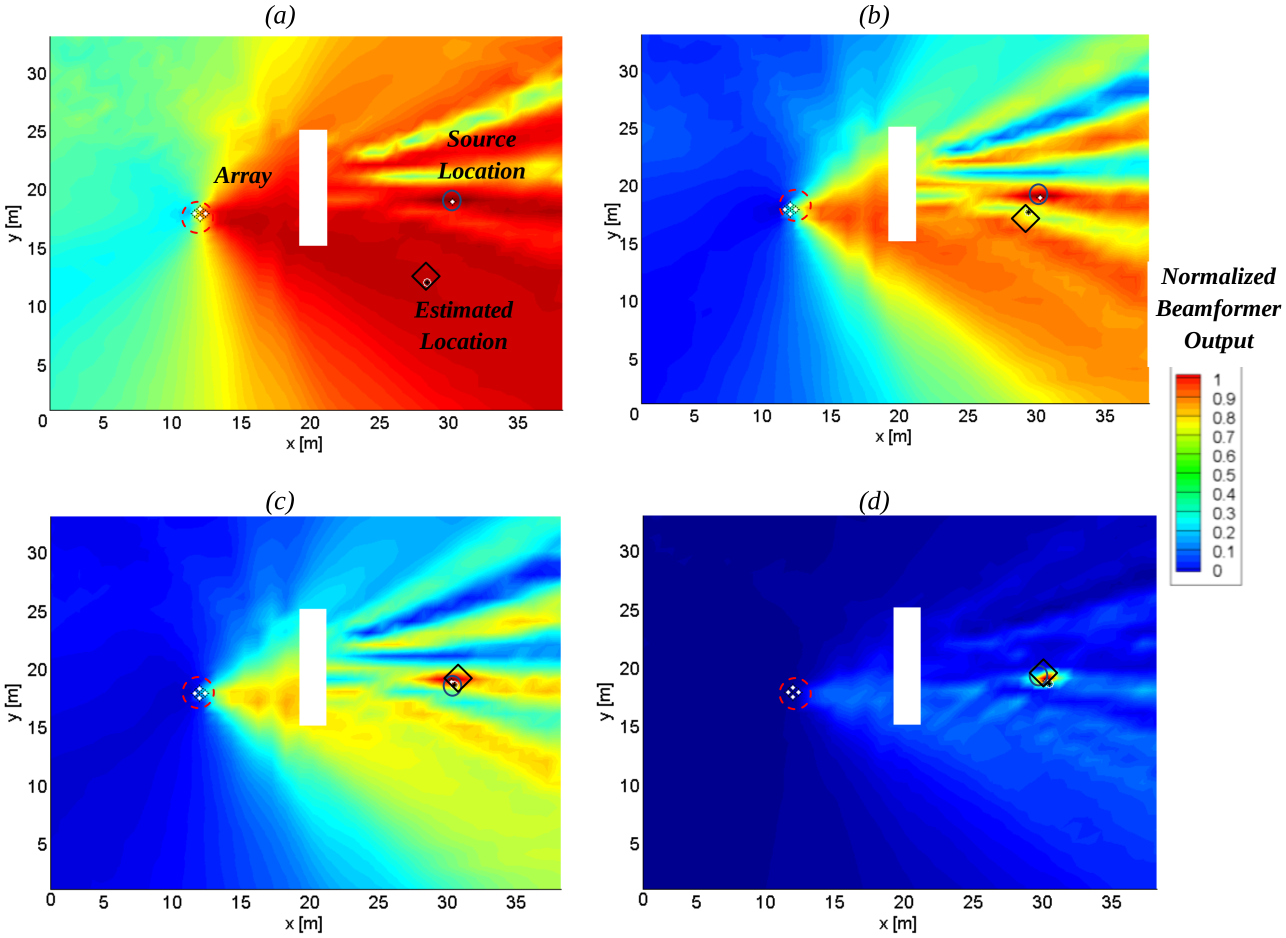


Figure 5.3: Localization maps of the EMVDR beamformer for 2D single-building environment with 5dB SNR noise with four adjustment factor: (a) $\beta = 0$, (b) $\beta = 0.3$, (c) $\beta = 0.7$ and (d) $\beta = 1$.

b. 2D Three-building model

Here the beamformer performance is investigated for the case of multiple buildings in the environment. Unlike the previous case, the 3 buildings introduce multiple reflections in addition to the edge diffraction effects thus resulting in many more propagation paths from the source to the array. The array is located at (7, 38) *m* and the source is at (40, 30) *m* where three buildings obstruct the direct LoS from the source to the array and introduce cluttering through diffraction and reflection effects. The results for the beamformers are presented in the same format as illustrated in Figure 5.1. The localization maps without and with 10dB SNR noise in terms of peak values are depicted in Figure 5.4 and 5.5, respectively. Due to the coarse beamforming grid resolution (1 *m*) used, the maps show step-like edges of the buildings. Hence, the actual edges of the buildings are additionally drawn on the maps.

Figure 5.4a and 5.4b shows that the DS and Mennitt-MVDR beamformers suffer from the cluttering in the environment resulting in higher localization error, even in the absence of noise. Note that for a potential user, the ambiguity shown in these two localization maps would be of little use to assess the actual source location. On the other hand, the MVDR and EMVDR beamformers are still very accurate in spite of the clutter (Figure 5.4c and 5.4d).

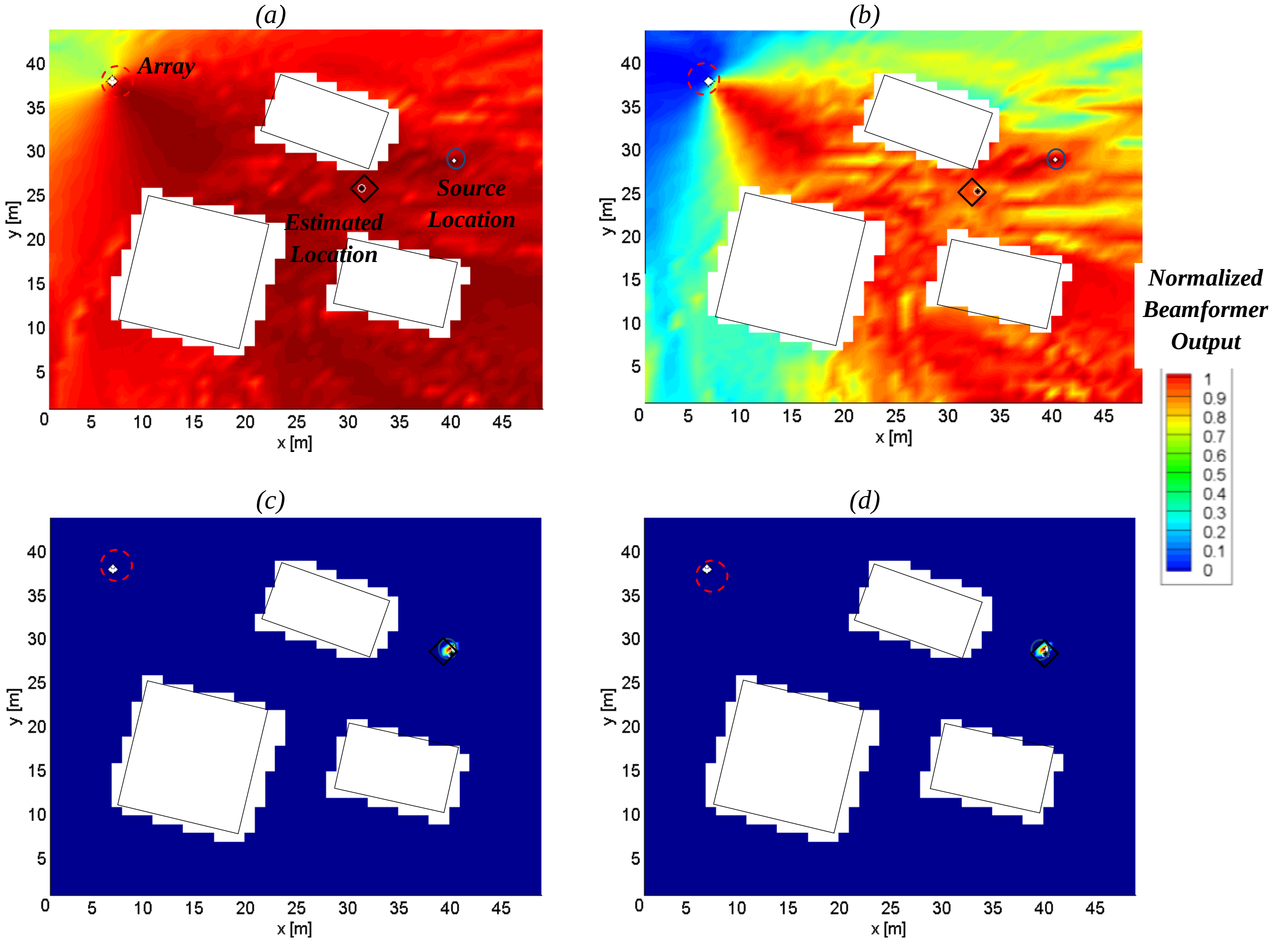


Figure 5.4: Localization maps for 2D three-building model without noise: (a) conventional DS (b) Mennitt-MVDR (c) Mennitt MVDR and (d) EMVDR beamformers.

The results in Figure 5.5a and 5.5b for a SNR of 10 dB show the robustness of the DS and Mennitt-MVDR to uncorrelated noise as the results are very similar in Figure 5.4a and 5.4b. A more important observation is, however, that the localization performance of the beamformers is not very acceptable in the presence of noise with the exception of the EMVDR beamformer. For EMVDR, the estimated source location is less than 2 *m* from the actual source placement while the other beamformers yield more than 10 *m* localization errors.

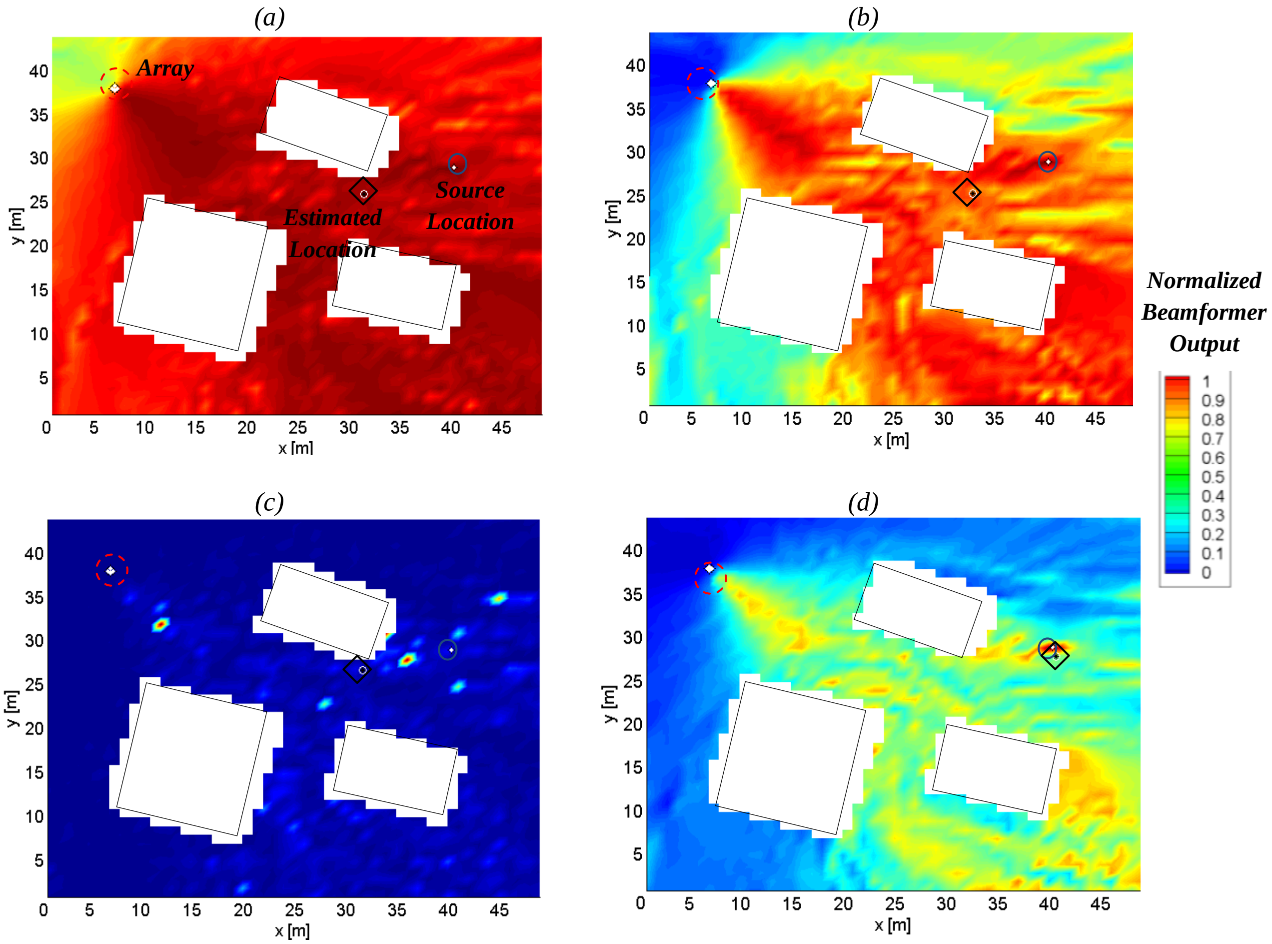


Figure 5.5: Localization maps for 2D three-building model with 10dB SNR noise: (a) conventional DS (b) Mennitt-MVDR (c) MVDR and (d) EMVDR beamformers.

In order to assess the source localization performance, the localization error for different levels of uncorrelated noise is computed. Figure 5.6 compares the localization performance of the beamformers for the single and three-building models for SNR ranging from -5 to 20 dB. It is noteworthy that SNR of 5 dB or lower are extreme cases. For example, a 0dB SNR indicates that the peak value of the noise is identical to that of the signal. The results in Figure 5.6 were obtained by performing Monte Carlo trials by introducing 300 sets of white noise for different SNRs while keeping the same position of the source and the array. The localization error, as given by eqs. (4.24) and (4.25), is then measured and averaged over the number of trials. Thus, the results in Figure 5.6 represent the expected localization errors.

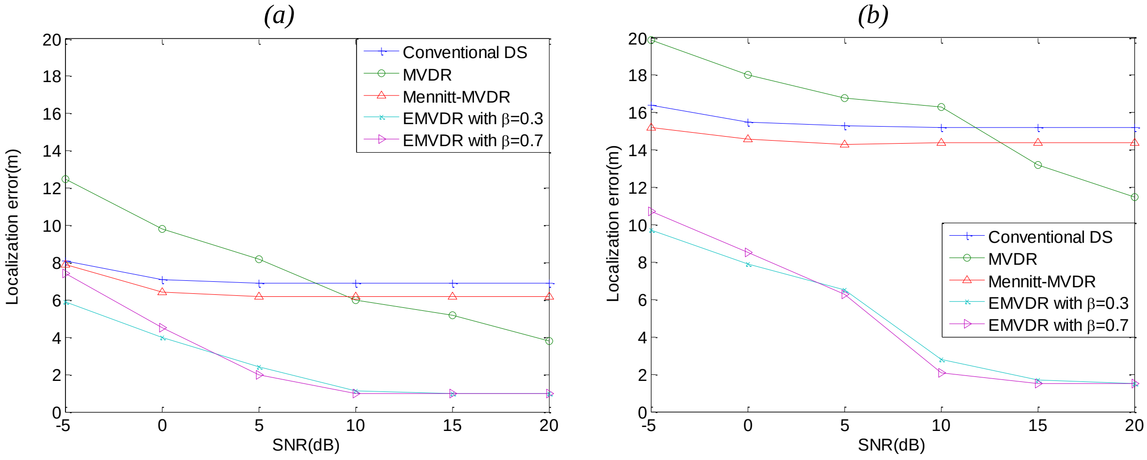


Figure 5.6: Localization error versus SNR using the cut-off threshold $\hat{b}_{cut-off} = 0.8$: (a) 2D single-building model and (b) 2D three-building model.

The results in Figure 5.6a for the single building model show that the conventional DS and the Mennitt-MVDR are relatively insensitive to uncorrelated noise (i.e. localization error is almost constant as a function of SNR). This is because of the *WNG* of both beamformers (i.e. 6dB for the conventional DS), indicating the strength of the signal is boosted due to the microphones in the array so that actual SNR becomes higher. At very low SNR (-5dB), however, it is found that the localization errors of both beamformers begin to increase. The only difference between the conventional DS and the Mennitt-MVDR is that the localization error of the Mennitt-MVDR is slightly less than the conventional DS. However, both beamformers are incapable to accurately locate the source, causing relatively large average localization errors. On the contrary, MVDR shows its sensitivity to noise such that localization error increases as SNR decreases. For large SNR (i.e. >15dB SNR), the localization error of the MVDR is actually smaller than the conventional DS. However, it gets worse than that of the conventional DS as the noise level increases. Most importantly, the EMVDR beamformer demonstrates its accurate localization capability in particular for high SNRs, i.e. localization error is less than 2 *m* for SNR > 10 dB. It is also observed that the localization accuracy of the EMVDR degrades as SNR decreases. However, the localization performance of the EMVDR is still much better than the other beamformers even at low SNRs. At 0dB SNR, the EMVDR beamformer with higher adjustment factor becomes more sensitive to noise, resulting in slightly larger localization error.

The results for the three building model are presented in Figure 5.6b. They show the same trends as for the single building case. Comparing Figure 5.6a and Figure 5.6b shows, however, the detrimental effect of the cluttering in the three-building model (i.e. larger localization errors in all beamformers) in particular for the DS, MVDR, and Mennitt-MVDR beamformers. For the DS, the localization error doubles from 7 *m* to 15 *m*. The localization performance of the Mennitt-MVDR and MVDR follow the similar pattern. On the other hand, the localization error for the EMVDR is always substantially less depending on the SNR.

These first set of results shows that the EMVDR beamformer clearly outperforms the others. It basically shows good robustness to noise while demonstrating super-directivity characteristics critical for accurate localization.

5.1.2 3D Models

The results of the 3D version of the previous 2D models are presented here for a direct comparison between 2D and 3D localization. To this end, the top-view of the localization map is shown in the figure, i.e. localization map of the horizontal plane at z-coordinate of the source.

Similar to the previous section, the localization error results are also presented. It is important to note that the number of microphones in the array used for 3D models is 6 instead of 4 for 2D models.

a. 3D Single building model:

The array and source are located at the same position in the x-y plane and 2 *m* above the ground. Here, the localization maps of the beamformers with 5dB SNR noise are shown in Figure 5.7 for the 3D (first column) and the corresponding 2D (second column) model. The 2D results are the same as in Figure 5.2. Each row corresponds to each beamformer used: *(a)* Conventional DS, *(b)* Mennitt-MVDR, *(c)* MVDR and *(d)* EMVDR beamformer with $\beta = 1$.

The results in the first column of Figure 5.7 show that the beamformers follow the same trends in 3D. However, comparing the first to the second column demonstrates that all the beamformers in 3D case show less ambiguity in the source location. This can be explained by an additional propagation path over the top of the building (dotted line in Figure 5.7a) to the original edge diffractions around the building (solid line in Figure 5.7a) as mentioned in Chapter 3.3. In other words, the small height of the building (3*m* height) allows large amount of sound (and thus information) over the building to improve the localization. Hence, the improvement of the localization is expected to degrade with the larger building.

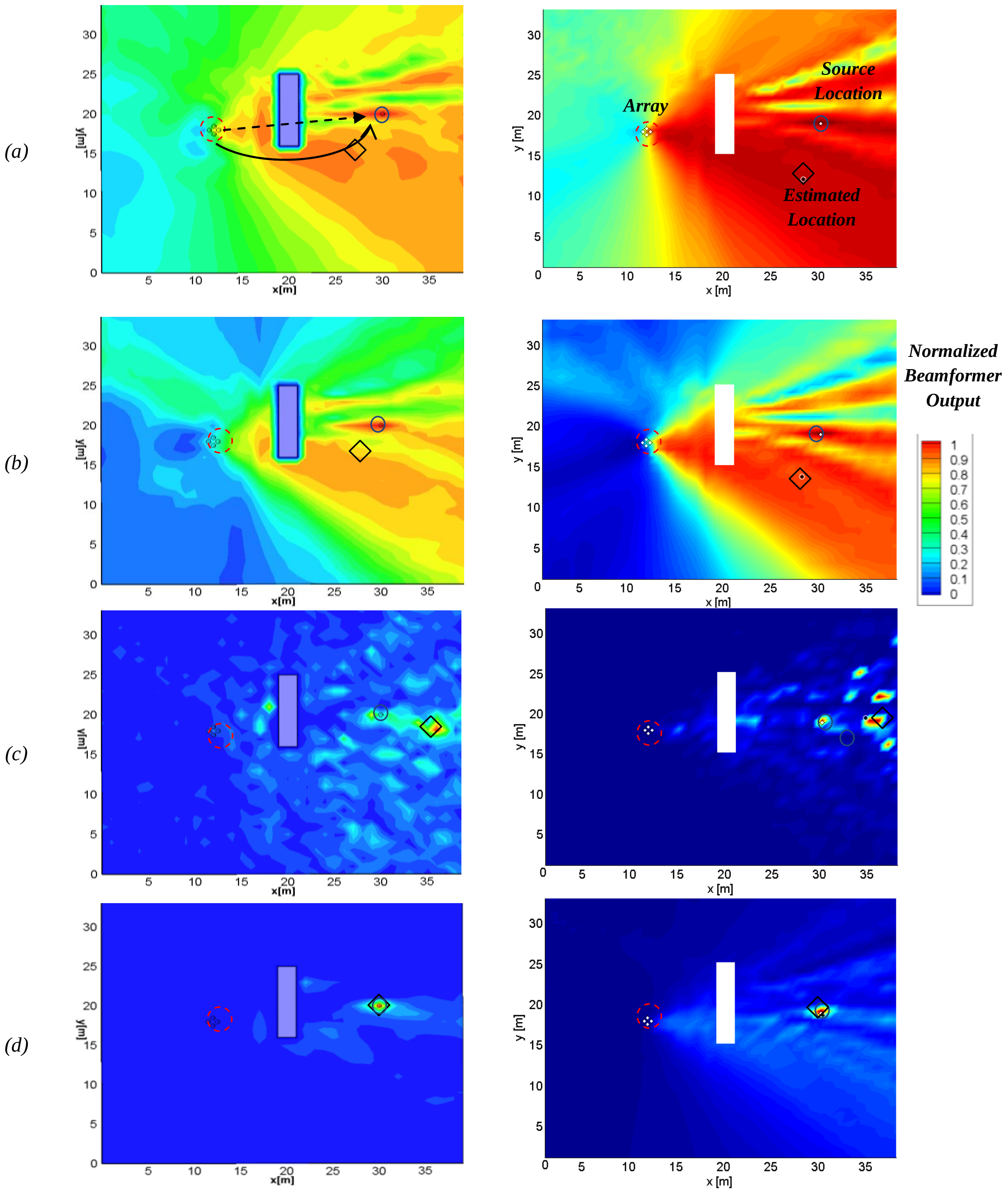


Figure 5.7: Localization maps for 3D single-building model (left) and the corresponding 2D model (right) with 5dB SNR noise: (a) conventional DS, (b) Mennitt-MVDR beamformer, (c) MVDR and (d) EMVDR beamformers.

Figure 5.8 compares the localization performance of the beamformers between the 3D single building and the corresponding 2D model again as a function of SNR. Firstly, the results in Figure 5.8a for the 3D single building model show the same trends regarding the robustness of the beamformers to uncorrelated noise among the beamformers as in the 2D case. Comparing Figure 5.8b to Figure 5.8a shows that 3D localization is more accurate than the corresponding 2D localization. For example, the localization error for the conventional DS beamformer decreases from 7 *m* to 3.5 *m*. Moreover, the localization error of the EMVDR at low SNRs become no longer an issue in 3D (i.e. the error is less than 2 *m*).

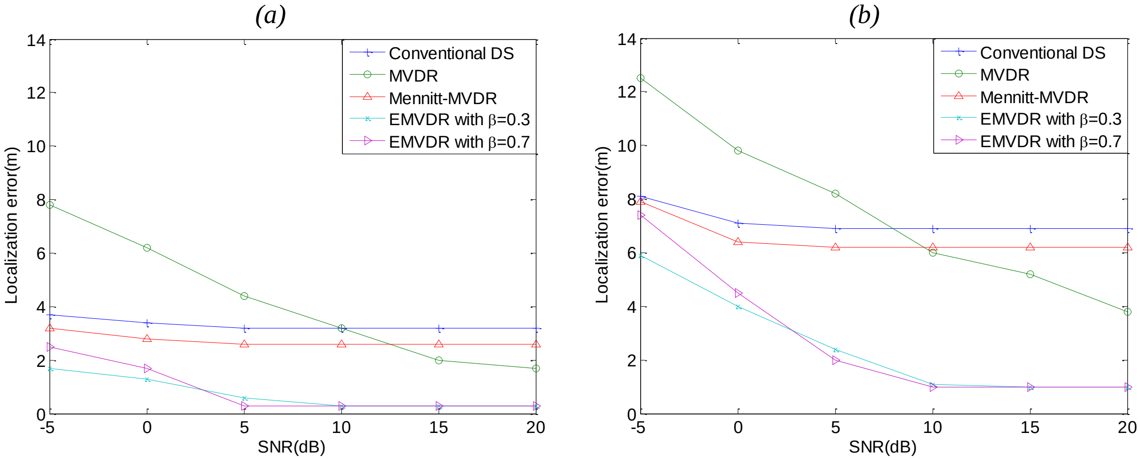


Figure 5.8: Localization error versus SNR using the cut-off threshold $\hat{b}_{cut-off} = 0.8$:
(a) 3D single-building model and (b) corresponding 2D model.

b. 3D Three-building model:

Similar to the single building case, the array and source are located at the same position in the x-y plane and 2 *m* above the ground. The results for the beamformers are presented in a format with 2 columns and 3 rows as illustrated in Figure 5.9. The results on the left column of the figure show top-view of the localization maps in the 3D three-building model while the ones on the right column shows the corresponding 2D localization maps. The three rows correspond to the beamformers used: conventional DS, Mennitt-MVDR, and EMVDR beamformer with $\beta = 1$.

From this point forward, the MVDR beamformer is eliminated from consideration due to its high sensitivity to noise.

The results show that the beamformers perform much better in 3D. This is partly because of the diffraction effects over the buildings as discussed previously in the 3D single building model. Another reason is that there are a lot more propagation paths between the source and the array as compared to the 2D model when the third dimension is added in a cluttered environment. This means that performance of the source localization improves as it takes advantage of additional information on the sound propagation in the 3D environments.

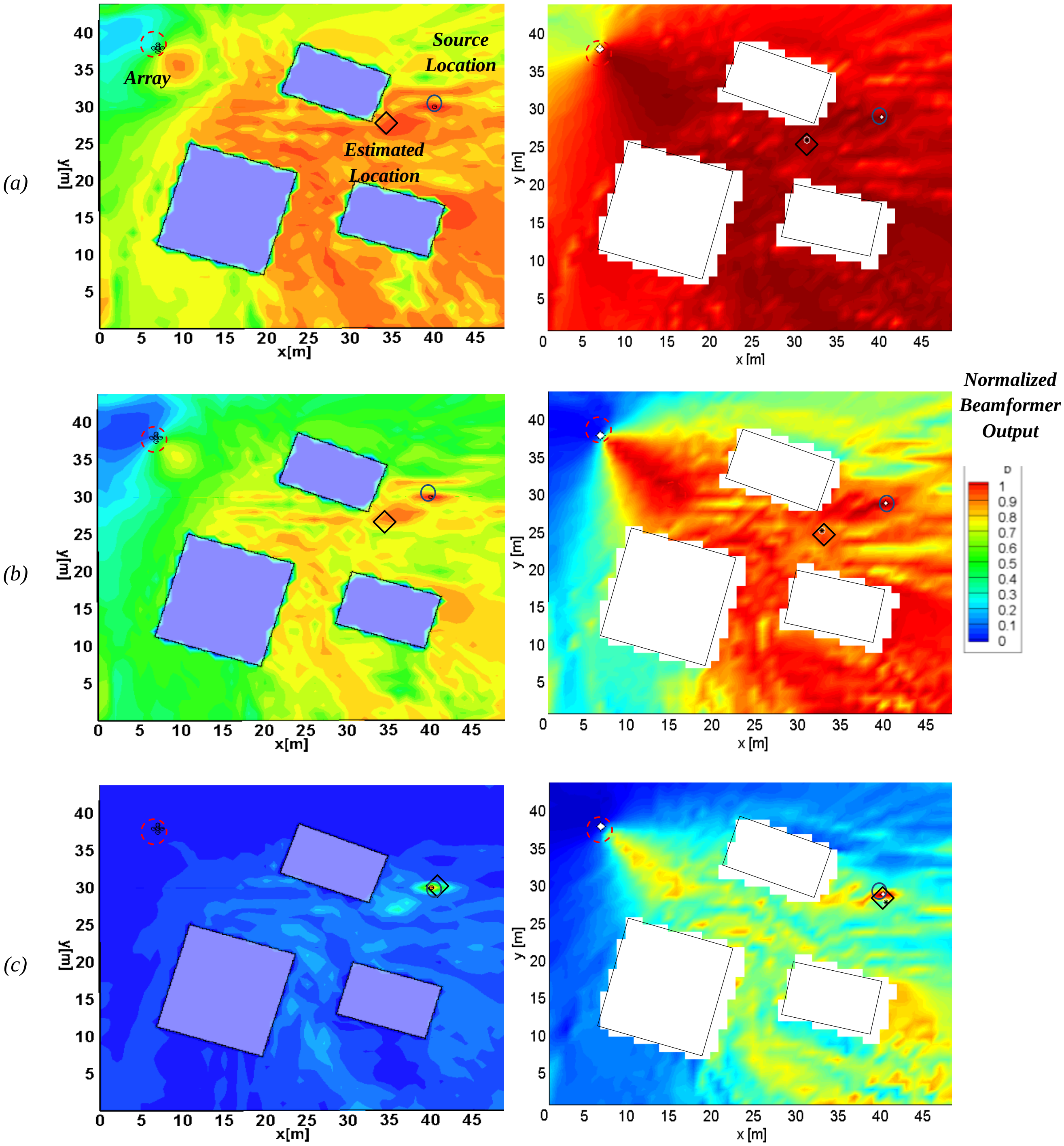


Figure 5.9: Localization maps for 3D three-building model (left) and the corresponding 2D model (right) with 10dB SNR noise: (a) conventional DS, (b) Mennitt-MVDR beamformer and (c) EMVDR beamformers.

Figure 5.10 compares the localization performance of the beamformers between the 3D three-building model and the corresponding 2D model as a function of SNR. Figure 5.10a shows that only the EMVDR is capable of localizing the source accurately. Compared to Figure 5.10b, the results in Figure 5.10a demonstrate that 3D localization yields smaller errors that the corresponding 2D localization in the three-building model. For conventional DS, the localization error is reduced from 16 *m* to 11 *m*. On the other hand, the performance improvement of the EMVDR is more remarkable especially at low SNRs.

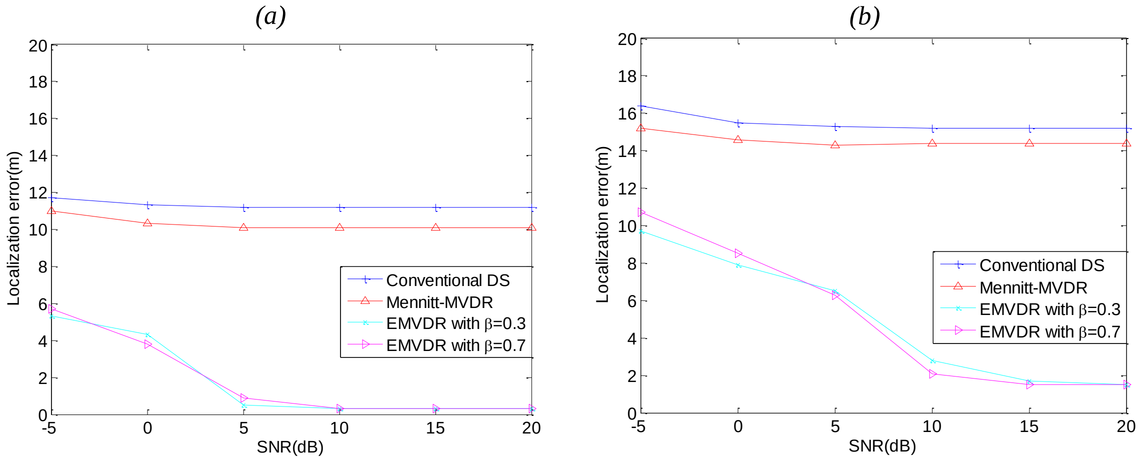


Figure 5.10: Localization error versus SNR using the cut-off threshold $\hat{b}_{cut-off} = 0.8$:
(a) 3D three-building model and (b) corresponding 2D model

From these preliminary results, there are two important observations. Firstly, 2D results show larger errors in the localization because the 3D models have more propagation paths (i.e. more information about the environment) to yield more accurate localization. Therefore, it is concluded that evaluation in 3D models is critical to assess correctly the potential of this localization technique. Hence, 2D model is no longer considered for further evaluation due to its inadequate representation of the realistic environment. Secondly, the EMVDR is the only beamformer capable of accurate source localization and still robust to noise (at least using a single array). The Mennitt-MVDR is essentially the same as the conventional DS and the MVDR is extremely sensitive to noise. Because of its superior performance, the EMVDR beamformer will be used for further evaluation. The conventional DS beamformer will still be evaluated for comparison purposes to the EMVDR beamformer.

5.2 Final Evaluation

In this section, the selected EMVDR beamformer is evaluated in the two more realistic large 3D urban models described in Chapter 3.2: simplified urban and Army Fort Benning model. As previously discussed, the localization performance is expected to degrade as the environment gets more complex. Results using the conventional DS beamformer are also obtained for comparison purposes.

The localization maps are generated in the same way as the preliminary evaluation. However, localization maps are presented from different angles to better visualize the results in the 3D environments. Firstly, a 3D view of the localization map plots the data on the three orthogonal planes passing through the position of the source. Secondly, the localization map of the vertical plane normal to the y-axis at y-coordinate of the source is presented (side-view). Finally, a top-view shows the localization map of the horizontal plane at z-coordinate of the source.

In addition, multiple-array processing using two arrays is considered for this study to investigate the performance enhancement due to the additional array. It is noteworthy that the array locations will remain the same in all the simulations.

a. *Simplified Urban Model:*

In these simulations, the source is at (33, 53, 2) *m* while the arrays are located at (112, 12.5, 2) and (20, 75, 2) *m*, respectively. Figure 5.11 shows the noisy pressure received at one of the microphones in each array. The source signal arrives to the microphone at different times due to the different propagation path lengths. In fact, the source signal arrives earlier to the microphone in array 2 than the microphone in array 1. The average SNRs for the microphone in array 1 and 2 are 5 and 12 dB, respectively.

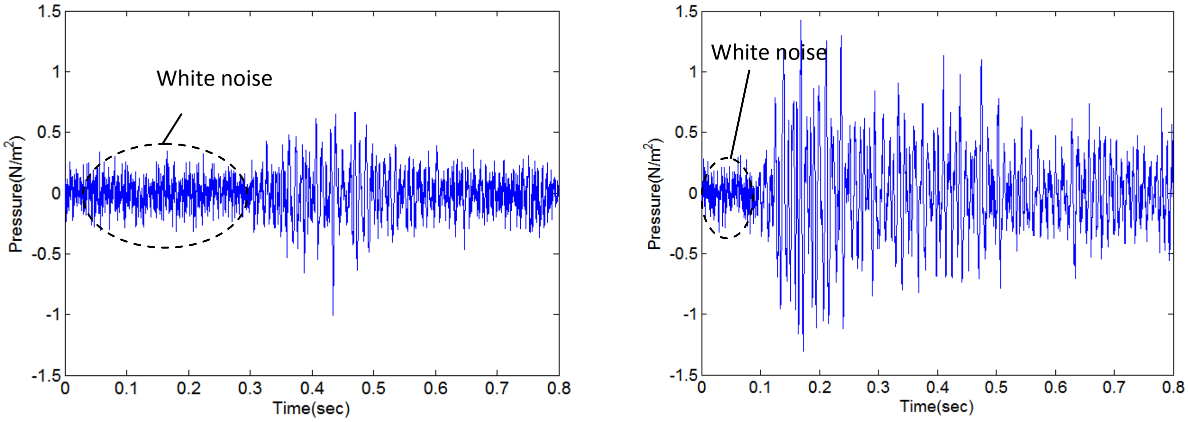


Figure 5.11: Time history of the microphone signal received at array 1(left) and array 2(right) in the simplified urban model. SNR values are 5 and 12dB, respectively.

To illustrate the performance of the EMVDR beamformer using array 1 only, Figure 5.12 presents the localization map with 5dB SNR noise in a matrix like format with 2 columns and 3 rows. The results in the left and right columns correspond to the conventional DS and EMVDR ($\beta = 0.3$) beamformers, respectively. The three rows correspond to three views of the localization maps described earlier.

The results in Figure 5.12a show that the localization performance of the conventional DS is very poor showing significant ambiguity in the environment. The localization error is more than 25 *m*. On the other hand, the results in Figure 5.12b show that the EMVDR beamformer ($\beta = 0.3$) provides relatively accurate localization capability, i.e. the estimated source location is less than 5 *m* from the actual source placement. Moreover, the EMVDR beamformer ($\beta = 0.3$) shows the expected super-directivity characteristics in this model.

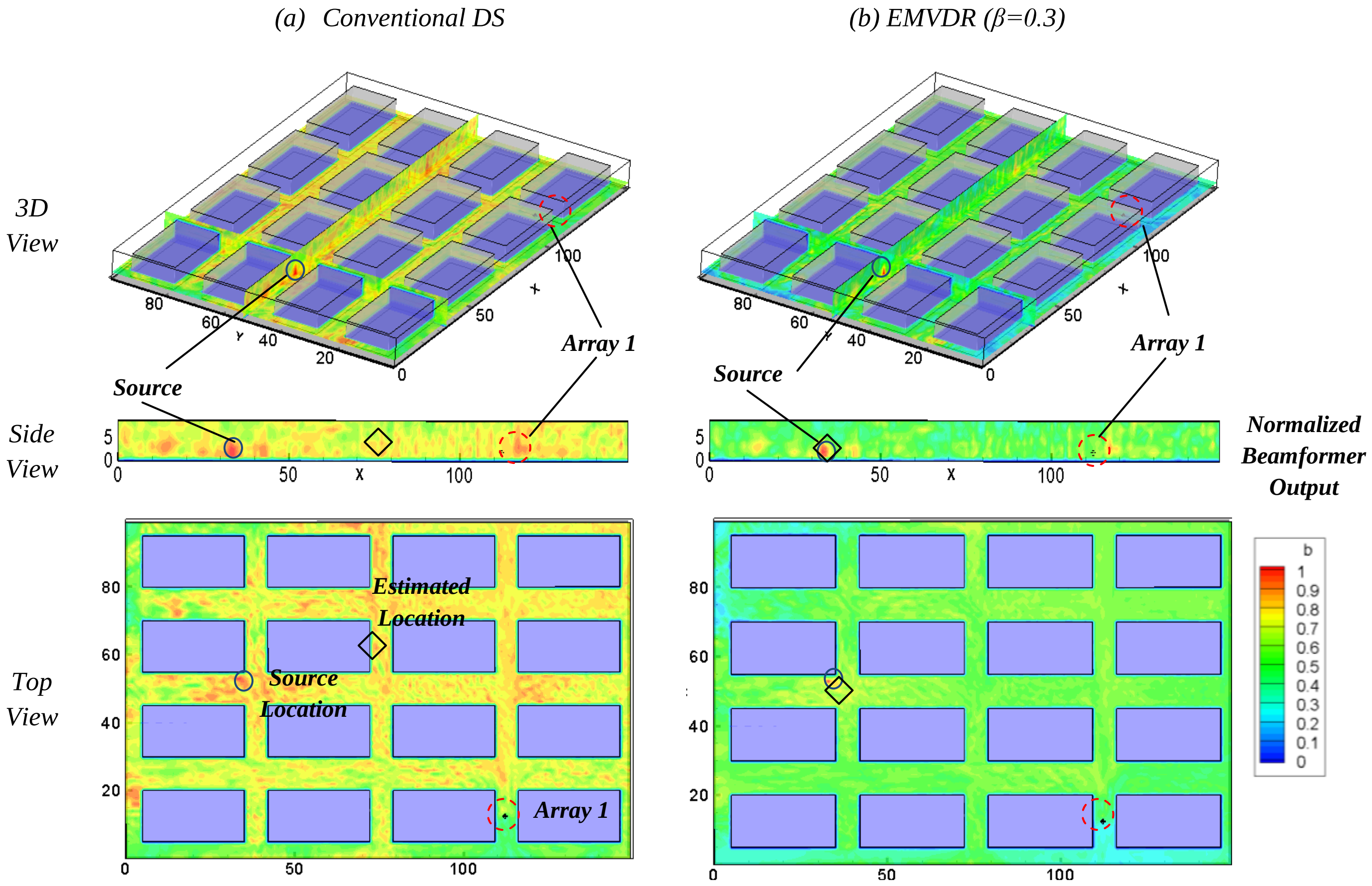
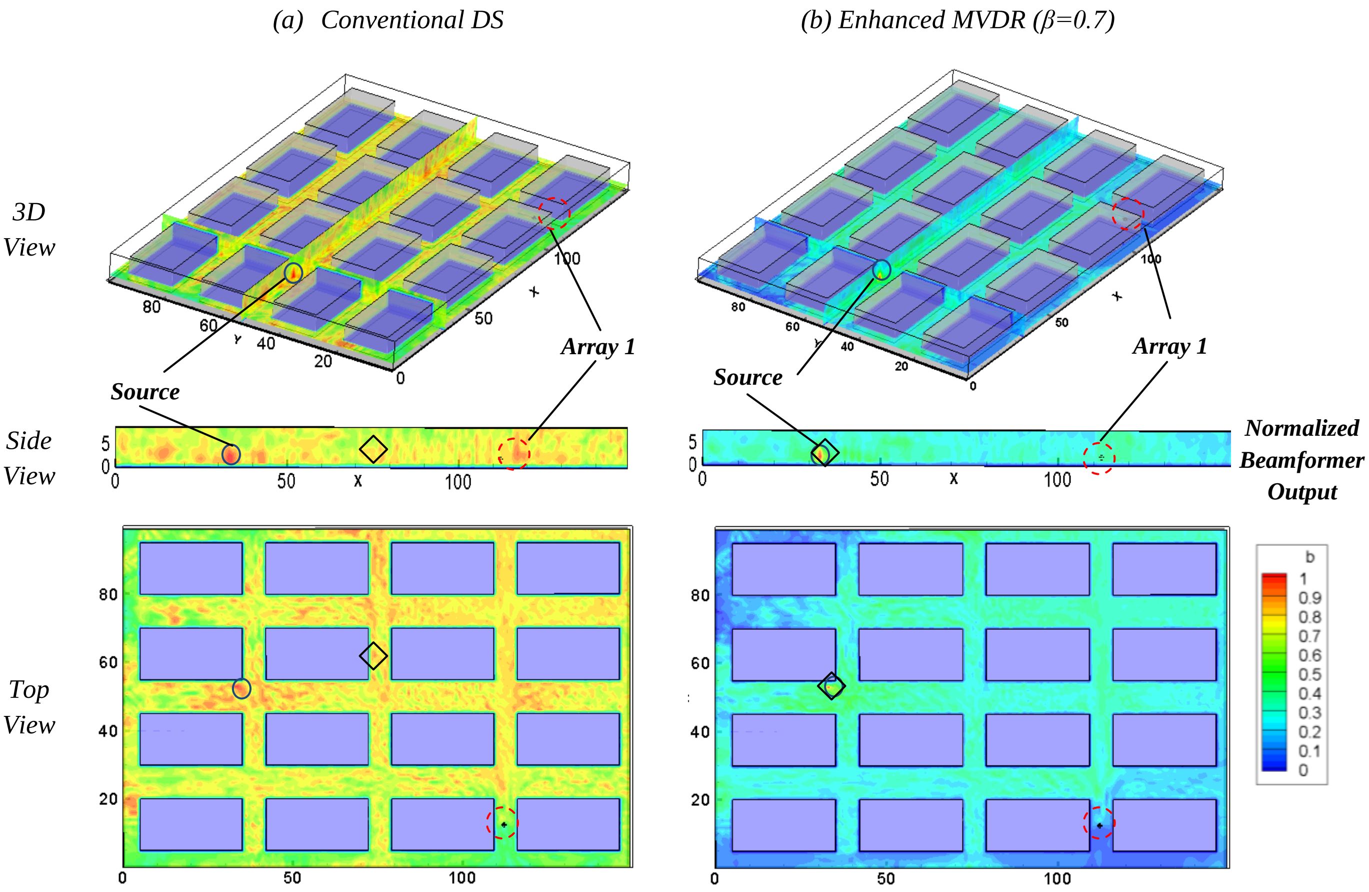


Figure 5.12: Localization maps for simplified urban environment model with 5dB SNR noise using (a) conventional DS and (b) EMVDR with β=0.3 beamformers. Localization error is (a) 26.2*m* and (b) 4.6*m*.

Figure 5.13 shows the results for the EMVDR beamformer using higher adjustment factor (β = 0.7) versus the conventional DS beamformer. As the adjustment factor is increased, the resolution is improved. In fact, the EMVDR beamformer (β = 0.7) was able to locate the source very accurately using a single array despite of the cluttering introduced by multiple buildings and the noise as shown in Figure 5.13b.


(a) Conventional DS

(b) Enhanced MVDR (β=0.7)

Figure 5.13: Localization maps for simplified urban environment model with 5dB SNR noise using (a) conventional DS and (b) EMVDR with β=0.7 beamformers. Localization error is (a) 26.2*m* and (b) 0.4*m*.

Figure 5.14 compares the results for the single and two array configurations for the conventional DS and EMVDR beamformer (β = 0.3). This figure shows the top view of the localization performance. In addition, a zoom of the 3D view in the vicinity of the source is also shown to provide more clear view on the estimated source location. The results show that the additional array significantly improves the localization performance. For example, the localization error of the conventional DS beamformer decreases from 26 *m* using a single array to around 7 *m* with two arrays (Figure 5.14a and 5.14b). Still, the localization performance of the conventional DS beamformer is not very accurate even with two arrays. However, the localization error of the EMVDR beamformer (β = 0.3) is reduce from 4 *m* using a single array to less than 1 *m* with two arrays (Figure 5.14a and 5.14b), indicating that the EMVDR beamformer (β = 0.3) has very accurate localization capability with the additional array.

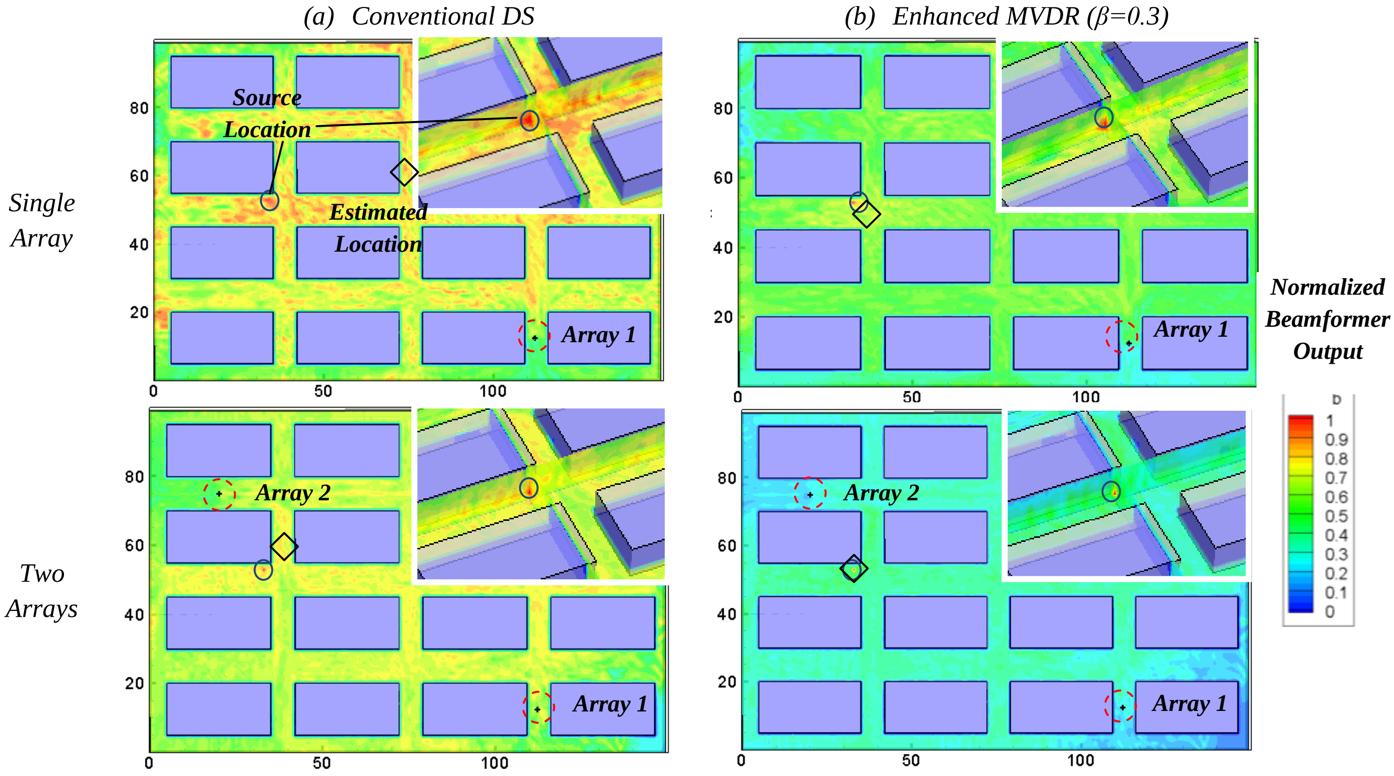
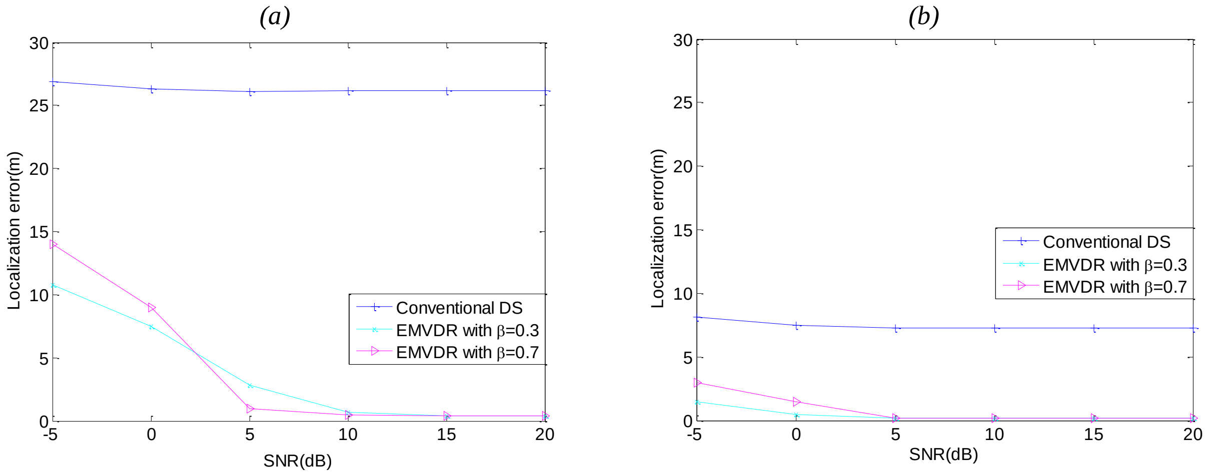


Figure 5.14: Localization maps (top-view) for simplified urban environment model using (a) conventional DS and (b) EMVDR with β= 0.3 beamformers with one and two arrays whose SNRs are 5dB and 12dB, respectively. Localization error using two arrays is reduced to (a) 7.3*m* and (b) 0.2*m*.

In order to compare the source localization performance for singe and two array configurations, the localization error of the beamformers for different SNRs (defined in terms of array 1) is computed via Monte Carlo simulation. Note that the average SNR at array 2 is around 7dB higher than the average SNR at array 1 due to its closer location to the source given the same amount of noise in the environment.

Figure 5.15 shows the localization performance for (a) the single array and (b) two-array configuration for three beamformers: conventional DS, EMVDR with β = 0.3 and 0.7. The results in Figure 5.15a for a single array show that the conventional DS is apparently incapable of locating the source, causing large localization error. However, the EMVDR beamformer demonstrates its accurate localization capability, i.e. localization error is less than 2 *m* for SNR > 10 dB. The localization accuracy of the EMVDR beamformer degrades as SNR decreases as expected, but the localization performance is still far better than the conventional DS even at low SNRs.

The results for two-array configuration are presented in Figure 5.15b. Comparing Figure 5.15a and Figure 5.15b shows that multiple arrays can make the conventional DS beamformer feasible. For the EMVDR, the localization becomes very accurate even at low SNRs while the conventional DS beamformer still causes relatively large localization error. It is noteworthy that the additional array is located at a more favored position, i.e. closer to the source than the original array. Hence, it is not guaranteed that an additional array always makes such a positive impact on the overall localization performance.


(a)

(b)

Figure 5.15: Localization error versus SNR at array 1 for 3D simplified urban model using $\hat{b}_{cut-off} = 0.8$: (a) single array and (b) two array configuration.

b. Fort Benning Urban Model:

For the Fort Benning model, the locations of the array 1 and 2 are (20, 10, 2) and (60, 60, 2) *m*, respectively. In this model, the source is located at (5, 72, 2) *m*. For this particular source-array configuration, the average SNRs for array 1 and 2 are 5 and 8 dB, respectively. The same types of results as in the previous case are presented in Figures 5.16 through 5.18.

In general, there exists more uncertainty on the location of the source due to the higher complexity of this model. As in the case of the simplified urban model, the conventional DS performs very poorly. Specifically, the top view for the conventional DS beamformer shows a large normalized beamforming output ≥ 0.8 (orange) over most of the domain (Figure 5.16a). On the other hand, the EMVDR beamformer ($\beta = 0.3$) shows significantly better performance. The beamforming output of the EMVDR beamformer is mostly ≥ 0.7 (yellow). More importantly, the higher beamforming outputs ≥ 0.9 (red) are fairly concentrated around the source location (Figure 5.16b). Obviously, it provides a relatively reliable estimate of the source location in this more complex urban environment, i.e. the localization error is less than 4 *m*.

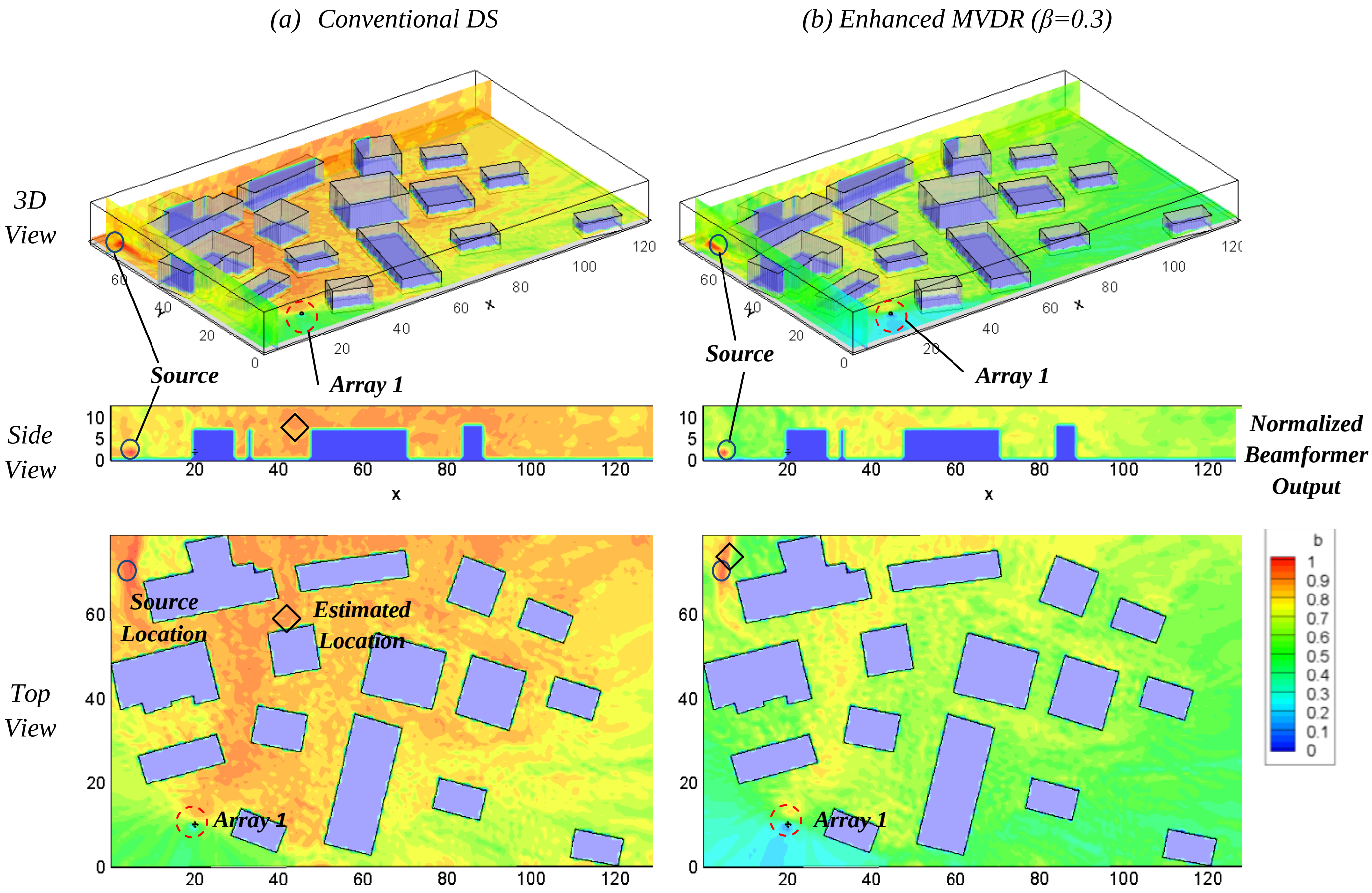


Figure 5.16: Localization maps for Fort Benning urban environment model with 5dB SNR noise using (a) conventional DS and (b) EMVDR with β=0.3 beamformers. Localization error is (a) 51.5*m* and (b) 4.2*m*.

Figure 5.17 shows the results for the EMVDR beamformer using higher adjustment factor (β = 0.7) versus the conventional DS beamformer. The EMVDR beamformer (β = 0.7) shows the better resolution, resulting in more accurate localization, i.e. the localization error decreases from 4 *m* to 2 *m* (Compare Figure 5.16b and 5.17b).

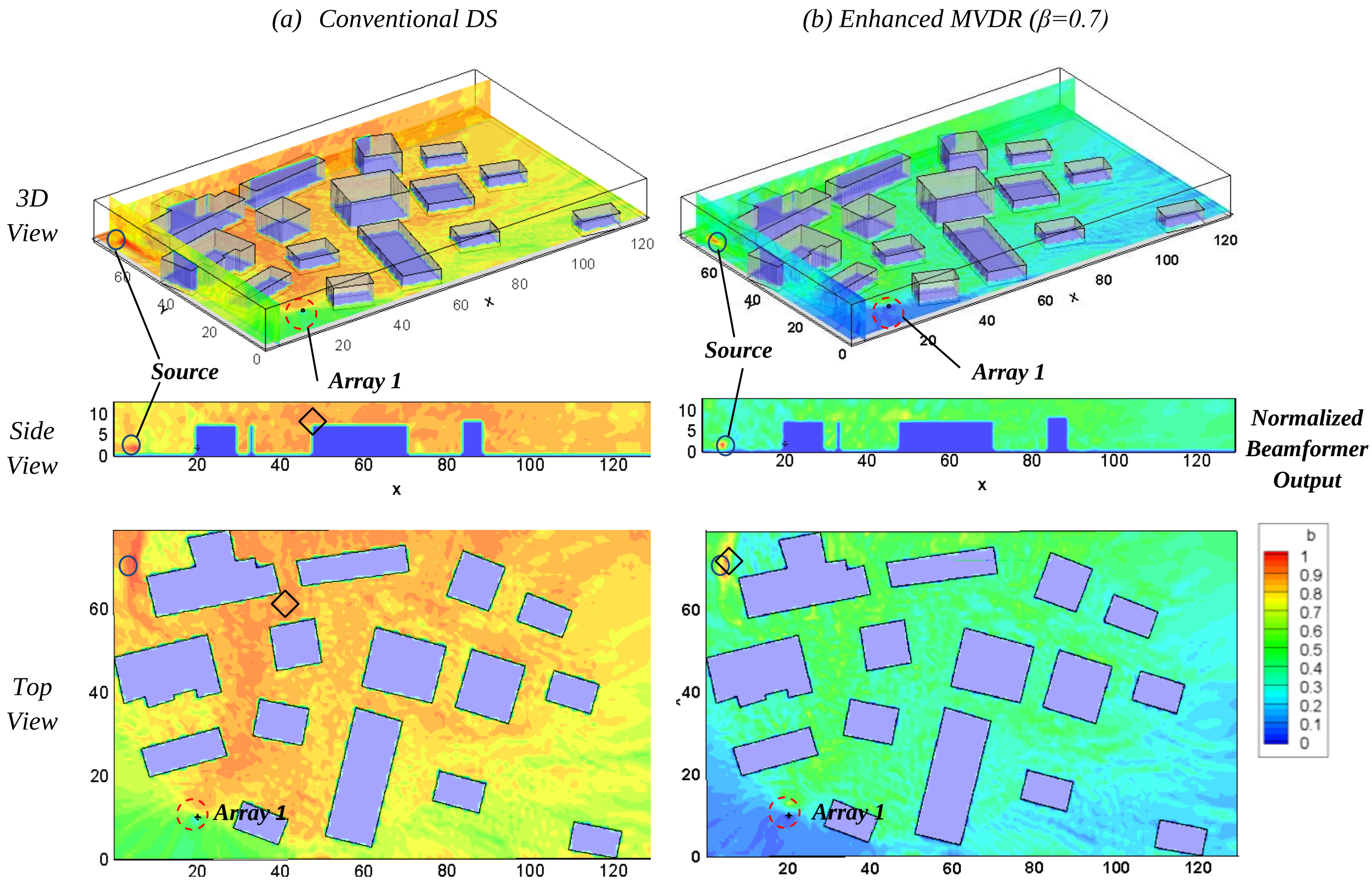


Figure 5.17: Localization maps for Fort Benning urban environment model with 5dB SNR noise using (a) conventional DS (b) EMVDR with β=0.7 beamformers. Localization error is (a) 51.5*m* and (b) 2.4*m*.

Figure 5.18 compares the results for the single and two array configurations for the conventional DS and the EMVDR beamformer (β = 0.3), respectively. Similarly, the results show that the additional array improves the localization performance due to its closer location from the source. For example, the localization error of the conventional DS beamformer decreases from around 50 *m* using a single array to 18 *m* with two arrays (Figure 5.18a and 5.18b), which is still poor. On the other hand, the localization error of the EMVDR beamformer with two arrays is reduced to less than 2 *m* from 4 *m* using a single array.

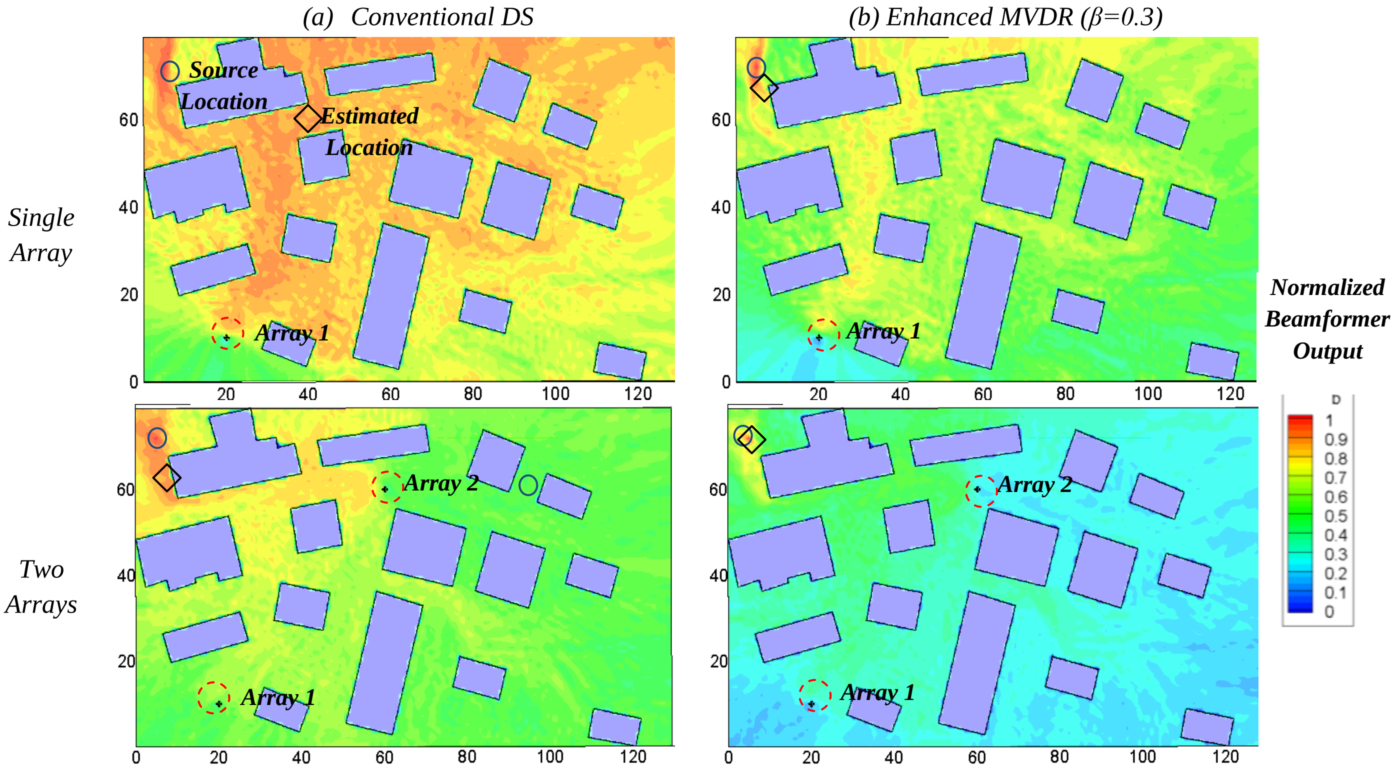


Figure 5.18: Localization maps (top-view) for Fort Benning model using the EMVDR with (a) conventional DS and (b) β=0.3 beamformers with one and two arrays whose SNRs are 5dB and 8dB, respectively. Localization error using two arrays is reduced to (a) 18.2*m* and (b) 1.5*m*, respectively.

Similar to the previous case, Figure 5.19 shows the localization performance for (a) the single array and (b) two-array configuration to compare their performance. Note that the average SNR at array 2 is about 3dB higher than the average SNR at array 1 due to the relative distance from the source. Basically, the results in Figure 5.19 show the same trends. It is shown that only the EMVDR beamformer is capable of localizing the source accurately with a single array, i.e. localization error is less than 2 m for SNR > 10 dB. The localization performance of the EMVDR at low SNRs is fairly poor, resulting in high localization error. However, this can be improved by the additional array as shown in Figure 5.19b. For the EMVDR beamformer (β = 0.3), the localization error goes below 5 *m* even at very low SNR.

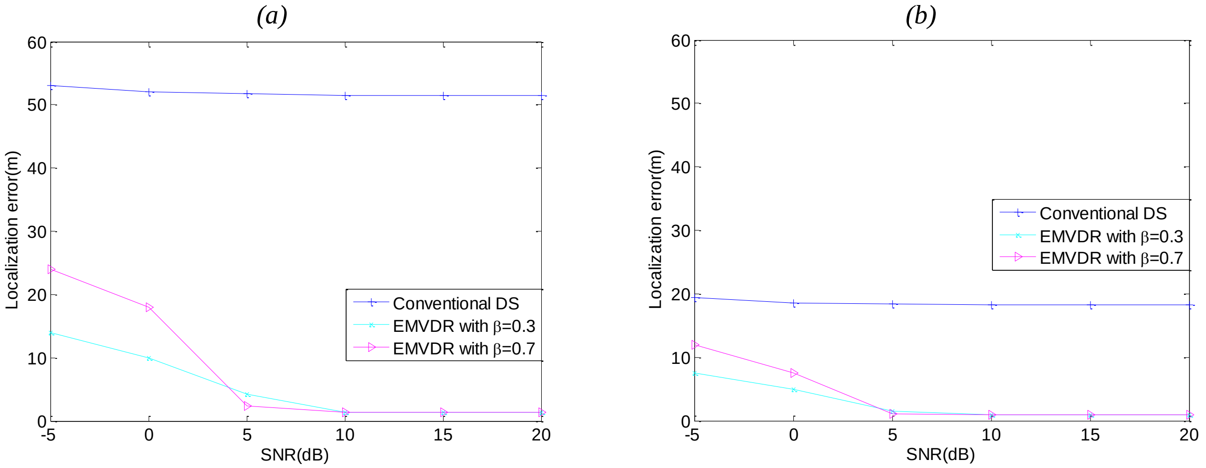


Figure 5.19: Localization error versus SNR at array 1 for 3D Fort Benning urban model using $\hat{b}_{cut-off} = 0.8$: (a) single array and (b) two array configuration.

c. Additional Results for the Fort Benning Urban Model:

It is important to note that most results so far are obtained using single and two arrays. Additional results with the number of arrays ranging from one to four arrays are provided in Appendix C.1. The locations of the additional arrays 3 and 4 are (100, 20, 2) and (120, 70, 2) *m*, respectively. Given the source location at (5, 72, 2) *m*, the average SNRs for array 1, 2, 3 and 4 are 5, 8, 2 and 0 dB, respectively. There are 15 possible combinations of arrays, i.e. 4 using single arrays, 6 using two arrays, 4 using 3 arrays, and 1 using all four arrays. Hence, 15 localization maps corresponding to the array combination are presented.

To summarize, the results in Appendix C.1 show that the localization capability depends on the amount of clutter between the source and the array. The array nearest to the source yields the best results and the localization error increases as the array is farther away from the source. In addition, the results for multiple arrays show that the localization performance is strongly dominated by the closest array to the source, as expected. It is also shown that additional arrays do not always improve the overall localization performance.

Based on the results presented in Appendix C.1, the localization error for increasing number of arrays from one to four is presented in Figure 5.20. Since there are several possible array combinations, the localization error for all possible combinations is computed and the average and range (i.e. maximum and minimum) are determined. For example, the localization error for two arrays case is computed for all possible six combinations. The average (5 *m*) is then plotted using the open circle symbol while the bar indicated the maximum and the minimum of the localization errors. As expected, the results in Figure 5.20 show that the averaged localization error decreases as the number of arrays increases. The localization error for one array case is around 15 *m* whereas the localization error for three array case decreases to less than 1 *m*. No significant improvement is observed over the three array case. The figure also shows a reduction in the uncertainty of the estimate. For example, the difference between the maximum and the minimum for one array is approximately 35 *m* as compared to 1.5 *m* for three-array case.

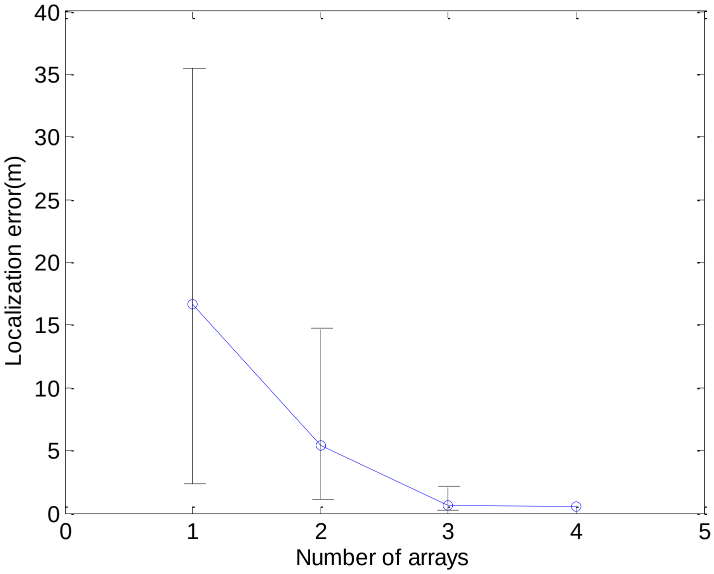


Figure 5.20: Localization error (averaged and range) for increasing number of arrays in the Fort Benning urban model using EMVDR (β= 0.7) beamformer whose SNRs are approximately 5 dB (array 1), 8 dB (array 2), 2 dB (array 3) and 0dB (array 4), respectively.

The results for different source position can be found in Appendix C.2. Basically, the results show the same trend as for the case of the original source position. The main observation of this study is that there is substantial benefit by deploying multiple arrays in the environment.

5.3 Effects of Uncertainty

In previous simulations, it was assumed perfect knowledge of the acoustic environment (transfer functions) to implement the beamformers. Naturally, localization performance of the beamforming approaches is correlated with the quality of a prior knowledge of the environment to accurately compute the steering vectors. However, there are always some uncertainties between the numerical model and the actual environment. Therefore, it is important to study the sensitivity of the beamformers to errors in the model. The main objective of this uncertainty study is to determine the severity of this problem and how to overcome it. In this section, the uncertainties investigated are presented first. As before, microphone self-noise is added for more realistic results. The effect on the performance of the beamformers is then assessed and the results are discussed. It is also investigated if additional arrays can alleviate the problem.

5.3.1 Uncertainty Parameters

It is important to note that this uncertainty study does not investigate all possible sources of model errors. For example, the effect of wind is not considered in this study due to the limitation of the FDTD code to incorporate this effect into the modeling. Different positions of the buildings can be considered as a possible uncertainty in the modeling. However, current technologies enables to find the position and size of the buildings accurately.

In this study, two uncertainties are investigated: (a) temperature gradient and (b) unmodeled object (i.e. uncertainty in geometry). Firstly, temperature gradient is a common uncertainty in realistic 3D environments. It is difficult to measure and it changes through the day/night.

Secondly, in urban environments, there are always unpredictable geometric changes on the terrain such as movement of vehicles, new buildings not yet mapped, demolition of the buildings and so forth. Hence, it is important to investigate how the uncertainty in geometry due to unmodeled objects affects the localization performance. It is also noteworthy that both uncertainty cases have not been investigated in any previous work.

a. *Temperature Gradient*

The first uncertainty considered is a temperature gradient such as inversion (lapse), i.e. increase (decrease) in temperature with altitude. So basically, a uniform temperature (zero gradients) is assumed in the numerical model while there is a gradient (lapse or inversion) in the real environment. Since the speed of sound is a function of temperature, the temperature gradient uncertainty implies a spatial distribution of the speed of sound with height.

Based on observations from the lowest 100 *m* of the atmosphere, the air can be divided in two distinct parts: a layer above the ground in which the rate of change of the temperature gradient is log-linear, and a second layer in which the temperature gradient is constant with height (Geiger et al., 2009). The height of the first layer is a minimum of 4 m in winter and 30 to 40 m in summer. The second layer is a few hundred meters in height. The temperature profile used follows the equation

$$T_2 = T_1 + a \ln\left(\frac{z_2}{z_1}\right) \tag{5.1}$$

where T_1 and T_2 are the absolute temperature in Kelvin at two different height z_1 and z_2 respectively and a is the profile constant.

This expression requires the knowledge of the temperature at two different heights. Usually, the values used are the surface temperature (just above the ground) and the air temperature (usually 2 m above the ground). Here, 50˚C to 30˚C is chosen as a maximum and minimum temperature in the profile, respectively as shown in Figure 5.21 and the corresponding profile constant *a* is approximately 8.69. This temperature gradient is rather extreme. However, it was chosen because, as it will be shown, the beamformer is not very sensitive to common temperature gradients.

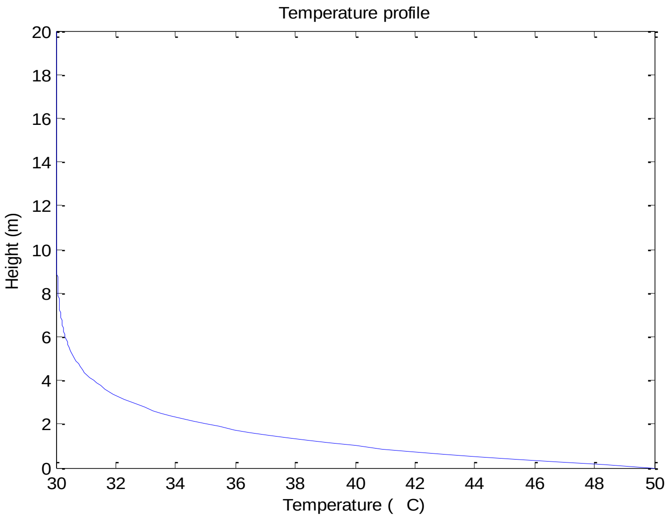


Figure 5.21: Temperature profiles for the temperature gradient from 50 ˚ C to 30 ˚ C.

b. *Unmodeled Object*

In urban environments, objects or structures can be mistakenly added to or omitted from the numerical model due to unpredictable changes on the environment (e.g. movement of vehicles, demolition of buildings, and so forth). This represents an uncertainty in the geometry. To study this type of uncertainty, an object with 10 *m* × 2.5 *m* × 3 *m* dimensions (such as a large bus or truck) is added to the model (See Figure 5.22).

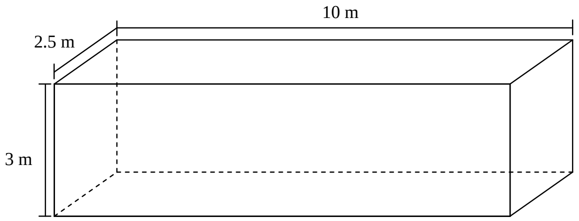


Figure 5.22: Unmodeled object geometry (e.g. a bus).

Originally, four positions were chosen for the unmodeled object: 2 positions near the array and 2 positions near the source. These positions were selected to be very close to the source or the array so that the influence of the object on the transfer function between the source and the array can be maximized.

5.3.2 Performance Degradation due to Uncertainty

To evaluate the performance degradation due to uncertainty, the simplified urban environment model without any uncertainty is considered as the baseline and used to compute the transfer functions. Errors due to uncertainties (temperature gradient or the un-modeled object) were then introduced as a modification of the baseline model. In other word, the beamformers are implemented using the information from the baseline model but the actual acoustic field recorded by the microphones is the one from the perturbed model as shown in Figure 5.23.

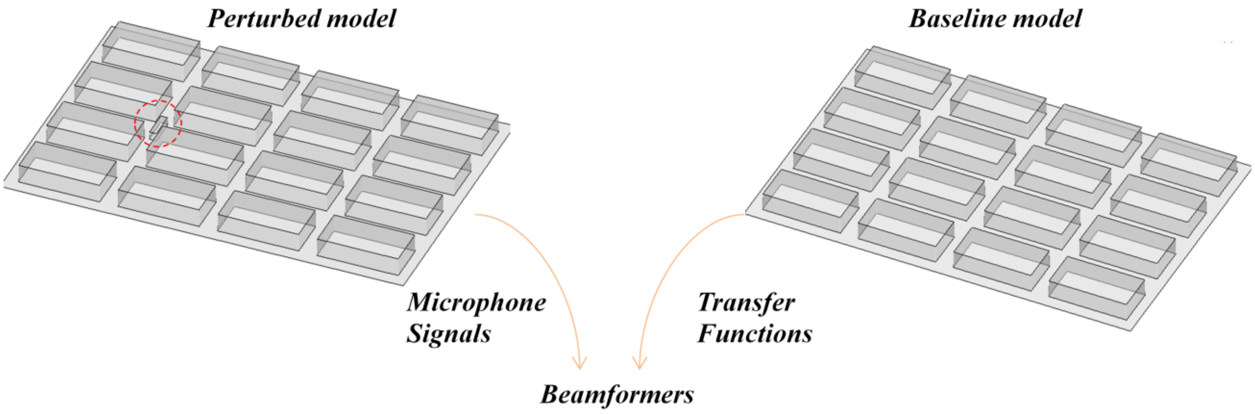
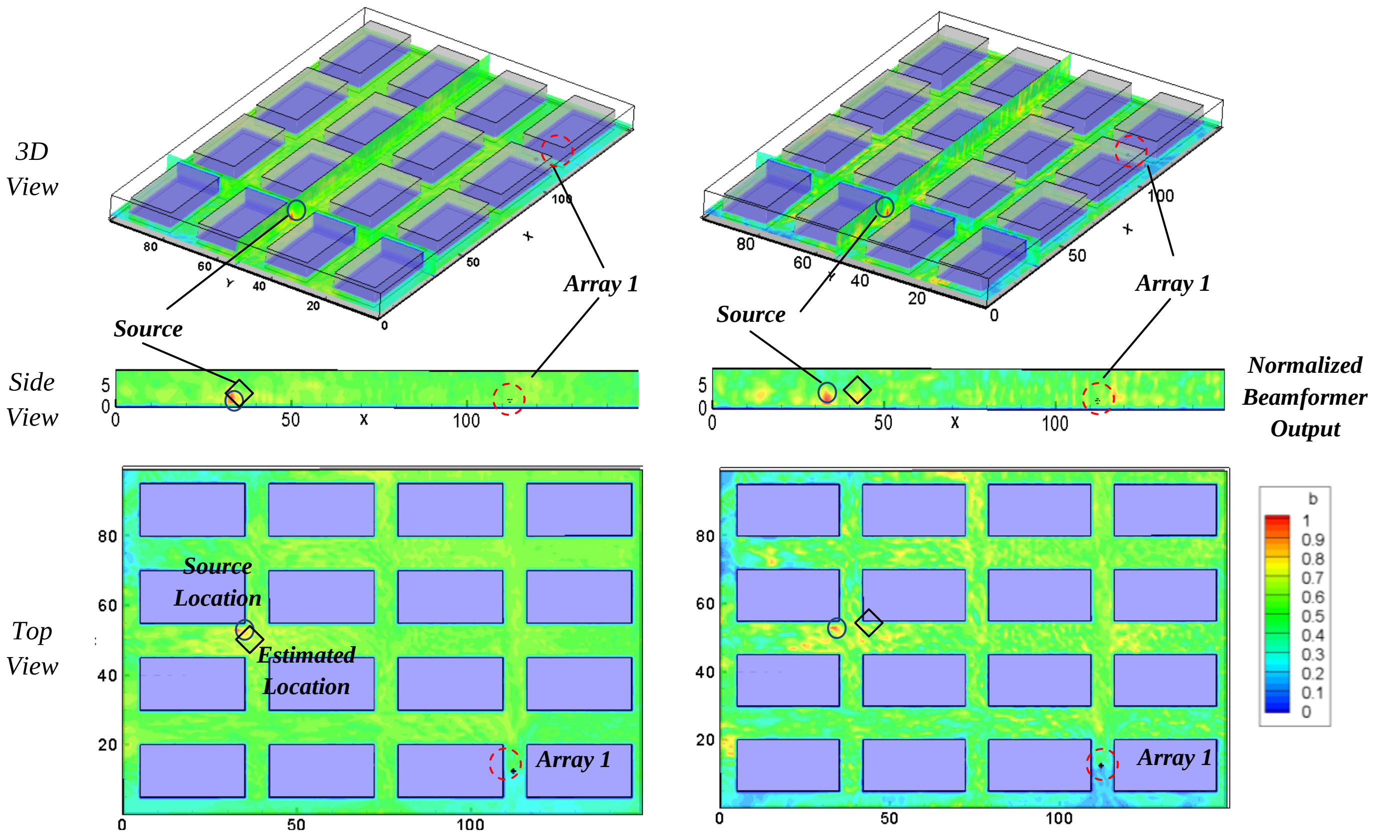


Figure 5.23: Illustration of the uncertainty study with the unmodeled object.

a. *Temperature Gradient*

To be consistent with the previous results, the array and source are located at the same positions. The average SNRs for array 1 is 5 dB. The baseline case without any uncertainty (i.e. 20℃ uniform temperatures) is presented in Figure 5.24a for comparison purpose. Note that the results in Figure 5.24a are identical to the ones in Figure 5.12b. On the other hand, Figure 5.24b show the results for the case of temperature gradient present in the actual environment but not considered by the beamformer. Compared to the results in Figure 5.24a, it is clear that the EMVDR beamformer still shows relatively good performance, i.e. the un-modeled temperature gradient resulted in minor degradation of the performance. The minor degradation is noticeable as the localization error increases from 2 *m* without uncertainty to more than 4 *m* with uncertainty.


(a) Without Uncertainty

(b) With Uncertainty

Figure 5.24: Localization maps for 3D simplified urban model with 5dB SNR noise using the EMVDR with β=0.3: (a) no uncertainty and (b) 50 to 30°C gradient uncertainty. Localization error is (a) 2.1*m* and (b) 4.5*m*.

In order to compare the source localization performance with and without temperature gradient uncertainty, the localization errors of the beamformers for different SNRs are computed and presented in Figure 5.25. The results show that the overall performance of the beamformers with uncertainty (dotted line) degrades due to the inaccuracy of the estimated transfer functions. For instance, the localization error of the conventional DS increases by 5 *m* due to the uncertainty. For the EMVDR, the uncertainty causes the localization error to be 3 *m* higher for SNR > 5 dB and has more detrimental effect as SNR decreases. However, the EMVDR still works much better than the conventional DS for all SNRs.

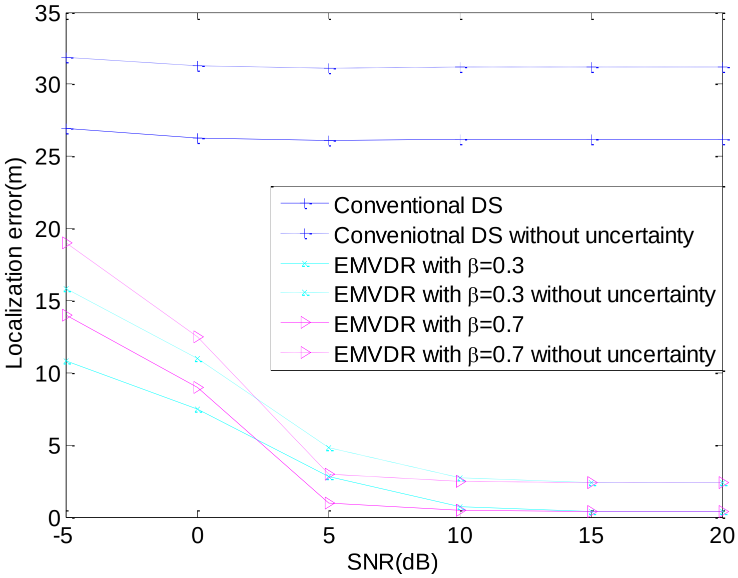


Figure 5.25: Localization error versus SNR for 3D simplified urban model using $\hat{b}_{cut-off} = 0.8$: without uncertainty and with 50 to 30°C gradient uncertainty.

It is important to note that minor degradation due to the temperature gradient uncertainty is inevitable. However, the negative impact on the localization performance turns out to be rather insignificant given that the temperature profile used is a relatively extreme case.

b. Unmodeled Object

Figures 5.26 shows the localization maps of the EMVDR beamformer with 5dB SNR noise for the case of the un-modeled object uncertainty at two different positions (showed in the figure as a dark blue rectangle). In this case, the unmodeled object is placed very close to the source. In general, it is observed that the presence of the un-modeled object near the source has a significant negative effect on the beamformer performance. For both cases, the localization error increases to at least 10 *m*. This observation is obvious by comparing these maps to the baseline case in the first column in Figure 5.24. The degradation is noticeable for the case of position 1 where a strong spurious source appears at the opposite side of the building likely due to the reflections caused by the object which is right below the source in the direction of the array. The case for position 2 is even worse showing a lot more ambiguity over the entire domain, i.e. the beamforming output is much higher. This is most likely because the object in position 2 blocks larger amount of the propagation from the source to the array.

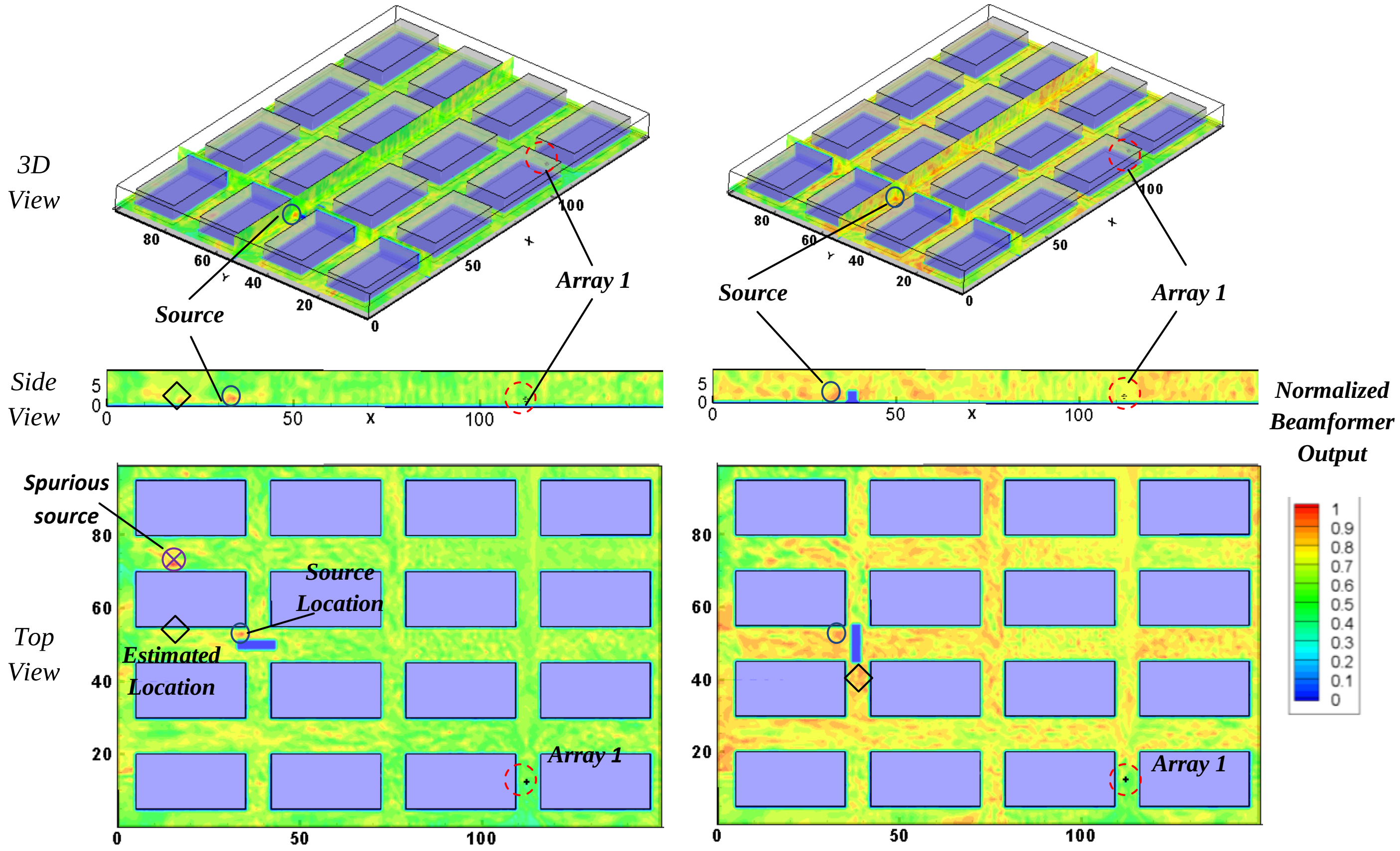

(a) Position 1   *(b) Position 2*

Figure 5.26: Localization maps (top-view) for 3D simplified urban model with 5dB SNR noise using the EMVDR with β=0.3 for unmodeled object in (a) position 1 and (b) position 2. Localization error is (a) 10.9*m* and (b) 10.3*m*.

It is noteworthy that the case where the object is very close to the array would result in similar results. Hence, the results for the object close to the array are not presented here.

The same type of the investigation is then performed for a different source location to see the effect of the un-modeled uncertainty when it is not maximized, i.e. the un-modeled object does not directly hinder the sound field between the source and the array. To this end, the source is chosen to be at (20, 25, 2) *m* while the array position is the same. For this study, the localization maps of the EMVDR beamformer with and without the un-modeled uncertainty are presented in Figure 5.27a and 5.27b, respectively. Again 5dB SNR microphone self-noise is added for more realistic results. The results in Figure 5.27b show that the EMVDR beamformer provides accurate localization capability with the unmodeled object not in the proximity of the source, i.e. the estimated source location is less than 2 *m* from the actual source placement. The comparison between Figure 5.27a and 5.27b shows that the presence of the object far from the source has negligible effect on the beamformer performance, i.e. the results of Figure 5.27a and 5.27b are similar. This implies that the performance degradation due to the uncertainty in geometry depends heavily on the relative location of the object given the source-array configuration.

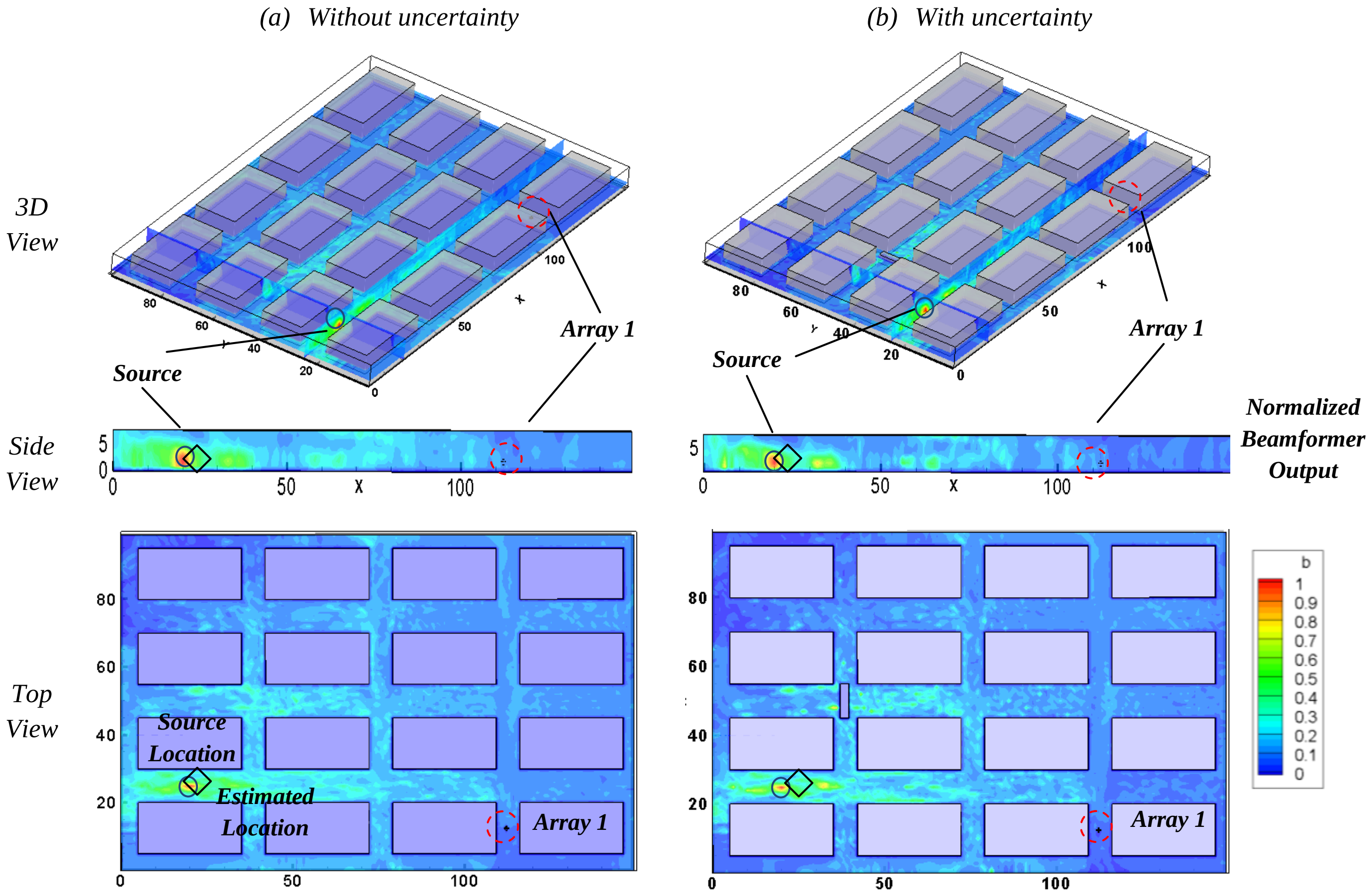


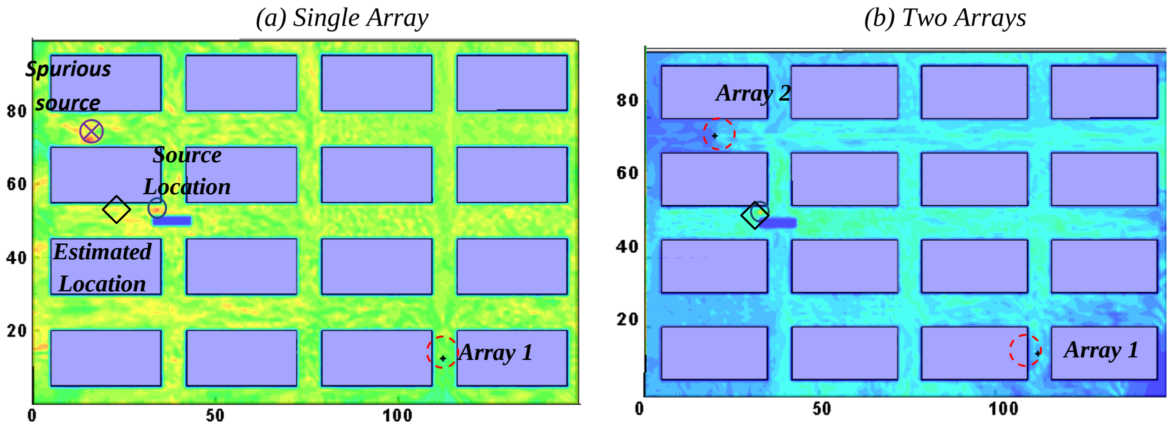
Figure 5.27: Localization maps (top-view) for 3D simplified urban model with 5dB SNR noise using the EMVDR with β=0.7 for (a) no uncertainty and (b) unmodeled object in position 2. Localization error is (a) 1.4*m* and (b) 1.9*m*.

c. Multiple Arrays to Minimize Effect of Uncertainties

Previous simulation results have shown that the addition of the second array improved the performance of the beamformer assuming perfect knowledge of the environment (see Figure 5.15 and 5.19). It is important to find out if multiple arrays can help in compensating for lack of knowledge in the model.

Here, two arrays configuration is investigated to assess if multiple arrays can overcome the uncertainties in the model used. Since the detrimental effects of the temperature gradient uncertainty are inconsequential, this study is focused on the uncertainty of the unmodeled object.

Figure 5.28 compares the single and two array cases corresponding to the unmodeled object at position 1. As shown in the figure, the negative effects of the uncertainty are virtually compensated by the second array. Since the spurious source was found near the second array, it disappears completely using two arrays. Compared with the baseline case without uncertainty (See Figure 5.14a and 14b), the improvement obtained by the two array configuration is significant, i.e. the localization error decreases from 10 *m* to only 2 *m*.


(a) Single Array

(b) Two Arrays

Figure 5.28: Localization maps (top-view) for simplified urban environment model with unmodeled object uncertainty (position 1) using the EMVDR with β= 0.3 beamformer with one and two arrays whose SNRs are 5dB and 12dB, respectively. Localization error is (a) 10.9*m* and (b) 1.8*m*.

Figure 5.29 compares the source localization performance with the unmodeled uncertainty at position 1 for (a) single and (b) two array configurations. The results in Figure 5.29a show the performance degradation due to the uncertainty for the single array. For conventional DS, the localization error increases by around 10 *m*, from 27 to 37 *m*. The localization error using the EMVDR beamformer (β = 0.3) also increases up to 7 *m* at very low SNR. Similarly, the results in Figure 5.29b show the performance degradation due the uncertainty for the two arrays. The localization error does not increase as much as for the single array case. For instance, the localization error of conventional DS increases by only 3 *m*, from 6 to 9 *m*. The performance degradation of the EMVDR beamformer is even smaller.

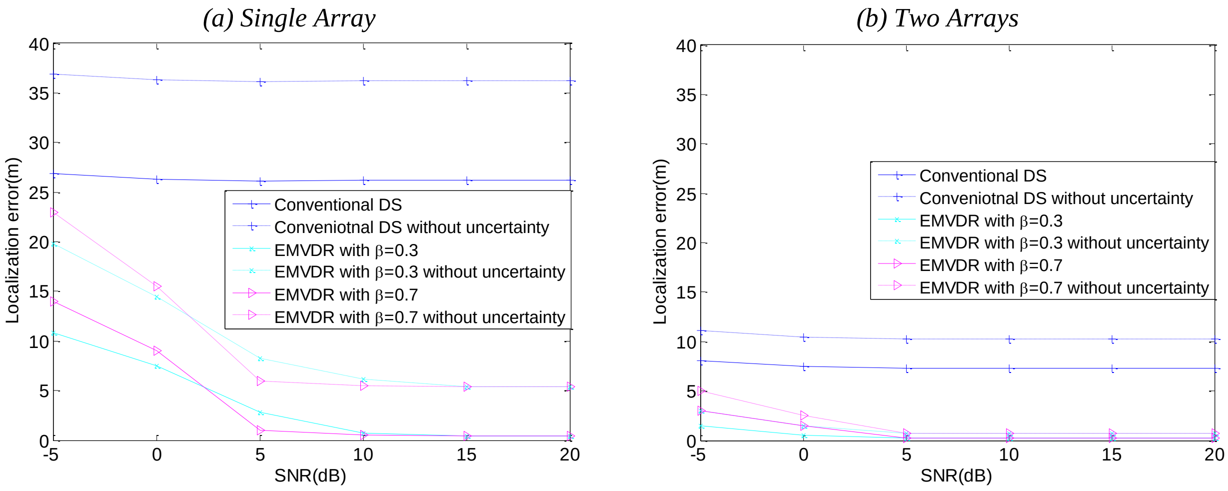


Figure 5.29: Localization error versus SNR for 3D simplified urban model using $\hat{b}_{cut-off} = 0.8$ for an un-modeled object uncertainty at position 1: (a) single array and (b) two arrays configuration.

From these results, the key observation is that two-array configuration can help significantly to compensate the detrimental effects of uncertainty (unmodeled object).

Chapter 6 Localization technique: Acoustic Fingerprinting

This chapter discusses a source localization technique using microphones dispersed in a cluttered environment such as urban centers in conjunction with a low fidelity acoustic model. The approach implemented here is to integrate a fingerprinting method with the energy-based acoustic model of the environment.

The fingerprinting approach has been used as an alternative solution to the source localization problem in RF signals (LaMarca et al., 2005; Ferris et al., 2006). However, it has not been investigated in acoustic source localization combined with the numerical model of the environments.

The source localization system using this approach is described in section 6.1. Mathematical formulation on integration of the approach with the energy method is then provided in section 6.2. This proposed approach is validated using the two large 3D realistic complex urban models (Figure 3.6 and 3.7). Since it requires several single microphones dispersed in the environment, the relationship between the localization accuracy and the number of dispersed microphones is also investigated. Finally, the effect of the accuracy of the model is addressed by comparing the results based on the estimated fingerprints (energy-based method) to the results by using the exact fingerprints (FDTD).

6.1 Localization Approach

Figure 6.1 depicts the localization system for a sound source in a complex urban environment using multiple dispersed microphones. Figure 6.1a represents an example of a typical scenario where an acoustic event (such as a gunshot) occurs at an unknown position in the environment (solid circle). Due to the either impulsive or continuous noise produced by the source, the sound would propagate through the environment and be recorded by the microphones (dotted circles) in Figure 6.1c. These signals are then transformed to the frequency domain by taking the Fast Fourier Transform. In this application, the magnitudes of the spectrums of the microphone signals are the measured signal strengths (Figure 6.1d). The fingerprinting method requires estimating the signal strength (spectrum) at the microphones due to a source at grid points (Figure 6.1e). The estimated spectrums obtained using a predictive propagation model become the reference fingerprints and are saved in the database (Figure 6.1f). Next, the best match between the currently observed (or measured) signal strength and the reference fingerprints in the database is found (Figure 6.1g). Finally, the coordinates associated with the fingerprint that provides the smallest Euclidean distance is returned as the estimate of the source location (Figure 6.1h).

The predictive acoustic propagation model considered here is based on the 3D diffusion equation (i.e. the energy-based method) as mentioned in Chapter 3. The major advantage of this model is that it allows fast computation of the sound field in large and complex urban environments. Therefore, this propagation model is used in the present study to generate the reference fingerprints by efficiently predicting the SPL distribution in a given urban environment. Due to fast computation of the energy model, relatively dense coverage over the entire domain can be achieved.

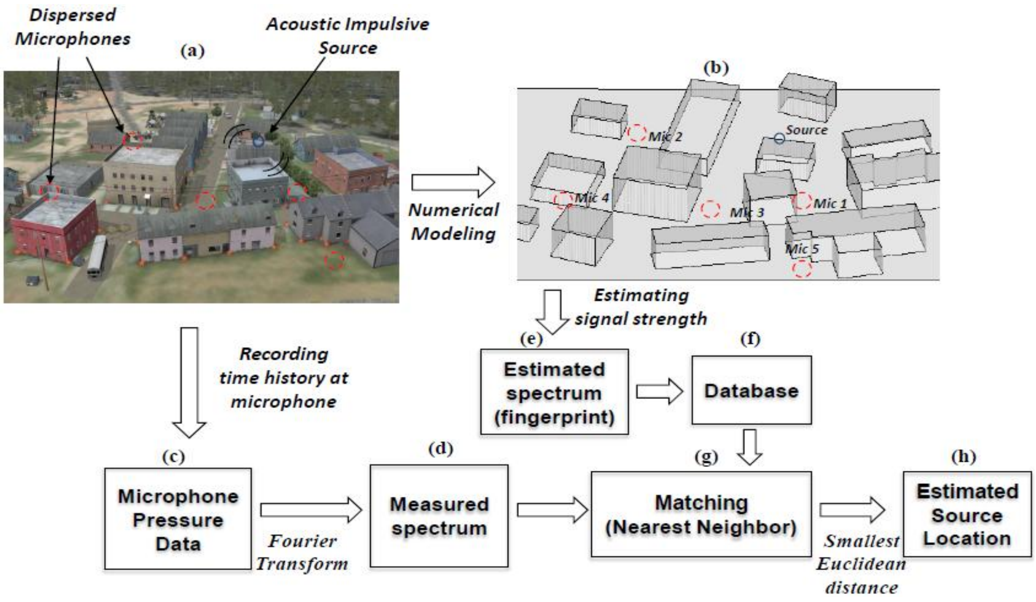


Figure 6.1: Schematic of localization system using fingerprinting method.

6.2 Source Localization Mathematical Formulation

Figure 6.2 illustrates the Fort Benning urban environment with *M* microphones dispersed throughout the environment. A noise source generates an impulsive noise that propagates through the environment and is recorded by the microphones.

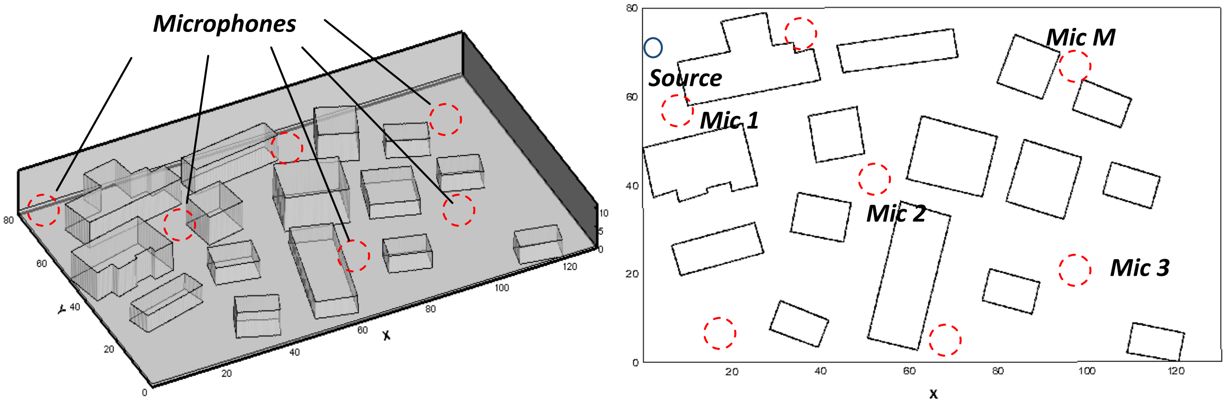


Figure 6.2: Illustration of an urban environment model with *M* microphones dispersed throughout the environment. 3D view (left) and top-view (right).

Since the source is a Ricker pulse with center frequency of 100Hz, most of the source information is concentrated between 50 and 200 Hz. Hence, this frequency range is considered for frequency domain source localization. The measured signal strength needed in the fingerprinting method is the microphone signal spectrum magnitude. Since the energy method is useful for predictions in frequency bands, the narrowband spectrum is used to compute the spectrum in wider frequency bands such as 1/3 octave bands. Within this range, there are six 1/3 octave bands with center frequencies between 63 and 200Hz. The magnitude of spectrum of the source for all six bands is presented in Figure 6.3. As observed, the band whose central frequency is 160Hz shows the maximum amount of energy among them. Although the spectrum concentrates around the 100Hz (see Figure 3.8b), there are more spectral lines in this band, resulting in greater sound pressure level.

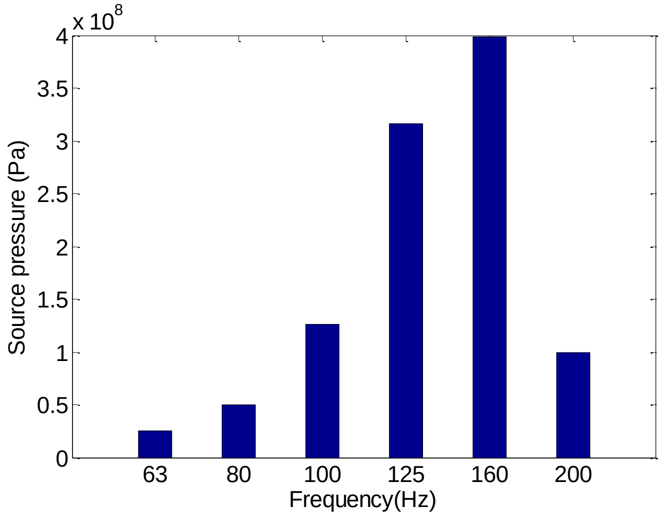


Figure 6.3: Magnitude of spectrum of the Ricker waveform in 1/3 octave bands.

Figure 6.4 illustrates the microphone noise signals at different positions and the corresponding spectrums in 1/3 octave bands. The first microphone position is chosen right next to the source (1m to the left). Hence, the source signal arrives to the microphone with a very small delay as shown in Figure 6.4a. Note that the magnitude of spectrum looks very much the same as the results in Figure 6.3. The positions of the second and the third microphones are further away from the source. As observed in Figure 6.4b and 6.4c, the source signal is distorted due to the environment and arrives to the microphones at different times due to the different propagation path lengths. It is clear that the original spectral content of the source is distorted by the presence of the cluttered environment (Compare Figure 6.3 and Figure 6.4b or 6.4c).

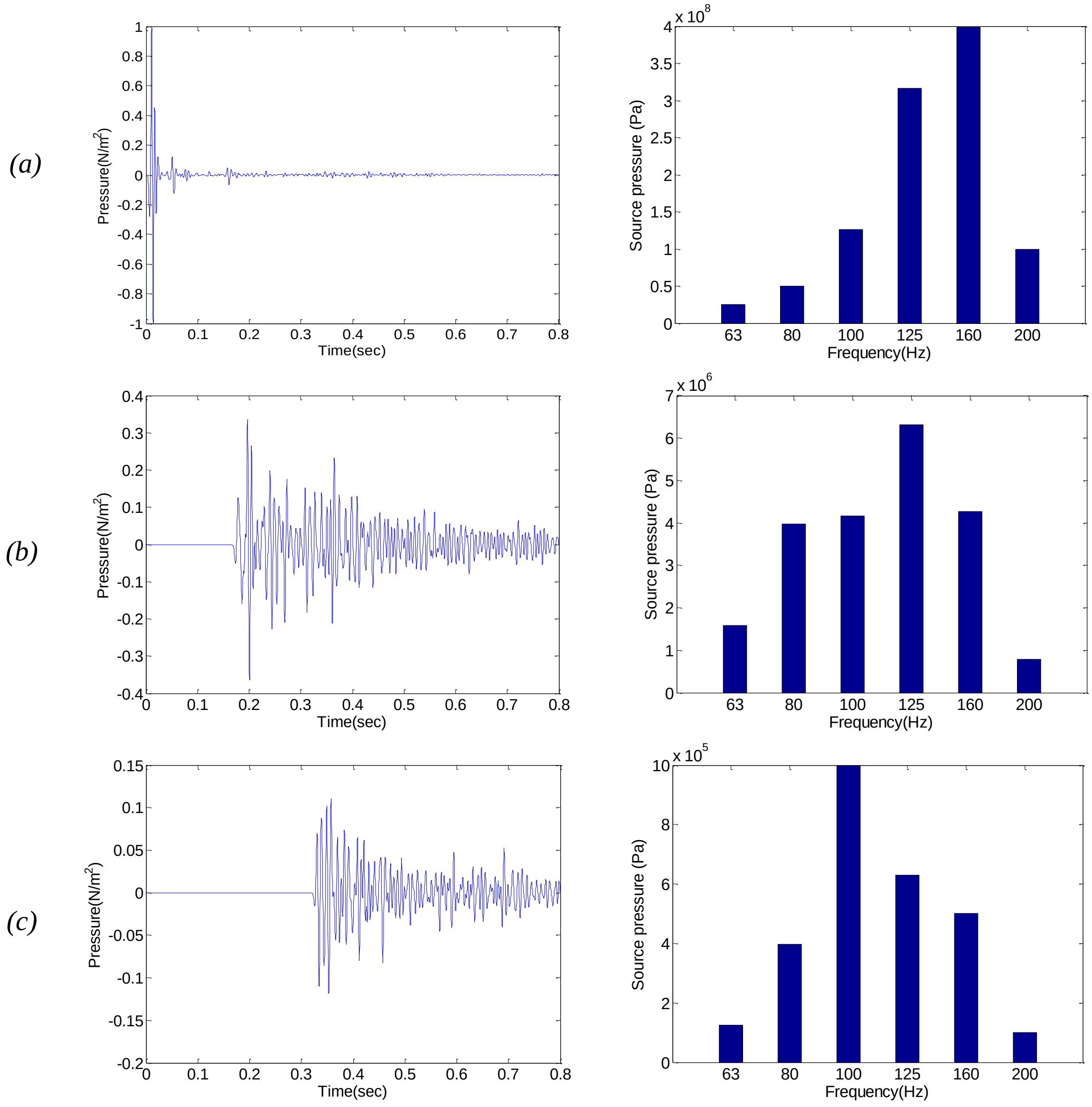


Figure 6.4: Time history and magnitude of the spectrum of the (a) near-source microphone signal and microphone signals at (b) position 2 and (c) position 3. Source location is (5, 72, 2) *m*.

As implemented here, the magnitude of the spectrum of the *m*-th microphone signal is written in vector form as,

$$\left\{P_m\right\} = \left\{P_m\left(f_1\right), \cdots, P_m\left(f_b\right), \cdots, P_m\left(f_B\right)\right\}^T \quad m = 1, 2, \cdots, M \tag{6.1}$$

where $f_b, b = 1, 2, \cdots, B$ is the *b*-th band center frequency. Each term $P_m\left(f_b\right)$ in the vector represents the mean-square-value (msv) of the acoustic pressure in the band, i.e. $p_{msv}\left(f_b\right) = p_{rms}^2\left(f_b\right)$.

The energy-based acoustic propagation model is then used to predict the signal strength at each microphone due to a source located at the grid points (Figure 6.1e). The output of the acoustic model is the magnitude of the spectrum at the microphone locations due a steady noise source at the j^{th} grid point at the same frequency bands f_b. That is

$$\left\{P_{mj}\right\} = \left\{P_{mj}\left(f_1\right), \cdots, P_{mj}\left(f_b\right), \cdots, P_{mj}\left(f_B\right)\right\}^T \quad m = 1, 2, \cdots, M; \quad j = 1, 2, \cdots, J \tag{6.2}$$

where $P_{mj}\left(f_b\right)$ is the predicted spectrum magnitude at the m^{th} microphone, for the b^{th} frequency band, and due to a source at the j^{th} grid point. These estimated microphone spectrums form the fingerprint in the fingerprinting method. The energy method is used to generate a database of spectrums for all grid points, i.e. a database of fingerprints (Figure 6.1f).

Figure 6.5 illustrates an example of the estimated spectrums in 1/3 octave bands at each microphone due to a source located at (5, 72, 2) *m* and compares it to the measured spectrums from Figure 6.4. As expected, the estimated spectrums at each microphone decrease as the distance from the source to the microphone increases. Note that the estimated spectrums at each microphone are the same for all the bands because the same amount of energy was inserted in each frequency band and the perfectly reflective boundaries were assumed. It is found that the spectral content of the source is shown only in the measured spectrum.

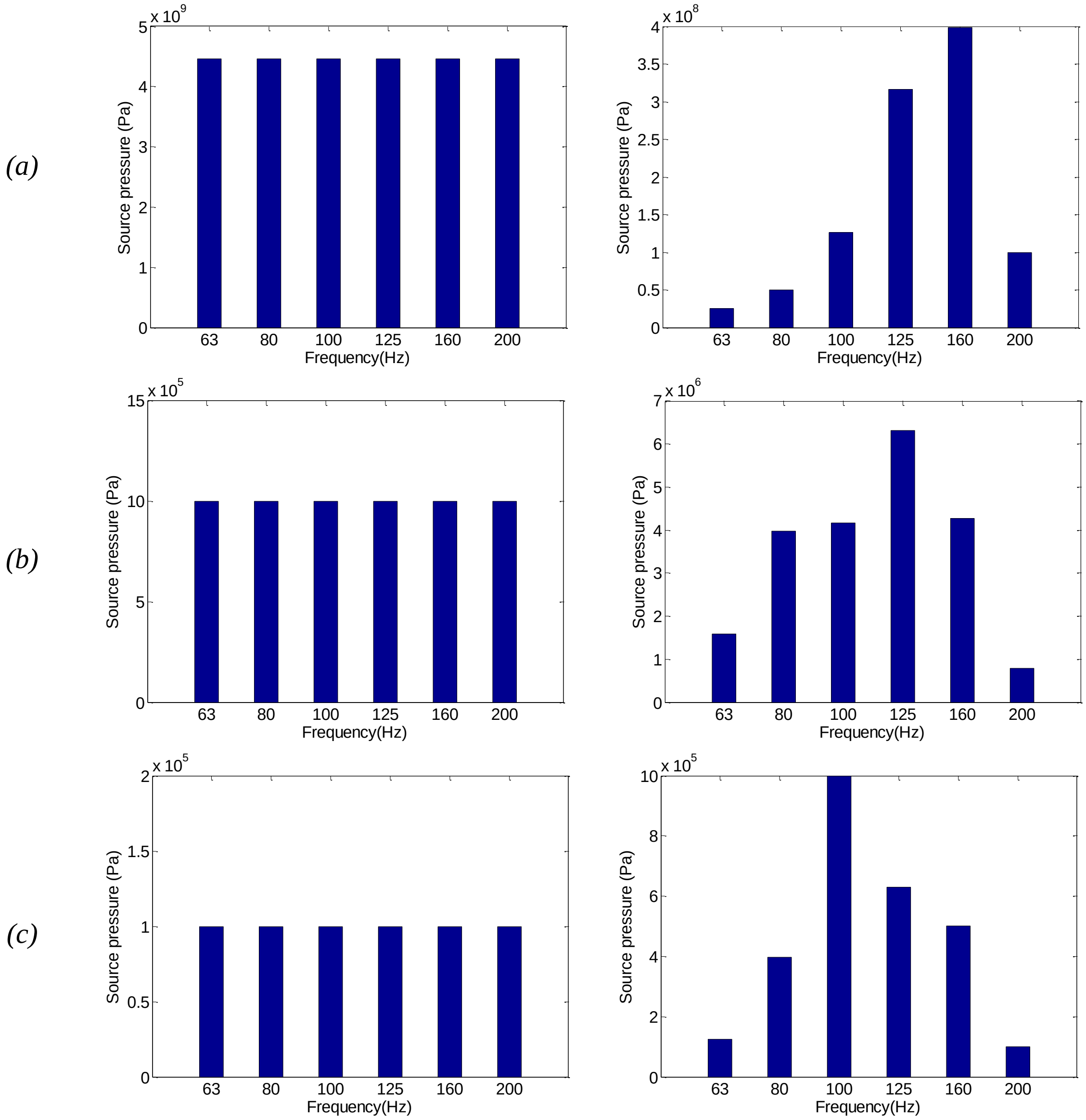


Figure 6.5: Estimated (left) and measured (right) spectrum of the (a) near-source microphone signal and microphone signals at (b) position 2 and (c) position 3. Source location is (5, 72, 2) *m*.

Source localization in the fingerprinting method is based on comparing the measured spectrums in eq. (6.1) to the estimated ones in eq. (6.2). However, these comparisons cannot be performed because the spectral content of the noise source is unknown which is needed in the computation of eq. (6.2). This can be also observed by comparing the estimated and measured spectrums shown in Figure 6.5.

This problem can be overcome by normalizing the spectrum. In other words, only the spectral shape is considered. To this end, the measured microphone spectrums are normalized as follows

$$\left\{\hat{P}_m\right\} = \left\{\hat{P}_m\left(f_1\right),\cdots,\hat{P}_m\left(f_b\right),\cdots,\hat{P}_m\left(f_B\right)\right\}^T \quad m = 1,2,\cdots,M \tag{6.3}$$

where the normalized level in the b^{th} frequency band $\hat{P}_m\left(f_b\right)$ is given as

$$\hat{P}_m\left(f_b\right) = \frac{P_m\left(f_b\right)}{\sqrt{\sum_{m=1}^{M} P_m^2\left(f_b\right)}} \tag{6.4}$$

Since the msv of the pressure is proportional to the sound power of the source, $P_m\left(f_b\right) \propto W\left(f_b\right)$, it is evident that the normalized spectrum in eq. (6.3) is independent of the source spectral content. Similarly, the predicted spectrums in eq. (6.2) are normalized to yield

$$\left\{\hat{P}_{mj}\right\} = \left\{\hat{P}_{mj}\left(f_1\right),\cdots,\hat{P}_{mj}\left(f_b\right),\cdots,\hat{P}_{mj}\left(f_B\right)\right\}^T \quad m = 1,2,\cdots,M; \quad j = 1,2,\cdots,J \tag{6.5}$$

where

$$\hat{P}_{mj}\left(f_b\right) = \frac{P_{mj}\left(f_b\right)}{\sqrt{\sum_{m=1}^{M} P_{mj}^2\left(f_b\right)}} \tag{6.6}$$

The fingerprinting method can now be implemented by comparing the normalized measurements, $\left\{RSS\right\}$, to the predicted fingerprints, $\left\{RSS_j\right\}$.

That is,

$$\begin{aligned}\{RSS\} &= \left\{\{RSS(f_1)\}^T,\cdots,\{RSS(f_b)\}^T,\cdots,\{RSS(f_B)\}^T\right\}^T \\ \{RSS_j\} &= \left\{\{RSS_j(f_1)\}^T,\cdots,\{RSS_j(f_b)\}^T,\cdots,\{RSS_j(f_B)\}^T\right\}^T\end{aligned} \qquad j=1,2,\cdots,J \tag{6.7}$$

where

$$\begin{aligned}\{RSS(f_b)\} &= \left\{\hat{P}_1(f_b),\cdots,\hat{P}_m(f_b),\cdots,\hat{P}_M(f_b)\right\}^T \\ \{RSS_j(f_b)\} &= \left\{\hat{P}_{1j}(f_b),\cdots,\hat{P}_{mj}(f_b),\cdots,\hat{P}_{Mj}(f_b)\right\}^T\end{aligned} \qquad b=1,2,\cdots,B; \quad j=1,2,\cdots,J \tag{6.8}$$

Figure 6.6 illustrates the normalized estimated spectrums in 1/3 octave bands at each microphone due to a source located at (5, 72, 2) *m* and the normalized measured spectrums. It is shown that both estimated and measured spectrums become independent of the spectral content of the source after normalization.

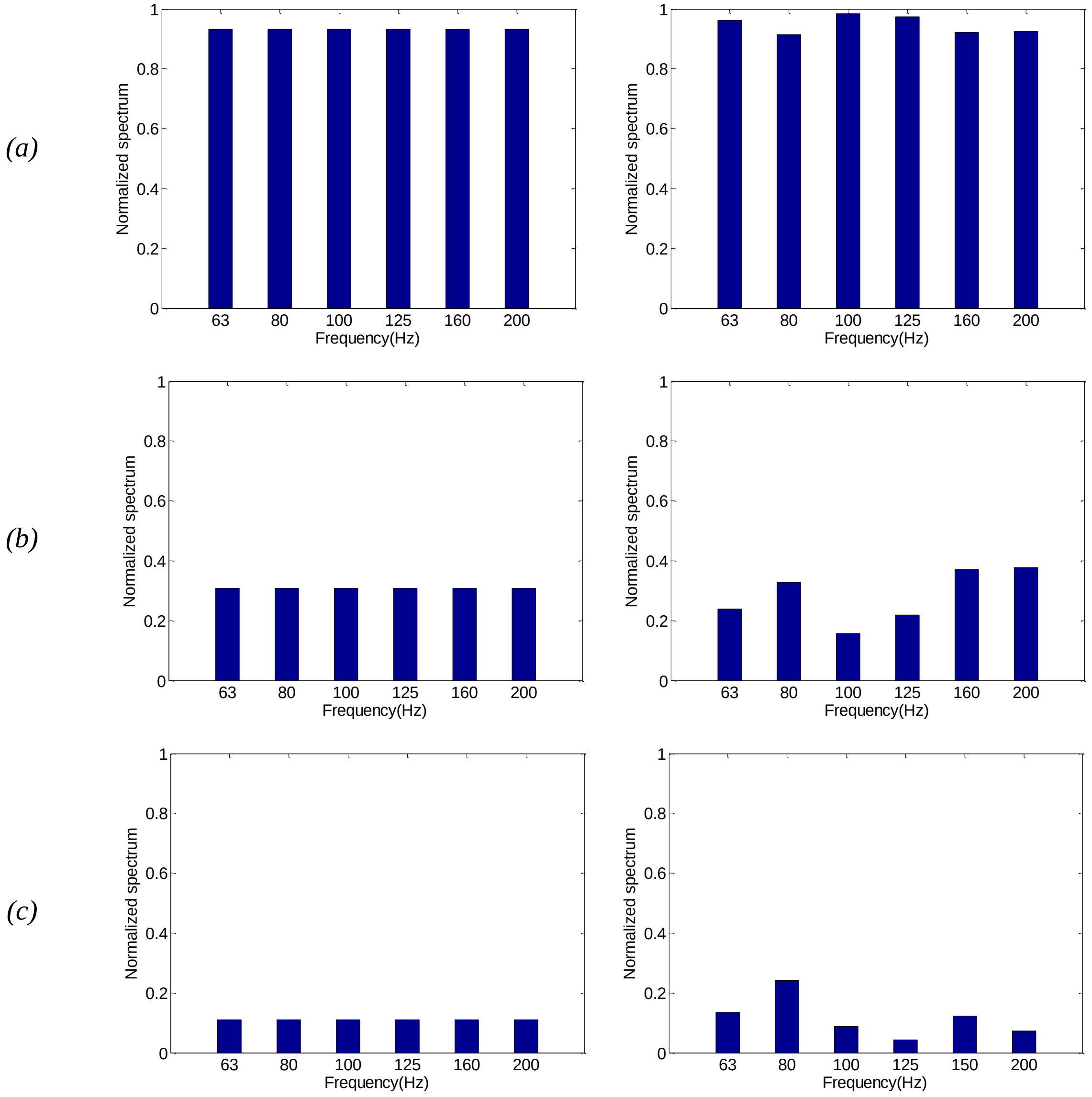


Figure 6.6: Normalized estimated (left) and measured (right) spectrum of the (a) near-source microphone signal and microphone signals at (b) position 2 and (c) position 3. Source location is (5, 72, 2) *m*.

The second issue regarding the implementation of the source localization technique is related to the type of sound source. The fingerprinting method was envisioned for steady-state RF sources. Here, the sound source can be either impulsive or continuous while the energy-based method assumes that the noise source is stationary. However, this is not an issue for a linear system and as long as the spectrum of the impulsive signal is computed consistently for all microphones. In particular, the time window used to compute the Fourier Transform (or the Fast Fourier Transform) must be of the same length for all microphones. In addition, the time window must be sufficiently long such that the impulsive signal is capture by all the microphones. Therefore, the normalized spectral shape estimated at the microphones will be independent of the character of the source, impulsive or steady-state.

Though several matching algorithms can be applied, here the Euclidean distance between the measured spectrum and each fingerprint in the database (Nearest Neighbor) is used (Figure 6.1g). This matching approach can be implemented in two different ways. In the first approach, the measured and predicted spectrum for each frequency band is used in the calculation of the Euclidean distance. That is

$$d_j\left(f_b\right)=\left|\left\{RSS\left(f_b\right)\right\}-\left\{RSS_j\left(f_b\right)\right\}\right|=\sqrt{\sum_{m=1}^{M}\left(\hat{P}_m(f_b)-\hat{P}_{mj}(f_b)\right)^2} \quad b=1,2,\cdots,B;\, j=1,2,\cdots,J \qquad (6.9)$$

The grid point that leads to the minimum $d_j\left(f_b\right)$ yields the source location. Thus, a source location estimate will result from each frequency band. To compute a single estimated source location independent of the frequency band, the result in eq. (6.6) can be average across the bands. That is

$$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j\left(f_b\right) \quad j=1,2,\cdots,J \qquad (6.10)$$

In the second approach, the measured and predicted spectrum for all the bands are used in the calculation of the Euclidean distance, that is,

$$d_j = \left| \{RSS\} - \{RSS_j\} \right| = \sqrt{\sum_{b=1}^{B} \sum_{m=1}^{M} \left(\hat{P}_m(f_b) - \hat{P}_{mj}(f_b) \right)^2} \quad j = 1, 2, \cdots, J \tag{6.11}$$

The comparison between these two methods is discussed further with the numerical results in the next section. The Euclidean distance can also be presented in a graphical form similar to the acoustic map as used in beamforming. To this end, the Euclidean distance is normalized relative to the maximum level in the environment, i.e. $0 \le d_j \le 1$.

Note that d_j plays a similar role as b_j in the beamforming source localization. Hence, the source location can be estimated in a similar way to the beamforming case, i.e. coordinate weighted by map level with a specific cut-off threshold. Since the larger Euclidean distance corresponds to the smaller beamforming output, the grid points with a normalized Euclidean distance greater than the cut-off threshold, $\hat{d}_{cut-off}$, are ignored. That is, the source coordinate is estimated as

$$\hat{\boldsymbol{l}}_s = \frac{\sum_{j=1}^{J} (1 - \hat{d}_j) \boldsymbol{l}_j}{\sum_{j=1}^{J} (1 - \hat{d}_j)} \tag{6.12}$$

where $\boldsymbol{l}_j = \{x_j, y_j, z_j\}$ is the coordinate vector for the j^{th} grid point and $\hat{d}_j$ is the corresponding Euclidean distance included only if $\hat{d}_j \le \hat{d}_{cut-off}$.

Similarly to the beamforming localization technique, it is very unlikely that a grid point would be situated right where the source is located. Since the simulations are limited to the case where the source is located on a grid point, the Euclidean distance at the actual source position is excluded for a more realistic assessment. The source localization error can then be computed as the Euclidean distance between the true and estimated source location as,

$$e = \left| \hat{\boldsymbol{l}}_s - \boldsymbol{l}_s \right| \tag{6.13}$$

where $\boldsymbol{l}_s$ is the true location vector of the source.

6.3 Numerical Validation

In this section, the validation of the proposed approach is performed using the large 3D models described in Chapter 3. The simplified urban environment model is first selected since the energy method should theoretically be fairly accurate in this type of environment (Pasareanu et al., 2011). The fingerprinting method is then implemented for Fort Benning model since it has been used in the previous investigation. Unlike the simplified model, the Fort Benning model is topologically different without any recognizable urban cannons. Since this is the first evaluation of the method, noise in the microphone signals is not considered.

Localization using one or two microphones is not considered here because at least three non-aligned microphones are required to determine the source position. A single microphone determines the source location to be on the surface of a sphere which is centered at the microphone. A pair of microphones helps to narrow the possible locations down to the intersections of two spherical surfaces, which is still insufficient to determine one unique location. For the same reason, three microphones that are aligned result in poor performance because microphones along a line will not provide any additional information to narrow down the possible locations.

Figure 6.7 shows the location of the five microphones and the source in the simplified urban model. The microphones are located at (x, y, z) = (75, 50, 2), (20, 75, 2), (112, 13, 2), (20, 25, 2) and (112, 88, 2) *m*. The source is located at (3, 40, 2) *m*. Similarly, Figure 6.8 illustrates the Army Fort Benning model with nine microphones dispersed in the environment. The source location is (52, 3, 2) *m*. Note that none of the microphones have a direct line of sight (LoS) to the source for both models.

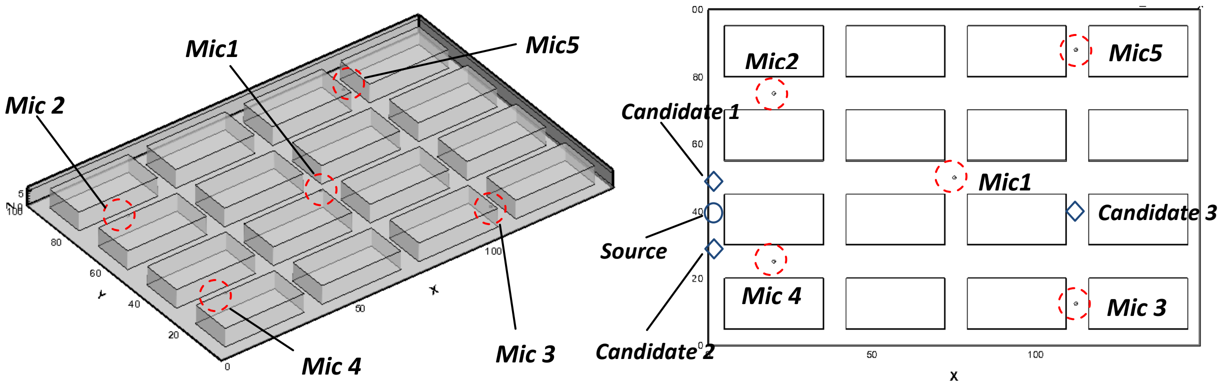


Figure 6.7: Simplified urban environment model: 3D view (left) and top view (right).

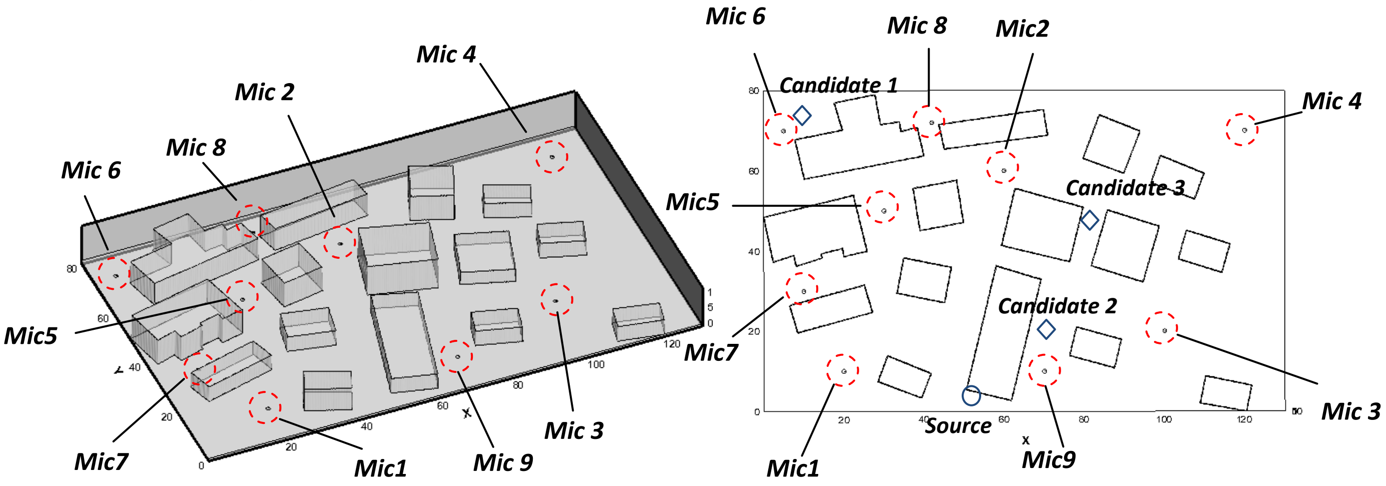


Figure 6.8: Realistic urban environment model based on the Army Fort Benning training facility: 3D view (left) and top view (right).

Here, illustrative results obtained from numerical simulations are presented.

A. Simplified Urban Environment:

The results for the simplified urban environment are presented here. Figure 6.9 shows the time signals (first column) and the corresponding 1/3 octave band spectrum in decibels (second column) for three microphones. Each row corresponds to microphone 1(a), 2(b) and 4(c). It is shown that the source signal arrives to the microphone at different times due to the different propagation path lengths. The source signal arrives to the microphone 4 first (see Figure 6.9). The peak value of the pressure also diminishes as the microphone is located further away from the source due to the spherical propagation. The measured spectrum at the microphone 4 is higher than the other two microphones due to the shortest propagation path.

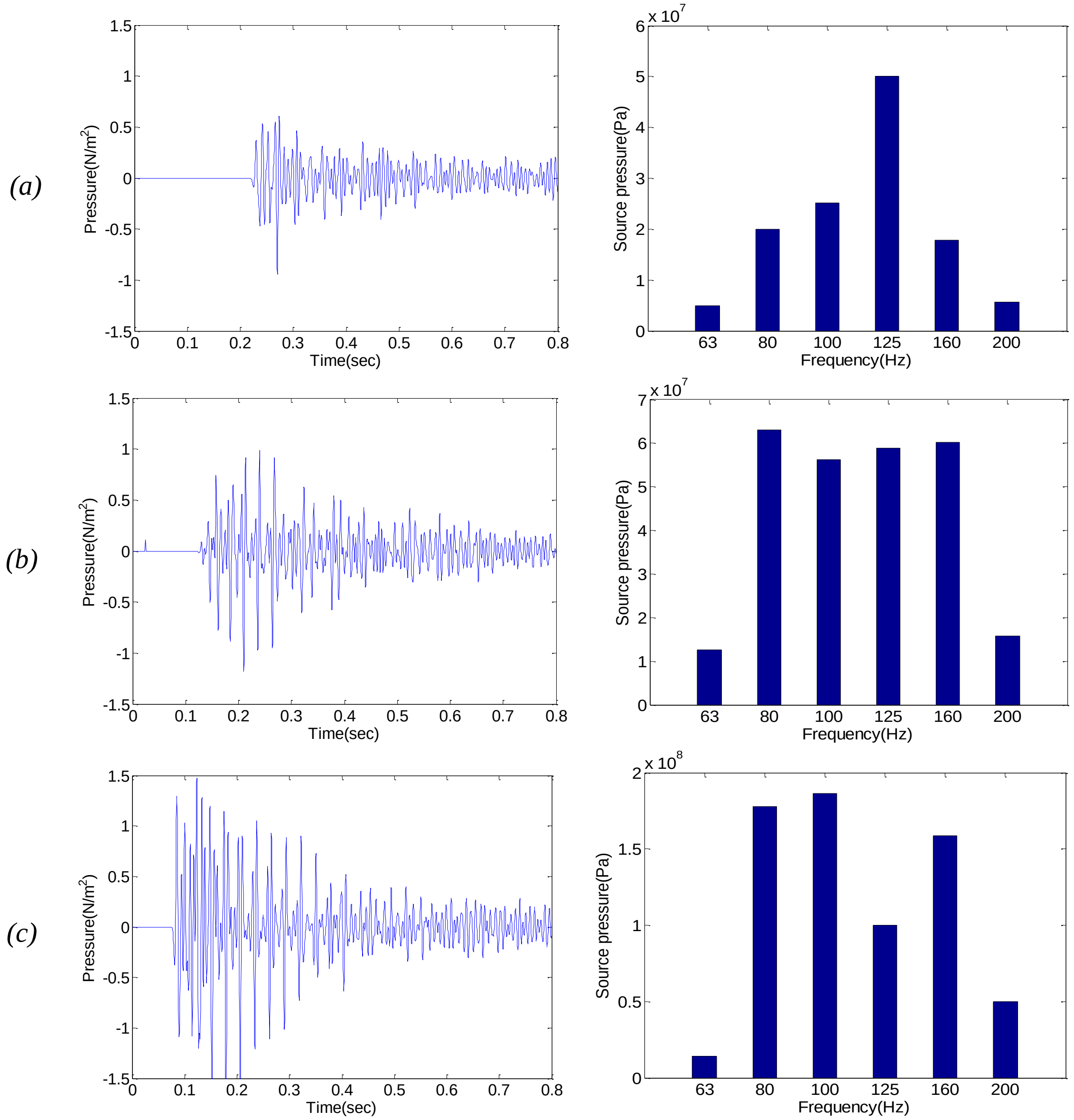


Figure 6.9: Time history and 1/3 octave spectrum for microphones (a) 1, (b) 2, and (c) 4 for simplified urban environment model. Source location is (3, 40, 2) *m*.

As mentioned earlier, the measured spectrums at these microphones are then normalized to compute the $\left\{ RSS\left(f_b \right) \right\}$ at each frequency band as shown in Table 6.1 and Figure 6.10. It is shown that the measured spectrum at microphone 4 is dominant at all six 1/3 octave band, indicating that the microphone 4 is most likely the closest one to the source.

Table 6.1: Microphone normalized measured spectrums at each frequency band.

	$\left\{RSS\left(f_b\right)\right\}=\left\{\hat{P}_1\left(f_b\right),\cdots,\hat{P}_m\left(f_b\right),\cdots,\hat{P}_M\left(f_b\right)\right\}^T$					
	f_b = 63Hz	f_b = 80Hz	f_b = 100Hz	f_b = 125Hz	f_b = 160Hz	f_b = 200Hz
Mic 1	0.279	0.106	0.175	0.210	0.085	0.168
Mic 2	0.613	0.239	0.146	0.185	0.248	0.505
Mic 4	0.739	0.965	0.973	0.959	0.964	0.846

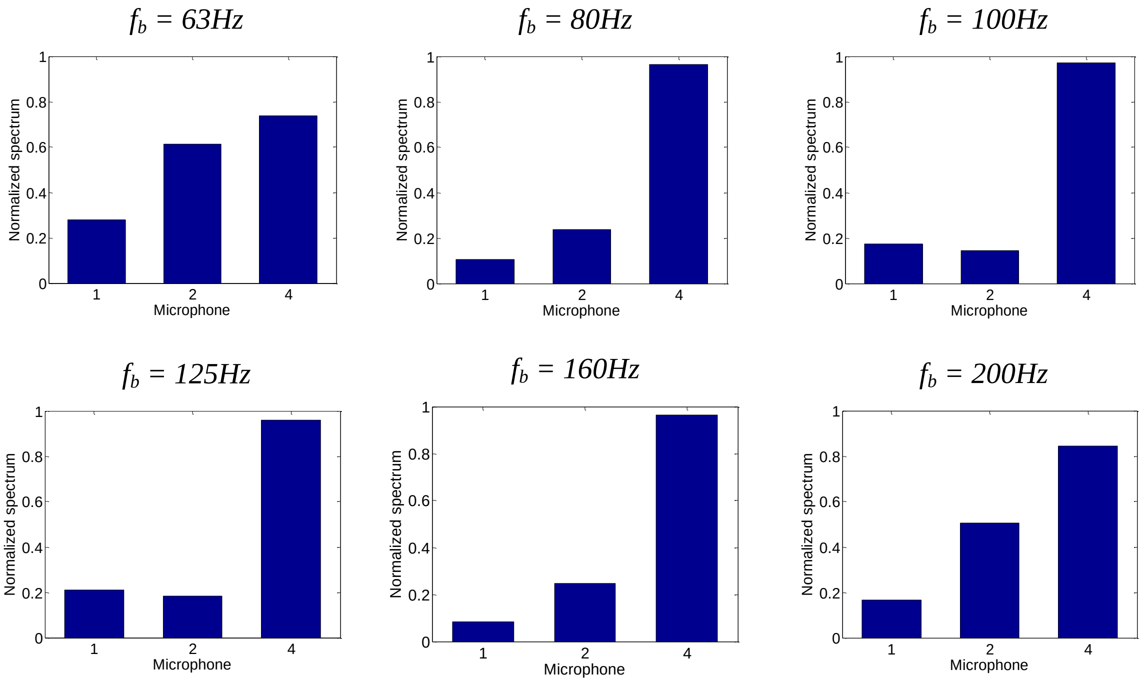


Figure 6.10: Microphone normalized measured spectrums at each frequency band.

As a next step, the database of fingerprints is constructed using the energy method. To this end, the grid point for source candidate location is selected to match the same grid points of the energy model, i.e. 150,000 uniformly distributed grid points with a uniform grid spacing of 1 *m*. Examples of these fingerprints for four candidate source locations (including the one where the source is located) are shown in Table 6.2. The locations of the candidate sources are (115, 40, 2), (3, 50, 2), (3, 40, 2) and (3, 30, 2) *m* which are shown in Figure 6.7.

Table 6.2: Examples of predicted fingerprints for different grid point, i.e. candidate source locations.

	Grid Point: (115,40,2)					
	$\left\{RSS_j(f_b)\right\}=\left\{\hat{P}_{1j}(f_b),\cdots,\hat{P}_{mj}(f_b),\cdots,\hat{P}_{Mj}(f_b)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.997	0.997	0.997	0.997	0.997	0.997
Mic 2	0.041	0.041	0.041	0.041	0.041	0.041
Mic 4	0.061	0.061	0.061	0.061	0.061	0.061
$d_j(f_b)$	1.141	1.285	1.232	1.203	1.300	1.232
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j(f_b)$	1.232					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m(f_b)-\hat{P}_{mj}(f_b)\right)^2}$	2.983					
	Grid Point: (3,50,2)					
	$\left\{RSS_j(f_b)\right\}=\left\{\hat{P}_{1j}(f_b),\cdots,\hat{P}_{mj}(f_b),\cdots,\hat{P}_{Mj}(f_b)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.117	0.117	0.117	0.117	0.117	0.117
Mic 2	0.703	0.703	0.703	0.703	0.703	0.703
Mic 4	0.701	0.701	0.701	0.701	0.701	0.701
$d_j(f_b)$	0.189	0.534	0.623	0.586	0.526	0.251
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j(f_b)$	0.451					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m(f_b)-\hat{P}_{mj}(f_b)\right)^2}$	1.163					
	Grid Point: (3,40,2) – Actual Source Location					
	$\left\{RSS_j(f_b)\right\}=\left\{\hat{P}_{1j}(f_b),\cdots,\hat{P}_{mj}(f_b),\cdots,\hat{P}_{Mj}(f_b)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.062	0.062	0.062	0.062	0.062	0.062
Mic 2	0.270	0.270	0.270	0.270	0.270	0.270
Mic 4	0.960	0.960	0.960	0.960	0.960	0.960
$d_j(f_b)$	0.462	0.054	0.163	0.171	0.032	0.282
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j(f_b)$	0.194					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m(f_b)-\hat{P}_{mj}(f_b)\right)^2}$	0.543					
	Grid Point: (3,30,2)					
	$\left\{RSS_j(f_b)\right\}=\left\{\hat{P}_{1j}(f_b),\cdots,\hat{P}_{mj}(f_b),\cdots,\hat{P}_{Mj}(f_b)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.027	0.027	0.027	0.027	0.027	0.027
Mic 2	0.063	0.063	0.063	0.063	0.063	0.063
Mic 4	0.997	0.997	0.997	0.997	0.997	0.997
$d_j(f_b)$	0.658	0.196	0.171	0.223	0.197	0.487
$d_\ell=\frac{1}{B}\sum_{b=1}^{B}d_\ell(f_b)$	0.322					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m(f_b)-\hat{P}_{mj}(f_b)\right)^2}$	0.892					

As shown in the Table 6.2, the fingerprints $\left\{RSS_j\left(f_b\right)\right\}$ are the same for all the bands. The reason is that perfectly reflective boundaries were assumed and thus the propagation is not a function of frequency, i.e. $\left\{RSS_j\left(f_b\right)\right\}=\left\{RSS_j\left(f_1\right)\right\}=\left\{RSS_j\left(f_B\right)\right\}$. Table 6.2 also shows the Euclidean distance computed using the two approaches presented in the previous section. The first approach (method 1) provides smaller Euclidean distances than the second approach (method 2). However, both approaches show the similar pattern in the relative difference among the candidate source locations in the Table 6.2. Among them, the grid point at (3, 40, 2) *m* results in the smallest Euclidean distance because that is the true location of the source. The predicted fingerprint for the actual source location is shown in Figure 6.11. As expected, it looks very similar to the microphone measured spectrums in Figure 6.10.

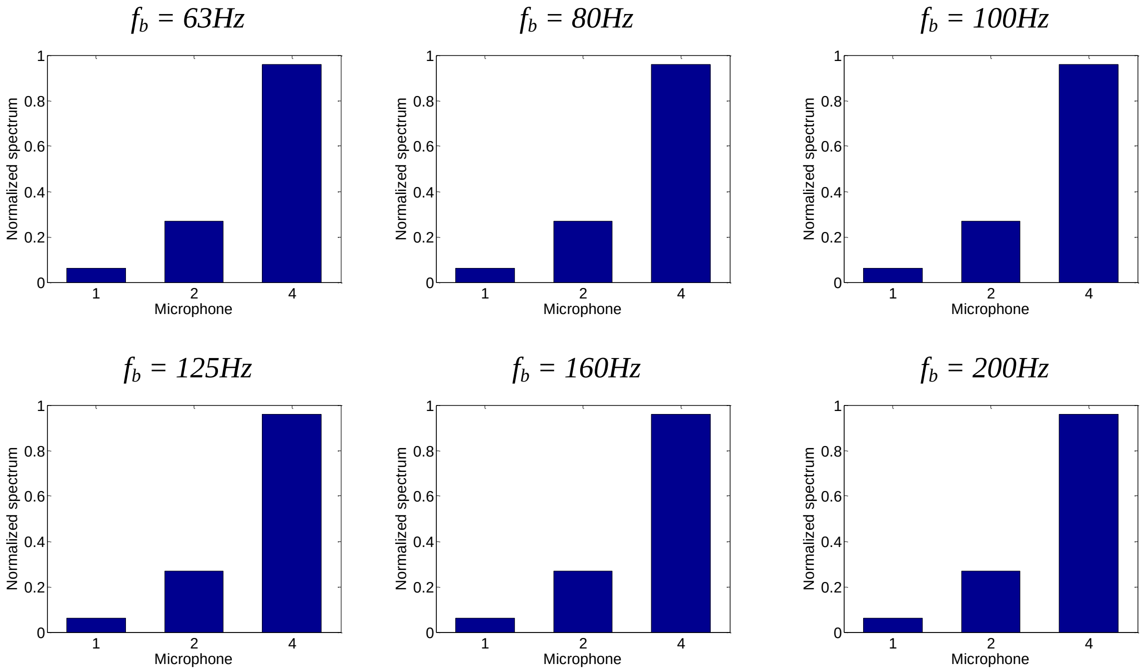


Figure 6.11: Predicted fingerprint for actual source location.

To visualize the Euclidean distance for all the grid points, the results for $d_j\left(f_b\right)$ normalized by the maximum value in the domain are presented in a color map which is similar to the localization map using beamforming in Chapter 5. Note that only three microphones (1, 2 and 4) are used for an initial study. Figure 6.12 shows the localization map for the entire frequency range that includes all six bands using both methods. The topology, source, and microphones are superimposed on the map. The estimated source location is shown in each localization map using a diamond shaped symbol. In these results the cut-off threshold used is $\hat{d}_{cut-off} = 0.1$. A color scale is then used to represent the numerical value: red represents a value of 0 (minimum) while the dark blue is 1 (maximum).

In method 1 (Figure 6.12a), the Euclidean distance computed for each band is averaged across the bands. In method 2 (Figure 6.12b), a single vector is constructed from the data at all frequency bands and the Euclidean distance computed. It is shown in Figure 6.12 that method 1 gives slightly better estimate than method 2, i.e. the region of the source becomes sharper. This observation was found to be true for all cases investigated (i.e. different source locations). Therefore, the normalized Euclidean distance maps presented from this point are obtained via method 1 (Choi et al., 2011b).

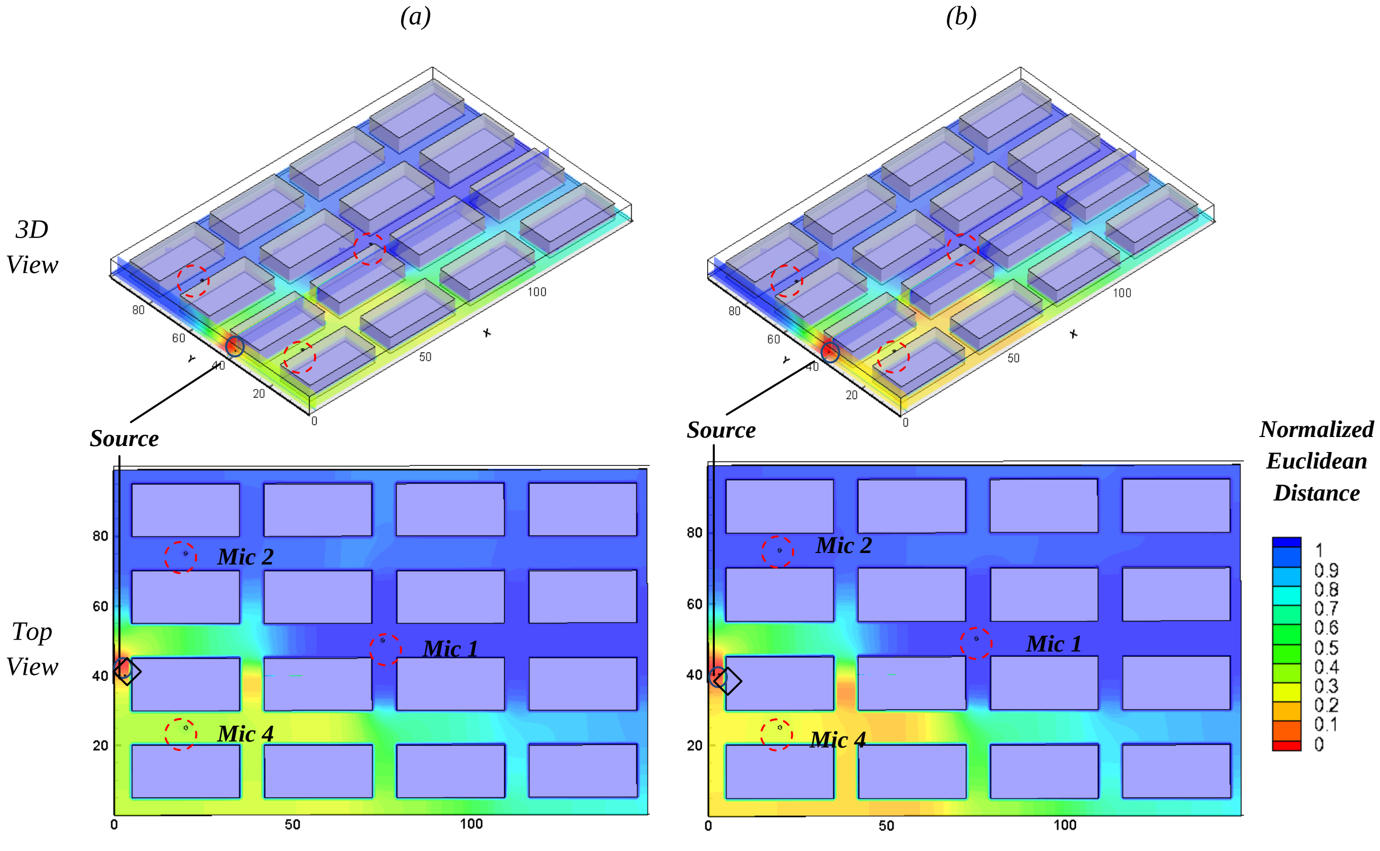


Figure 6.12: Localization map for the simplified urban model for the entire frequency bands using both methods; method 1 (a) and method 2 (b). Source is at (3, 40, 2) *m*.

So far, the results show only one particular case of three dispersed microphones. Figure 6.13 shows the performance of the source localization for several combinations of 4 and 5 dispersed microphones. As expected, the additional microphone does not improve the performance as long as the first three microphones are closer to the source than the additional one (see Figure 6.12a and compare to Figure 6.13c and 6.13d). Significant improvement is observed only if the new microphone added is the one closest to the source as demonstrated in Figure 6.13d (as compared to Figure 6.12a). Finally, the case where all five microphones are used is shown in Figure 6.13d. The performance is not improved as compared to the best three microphone case in Figure 6.12a indicating the overwhelming importance of the microphone 4 (closest to the source) in the source localization performance.

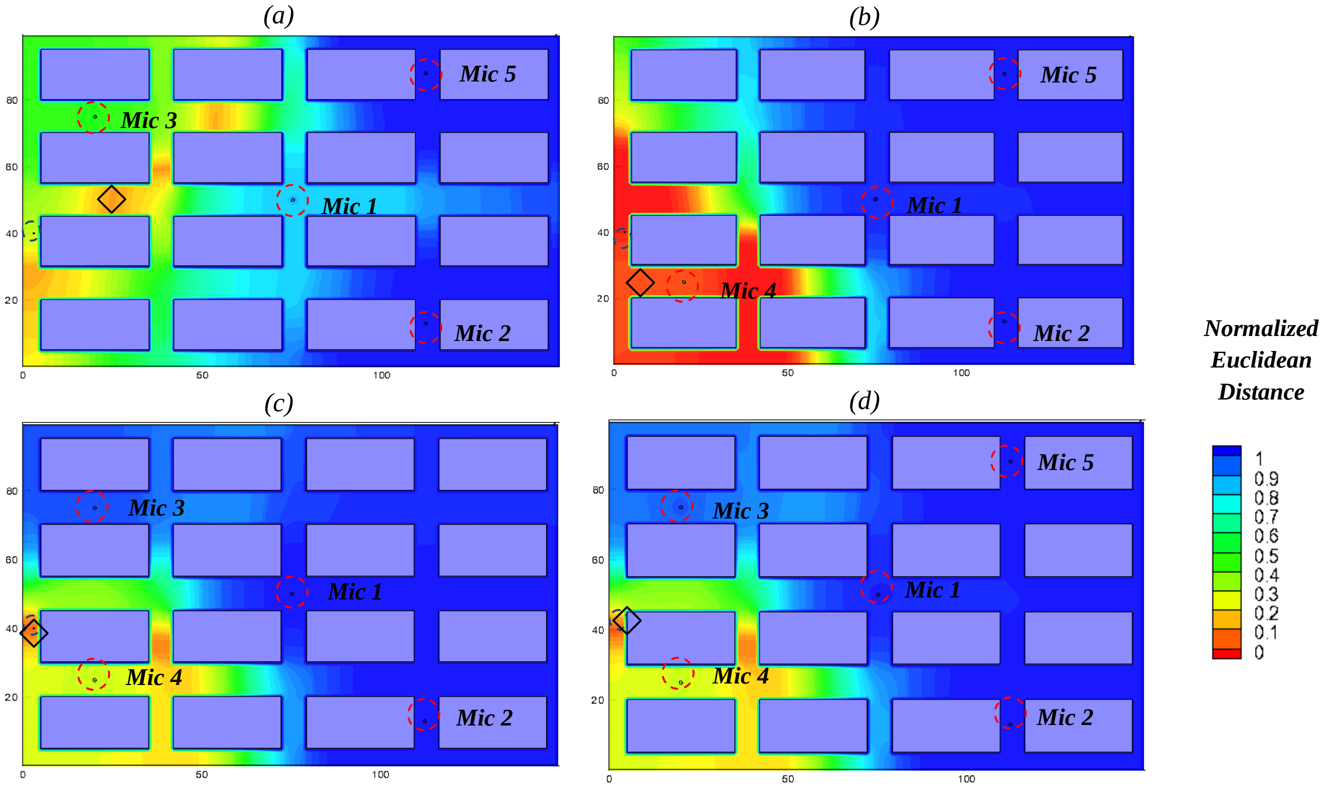


Figure 6.13: Examples of localization map (top view) for the simplified urban model using four and five microphones. Source is at (3, 40, 2) *m*.

B. Fort Benning Urban Environment:

The same type of results is presented here for the Army Fort Benning urban environment training facility shown in Figure 3.6. For this case, the number of microphones implemented in the simulations ranged from 3 to 9 (see Figure 6.8). The number of possible microphone combinations is too large to present all of them here. Therefore, a limited number of cases are shown in this section with the additional results found in Appendix D.

The time signals (first column) and corresponding 1/3th octave band spectrum (second column) for microphone 1, 2 and 3 are shown in Figure 6.14. Each row corresponds to microphone 1(a), 2(b) and 3(c). It is interesting to note that the fluctuation of the received signals is not as severe as in the simplified urban model in spite of the irregularity of the buildings. The larger space between the buildings seems to reduce cluttering in the model. The results in Figure 6.14 show that the source signal arrives to the microphone 1 first due to the shortest propagation path. For the same reason, the measured spectrum at the microphone 1 shows the highest source pressure level in all six bands. On the other hand, the source signal reaches later to the microphone 2 and 3. However, comparing Figure 6.14b and Figure 6.14c demonstrates that the spectral content of the microphone signals are heavily distorted by the clutter on their paths.

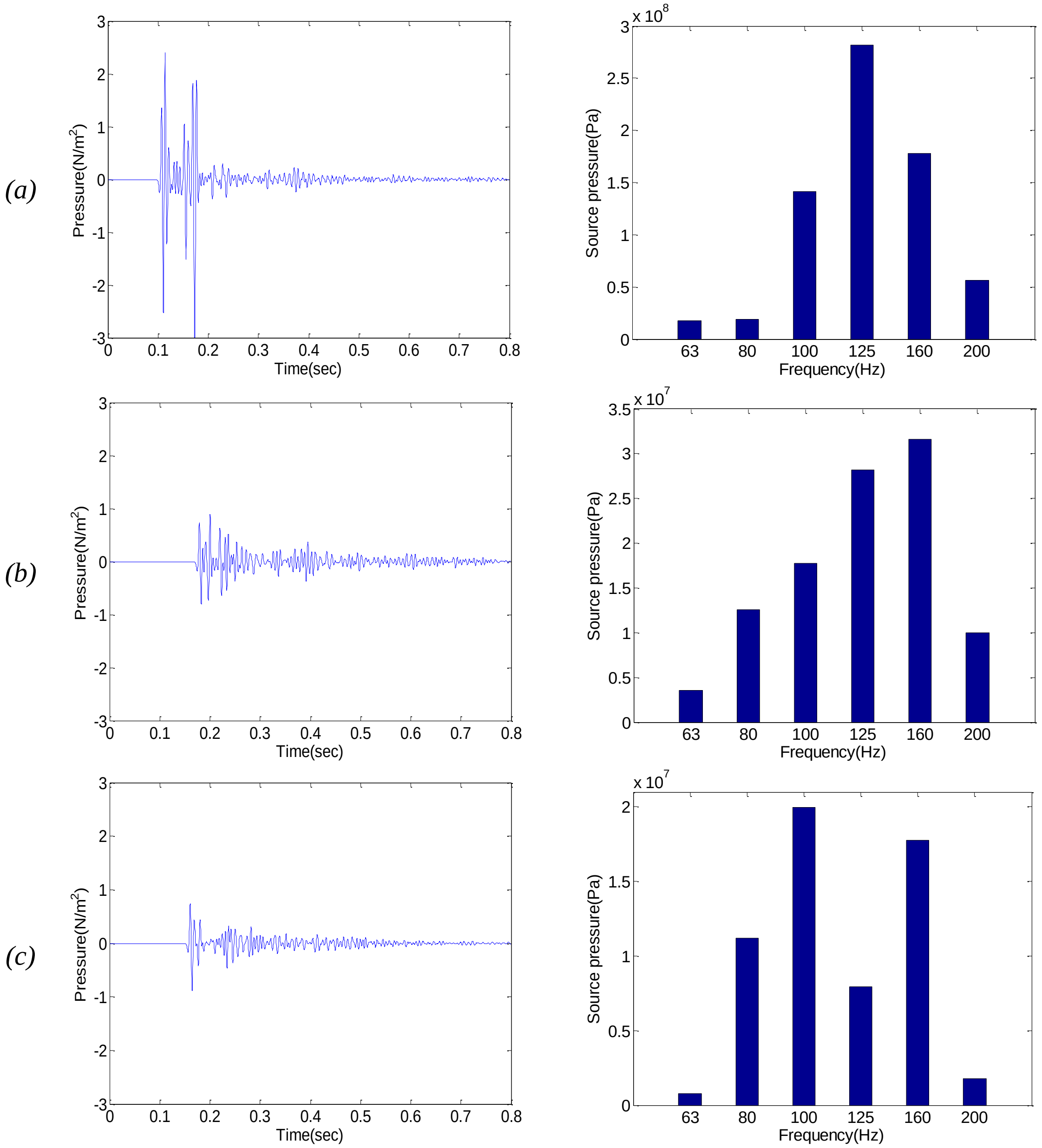


Figure 6.14: Time history and spectrum of the received signal at microphone 1(a), 2(b) and 3(c) for a Fort Benning urban model. Source location is (52, 3, 2) *m*.

The normalized measured spectrums at these microphones to compute $\left\{RSS\left(f_b\right)\right\}$ are shown next in Table 6.3 and Figure 6.15. It is easy to find out that the microphone 1 dominates the other microphones at all six 1/3 octave bands. On the other hand, the microphone 3 results in the smallest normalized measured spectrum at all six bands because it is far away and there is a lot of cluttering between the source and the microphone.

Table 6.3: Normalized measured spectrum for each microphone.

	$\left\{RSS\left(f_b\right)\right\}=\left\{\hat{P}_1\left(f_b\right),\cdots,\hat{P}_m\left(f_b\right),\cdots,\hat{P}_M\left(f_b\right)\right\}^T$					
	f_b = 63Hz	f_b = 80Hz	f_b = 100Hz	f_b = 125Hz	f_b = 160Hz	f_b = 200Hz
Mic 1	0.961	0.764	0.983	0.974	0.921	0.779
Mic 2	0.239	0.598	0.158	0.218	0.369	0.624
Mic 3	0.136	0.241	0.087	0.044	0.122	0.073

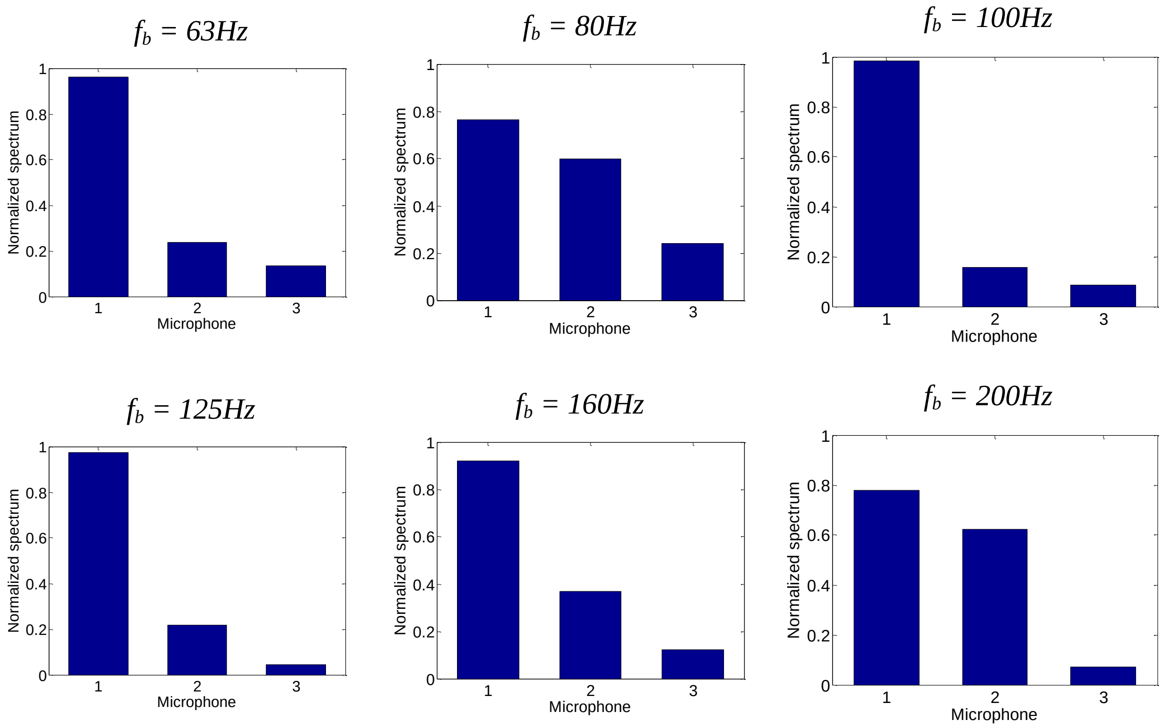


Figure 6.15: Microphone normalized measured spectrums at each frequency band.

The database of fingerprints is then constructed using the energy method. To this end, 145,600 uniformly distributed grid point for source candidate location was selected to match the same grid points of the energy model. Similarly, examples of these fingerprints for four candidate source locations including the true one are presented in Table 6.4. The locations of the candidate sources are (5, 72, 2), (85, 20, 2), (90, 50, 2) and (52, 3, 2) *m* which are shown in Figure 6.6. As expected, the grid point at (52, 3, 2) *m* provides the smallest Euclidean distance.

Table 6.4: Examples of predicted fingerprints for different grid points, i.e. candidate source locations.

	Grid Point: (5,72,2)					
	$\left\{RSS_j\left(f_b\right)\right\}=\left\{\hat{P}_{1j}\left(f_b\right),\cdots,\hat{P}_{mj}\left(f_b\right),\cdots,\hat{P}_{Mj}\left(f_b\right)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.468	0.468	0.468	0.468	0.468	0.468
Mic 2	0.882	0.882	0.882	0.882	0.882	0.882
Mic 3	0.043	0.043	0.043	0.043	0.043	0.043
$d_j\left(f_b\right)$	0.816	3.686	0.890	0.835	0.689	0.191
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j\left(f_b\right)$	0.646					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m\left(f_b\right)-\hat{P}_{mj}\left(f_b\right)\right)^2}$	1.683					
	Grid Point: (85,20,2)					
	$\left\{RSS_j\left(f_b\right)\right\}=\left\{\hat{P}_{1j}\left(f_b\right),\cdots,\hat{P}_{mj}\left(f_b\right),\cdots,\hat{P}_{Mj}\left(f_b\right)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.048	0.048	0.048	0.048	0.048	0.048
Mic 2	0.083	0.083	0.083	0.083	0.083	0.083
Mic 3	0.995	0.995	0.995	0.995	0.995	0.995
$d_j\left(f_b\right)$	1.261	6.301	1.307	1.324	1.259	1.264
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j\left(f_b\right)$	1.261					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m\left(f_b\right)-\hat{P}_{mj}\left(f_b\right)\right)^2}$	3.026					
	Grid Point: (90,50,2)					
	$\left\{RSS_j\left(f_b\right)\right\}=\left\{\hat{P}_{1j}\left(f_b\right),\cdots,\hat{P}_{mj}\left(f_b\right),\cdots,\hat{P}_{Mj}\left(f_b\right)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.036	0.036	0.036	0.036	0.036	0.036
Mic 2	0.876	0.876	0.876	0.876	0.876	0.876
Mic 3	0.479	0.479	0.479	0.479	0.479	0.479
$d_j\left(f_b\right)$	1.174	6.270	1.252	1.226	1.081	0.722
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j\left(f_b\right)$	1.045					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m\left(f_b\right)-\hat{P}_{mj}\left(f_b\right)\right)^2}$	2.582					
	Grid Point: (52,3,2) – Actual Source Location					
	$\left\{RSS_j\left(f_b\right)\right\}=\left\{\hat{P}_{1j}\left(f_b\right),\cdots,\hat{P}_{mj}\left(f_b\right),\cdots,\hat{P}_{Mj}\left(f_b\right)\right\}^T$					
	f_b *= 63Hz*	f_b *= 80Hz*	f_b *= 100Hz*	f_b *= 125Hz*	f_b *= 160Hz*	f_b *= 200Hz*
Mic 1	0.933	0.933	0.933	0.933	0.933	0.933
Mic 2	0.293	0.293	0.293	0.293	0.293	0.293
Mic 3	0.209	0.209	0.209	0.209	0.209	0.209
$d_j\left(f_b\right)$	0.161	1.291	0.217	0.252	0.234	0.683
$d_j=\frac{1}{B}\sum_{b=1}^{B}d_j\left(f_b\right)$	0.329					
$d_j=\sqrt{\sum_{b=1}^{B}\sum_{m=1}^{M}\left(\hat{P}_m\left(f_b\right)-\hat{P}_{mj}\left(f_b\right)\right)^2}$	0.829					

Figure 6.16 shows the predicted fingerprint for the actual source location. As expected, it looks relatively similar to the microphone measured spectrums shown in Figure 6.15.

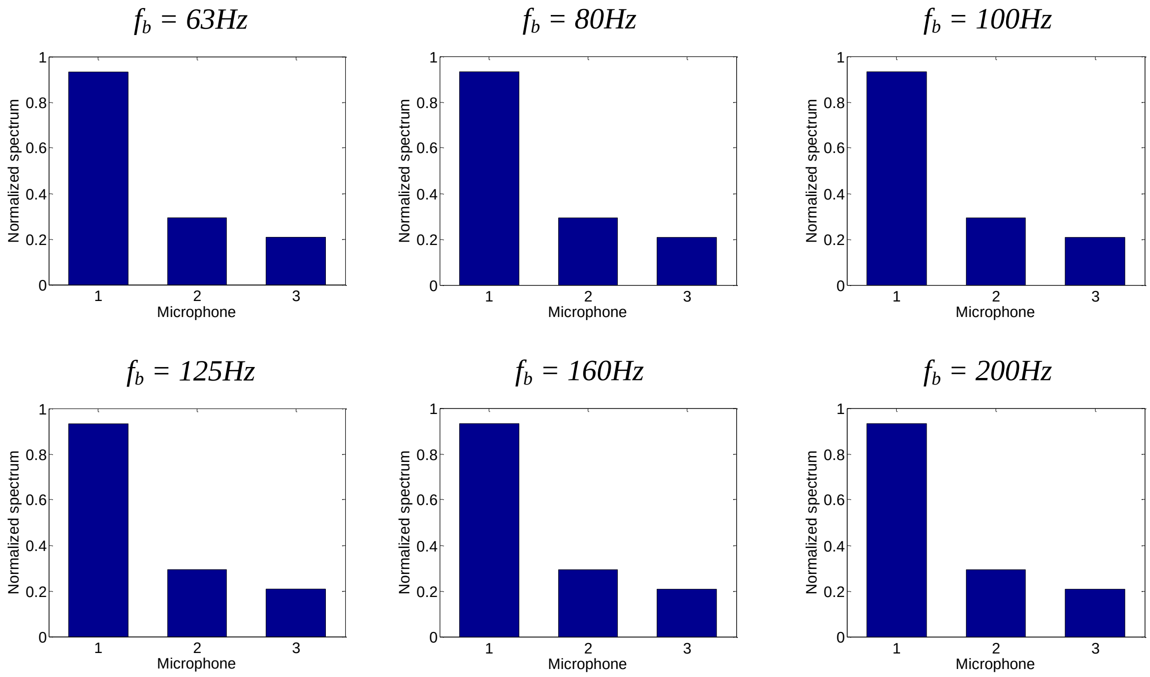


Figure 6.16: Predicted fingerprint for actual source location.

The localization maps for this case are then shown in Figures 6.17 and 6.18. Figure 6.17 shows the normalized Euclidean distance map (top view) for the entire frequency range that includes all six bands using both localization methods. As observed, the source location estimates are not as accurate as the ones for the simplified urban model even with a microphone close to the source. The localization error is more than 10 *m* as compared to 5 *m* for the simplified urban model. This is because the energy method models better the simplified urban model due to its regularly spaced streets and identical clutters. In other words, the energy method provides more accurate fingerprints for the simplified urban model. However, the results still identify the general area of the source location. Here, method 1 gives a closer estimate of the source placement than method 2. Hence, method 1 is used for the rest of the results.

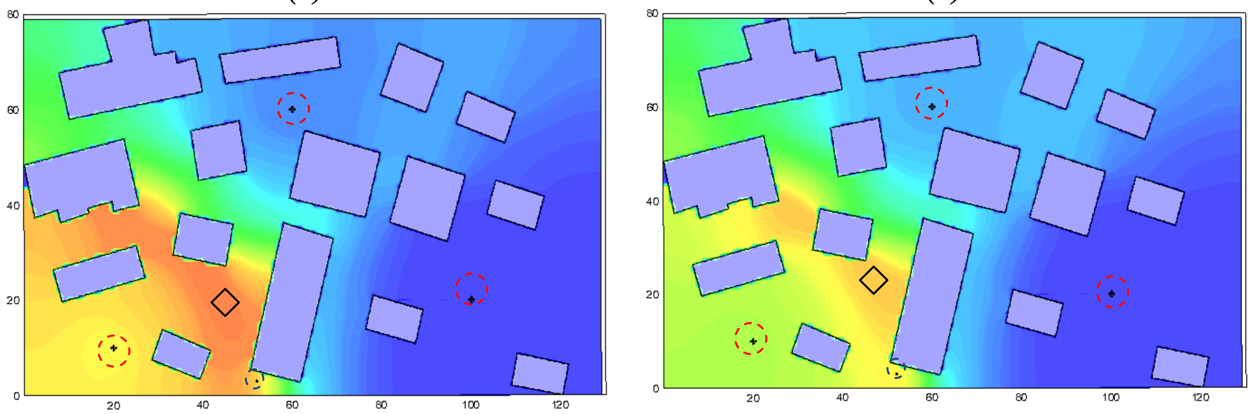

(a) *(b)*

Figure 6.17: Localization map (Top view) for the Fort Benning urban model for the entire frequency bands using both methods; method 1 (a) and method 2 (b). Source is at (52, 3, 2) *m*.

A few additional cases of three-microphone combinations are depicted in Figure 6.18. The results again shows a relatively good estimation of the region where the source is located although the localization accuracy still needs to be improved.

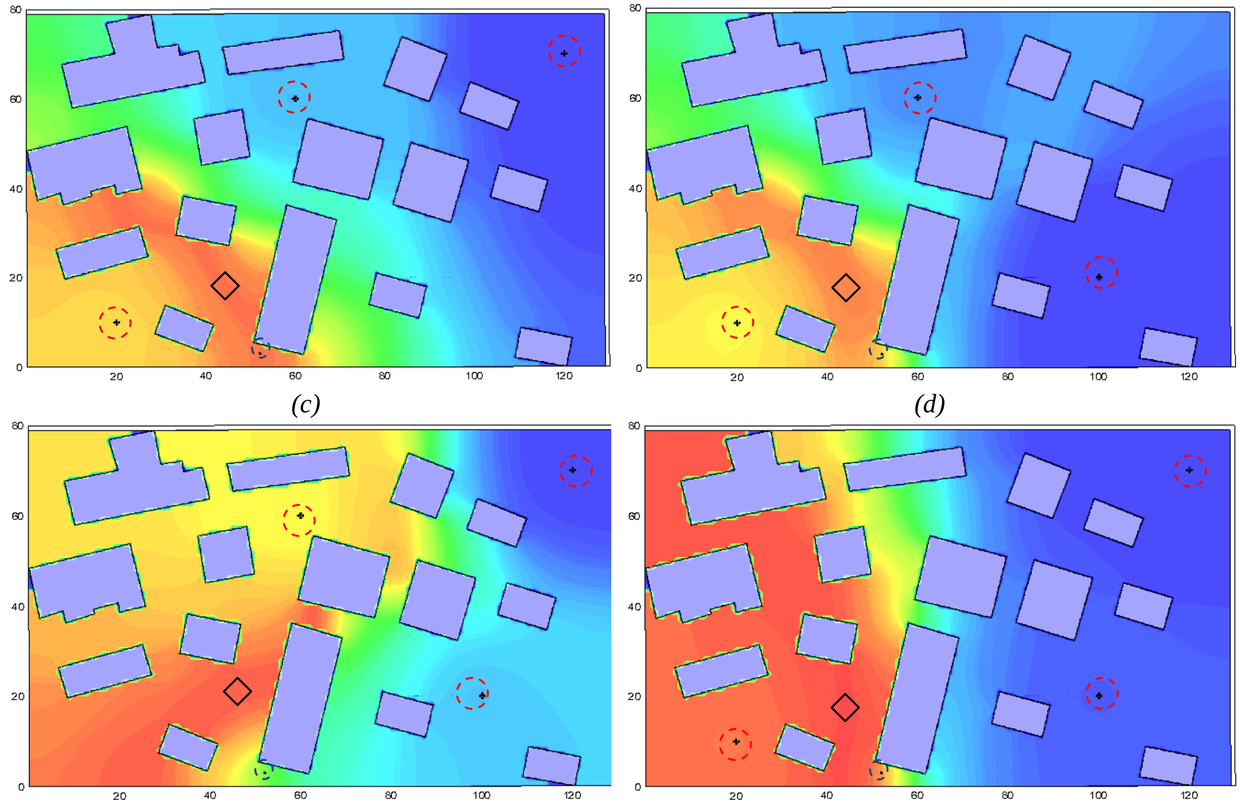

(a) *(b)* *(c)* *(d)*

Figure 6.18: Localization map (Top view) for Fort Benning urban model with three microphones using method 1. Source is at (52, 3, 2) *m*.

Since there are too many possible combinations of the nine dispersed microphones, Figure 6.19 shows localization maps for increasing the number of microphones from four (Figure 6.19a) to nine (Figure 6.19f). The microphones used in this figure are randomly selected. The results show a progressive improvement in the source localization capabilities as the number of microphones is increased. However, more microphones do not guarantee better performance as the case of eight and nine microphones in Figures 6.19e and 6.19f. In fact, adding the ninth microphone produces a slight degradation of the performance since the location of the additional microphone is far from the source placement.

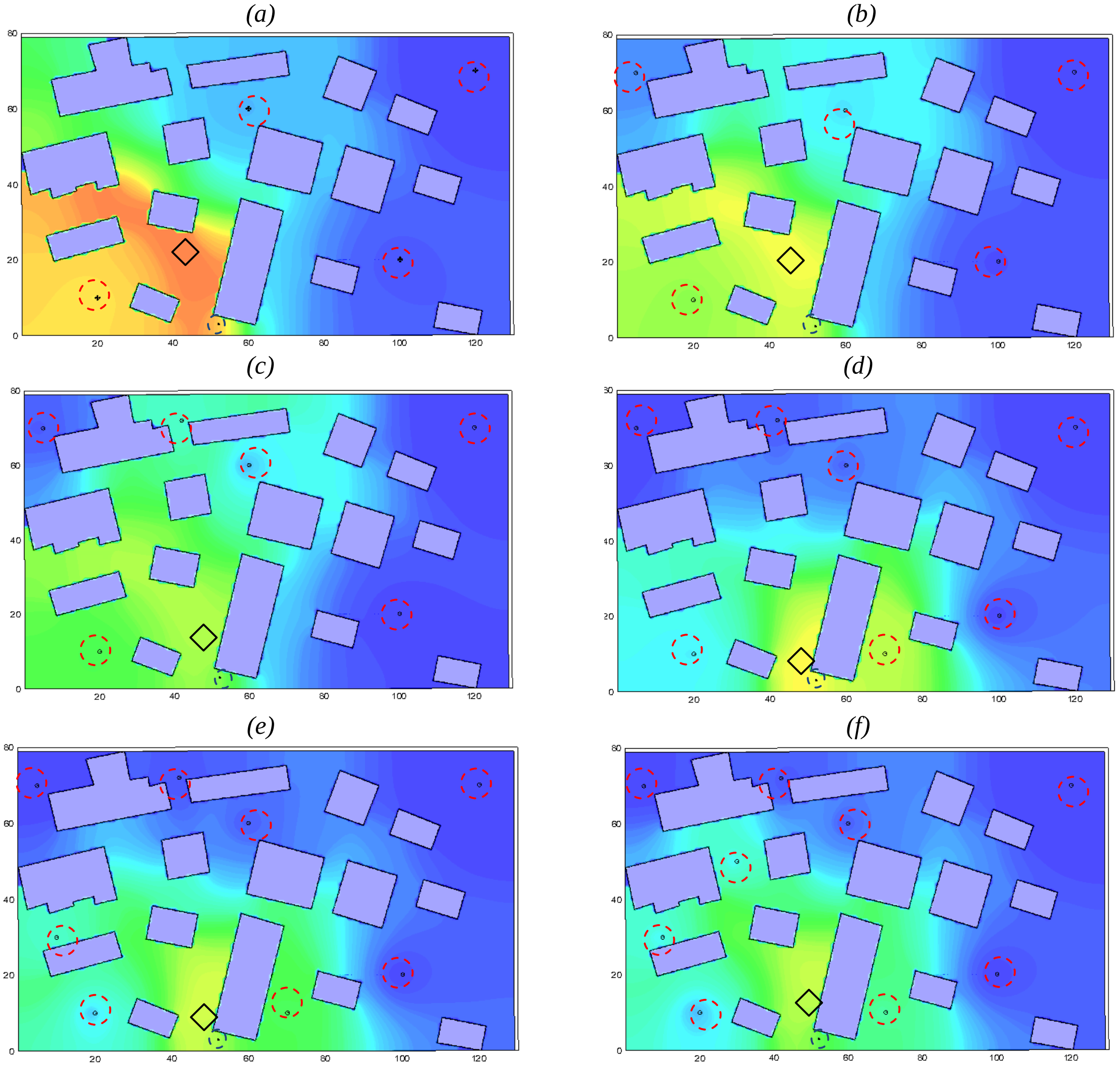

(a) *(b)* *(c)* *(d)* *(e)* *(f)*

Figure 6.19: Localization map (Top view) for the Fort Benning urban model with multiple microphones (from four to nine) using method 1. Source is at (52, 3, 2) *m*.

In order to assess the localization accuracy as a function of the number of microphones, the localization error (average and standard deviation) is then presented in Figure 6.20 for increasing the number of microphones (from four to nine). Since there are several combinations for each case of number of microphones, the localization error for all possible combinations is computed and the averaged and standard deviation determined. For example, the localization error for the eight-microphone case is computed for all nine possible combinations. The average (15 *m*) is then plotted using the open circle symbol while the bar indicated the mean ± the standard deviation. The results in Figure 6.20 show that the averaged localization error decreases as the number of microphones increases, i.e. the localization error for the four-microphone case is around 23 *m* whereas the localization error for the nine-microphone case decreases to less than 10 *m*. The figure also shows a reduction in the uncertainty of the estimate. For example, the standard deviation for the four-microphone case is 16.1 *m* as compared to 8.7 *m* for the eight-microphone case.

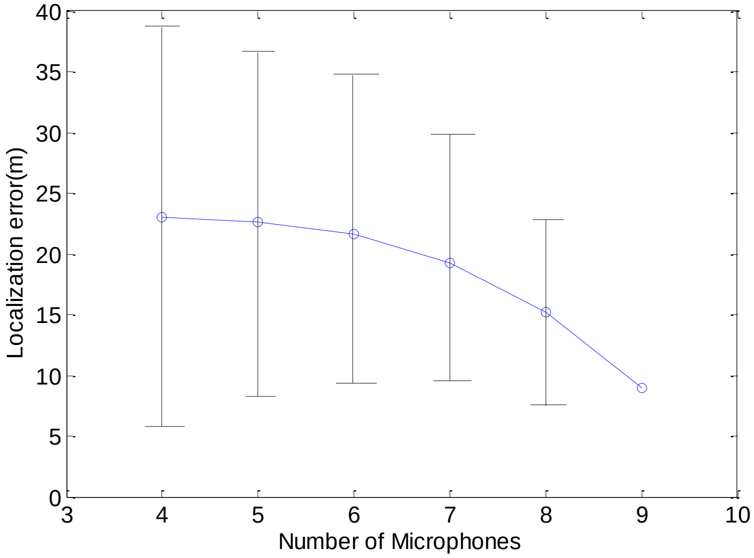


Figure 6.20: Localization error for the Fort Benning urban model for increasing the number of microphones.

6.4 Effect of the accuracy of the fingerprints

An important issue to be addressed is the effect of the accuracy of the fingerprints on the source localization performance. To assess this impact, the fingerprints are computed using both the FDTD code (exact fingerprints) and the energy-based method (approximate fingerprints). Then the fingerprinting localization performance using the exact and approximate fingerprints is compared.

A. Simplified Urban Environment:

Figure 6.21 shows this comparison for four three-microphone configurations. The results on the left and right columns of the figure are obtained using the exact and approximate fingerprints, respectively. As expected, the figure shows the better results when using the exact fingerprints. For example, the localization error when using the exact fingerprints is less than 5 *m* for all four configurations (Figure 6.21a, c, e and g). On the other hand, the localization error when using the approximate fingerprints increases to 6-8 *m* for three of them (Figure 6.21b, d and f). However, the overall performance difference between them is not remarkable. This comparison suggests that an extremely accurate model of the environment is not needed to implement the fingerprinting method, i.e. the energy method provides relatively accurate fingerprints.

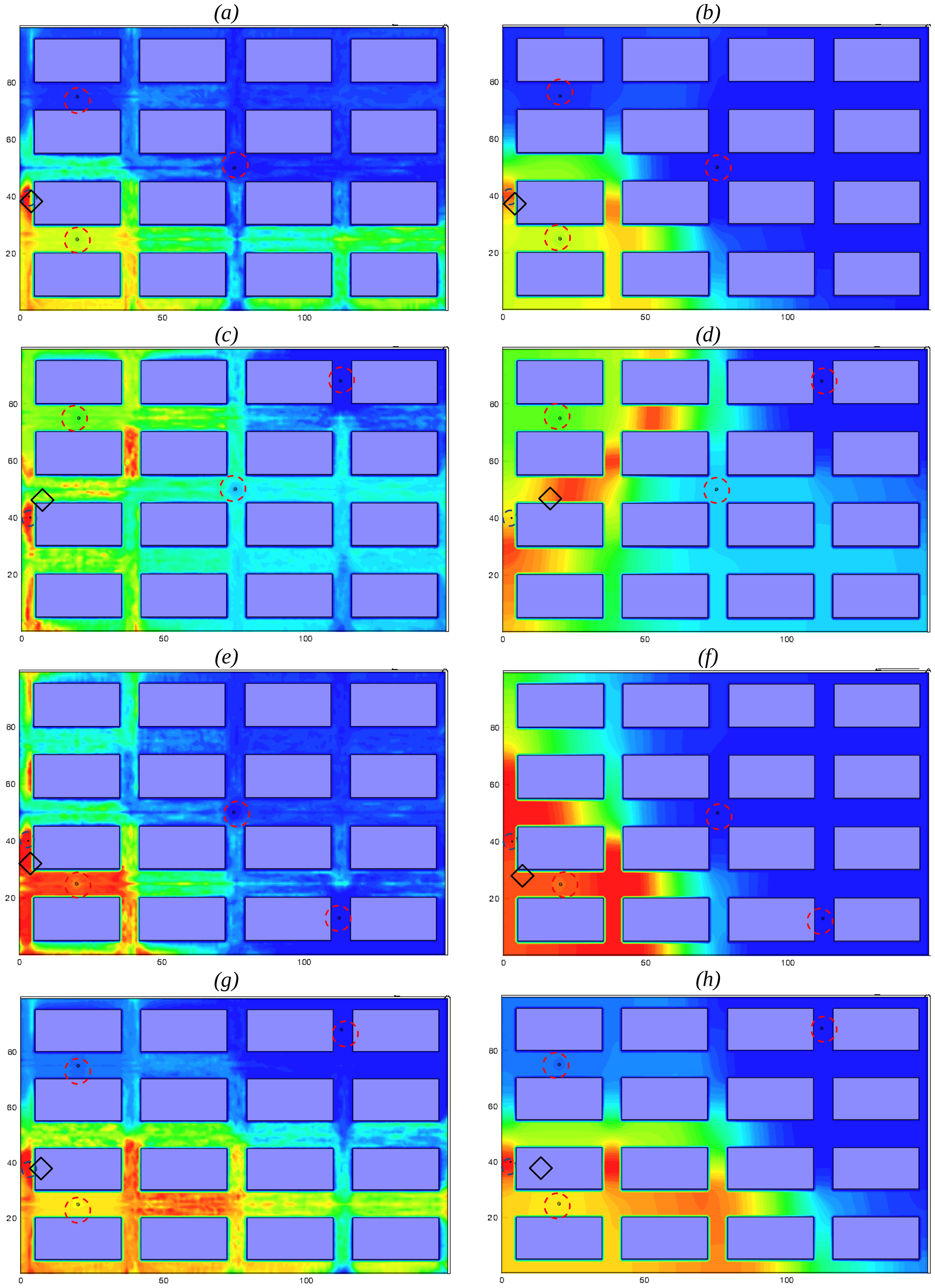


Figure 6.21: Localization map (Top view) for the simplified urban environment with three microphones using exact (left column, FDTD model) and approximate (right column, Energy model) fingerprints. Source is at (3, 40, 2) *m*.

B. Fort Benning Urban Environment:

The same type of comparison study using the approximate (energy-based acoustic model) and exact (FDTD acoustic model) fingerprints are presented next. Figure 6.22 shows this comparison for the configurations presented with increasing number of microphones. The results on the left and right columns of the figure are obtained using the exact and approximate fingerprints, respectively. The figure shows that the results are remarkably better using the exact fingerprints as the case of four microphones in Figures 6.22a and 6.22b. However, the performance for more microphones, from five (Figure 6.22c and 6.22d) to seven (Figure 6.22k and 6.22l), is not markedly different than using the approximate fingerprints computed using the energy model. This comparison suggests that an extremely accurate model of the environment is not needed to implement the fingerprinting method as long as enough microphones are deployed in the environment.

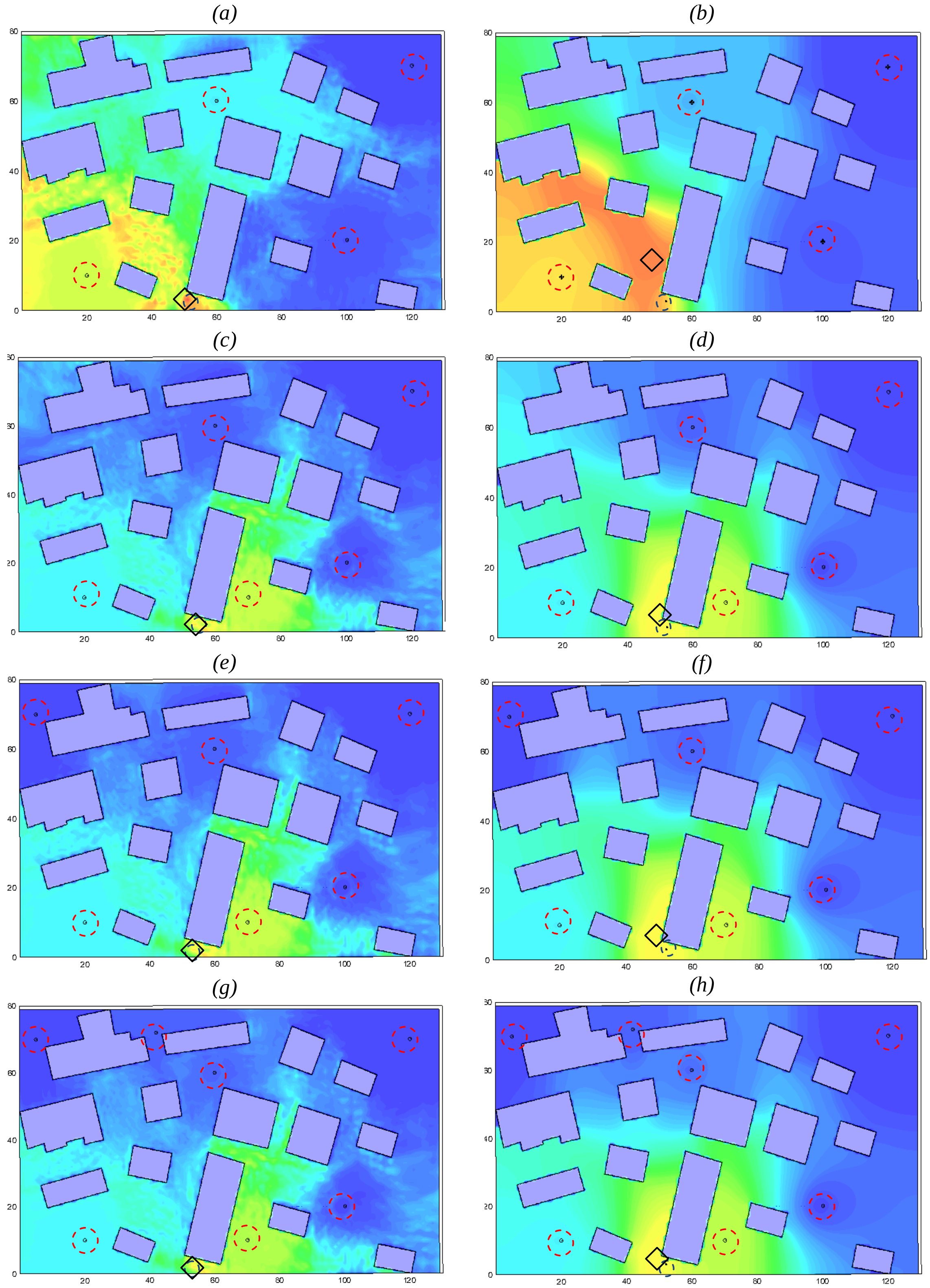


Figure 6.22: Normalized Euclidean distance map (Top view) for Fort Benning urban environment with four to seven microphones using both propagation models; FDTD (left) and Energy-based method (right). Source is at (52, 3, 2) *m*.

Chapter 7 Conclusions and Recommendations

The problem of acoustic localization of impulsive sources in cluttered urban environments has been studied. This final chapter reviews major observations in light of the work presented and offers some recommendations on future directions.

7.1 Conclusions

In this dissertation, two source localization techniques were developed. They are: (a) beamforming in conjunction with a high fidelity model (FDTD) and (b) acoustic fingerprinting in conjunction with a low fidelity model (Energy method). Here, several important observations from this investigation are addressed.

It is important to mention that in this work the methods were evaluated using 3D models and a full range of uncorrelated noise for more realistic simulations and to test the limits of the beamformers investigated. This should be contrasted to previous research where only 2D models were used and very small amount of uncorrelated noise added (Liu et al., 2007; Mennitt 2008). Also, the microphone SNR in previous research was defined as the ratio of the average mean-square-value of the signal to the noise. However, for an impulsive source this approach is erroneous since the SNR changes as a function of the length of the time window used. To better represent the impulsive characteristic of the source, the SNR in terms of the peak values was used in this dissertation. Moreover, in order to assess the capability of the beamformers to handle uncorrelated noise independent of the model, the WNGs of the beamformers were numerically compared.

Firstly, the EMVDR was identified as the only beamformer that has accurate localization capability as well as good robustness to uncorrelated noise. For example, the average localization error of the EMVDR beamformer for 3D three-building models was less than 1 *m* for SNR > 5dB. Even for the extreme case of 0 dB SNR, where the peak value of the noise was identical to that of the signal, the localization error was less than 4 *m* as compared to 12 *m* for the conventional DS beamformer.

Secondly, it was shown that localization in 3D models was more accurate than the corresponding 2D models. The reason for this observation is that the 3D models had additional propagation paths (i.e. more information about the environment) to make more accurate localization. For instance, the localization error of the EMVDR for 3D single-building model decreased from 5 *m* to less than 2 *m* for 0 dB SNR. For 3D three-building model, the localization error was reduced from 9 *m* to 4 *m* for the same configuration. Therefore, evaluation in 3D models is critical to assess correctly the potential of the localization technique (i.e. 2D models showed inadequate representation of the realistic environment).

Moreover, the results showed that the localization performance was strongly dominated by the closest array to the source for multi-array processing. It was also shown that additional array did not always improve the overall localization performance. However, averaged localization error has been found to decrease as the number of array increased. For EMVDR beamformer with β=0.7, the localization error for one array was around *15 m* whereas the average localization error for three arrays decreased to less than 1 *m* for low SNRs (<0 dB). Also, multiple arrays can make the conventional DS beamformer feasible since the localization error of this beamformer was reduced from more than 25 *m* (one array) to less than 8 *m* (two arrays).

In this research, uncertainties in the model due to a temperature gradient and unmodeled objected were investigated. The main observation was that these uncertainties had detrimental effects on the performance of the beamformers. For temperature gradient uncertainty, the detrimental effects were small as the localization error increased from 2 *m* without uncertainty to around 4 *m* with uncertainty using the EMVDR beamformer with β=0.3. On the other hand, the performance degradation due to an unmodeled object such as a large vehicle was heavily dependent on the relative location of the object to the source and array. When the unmodeled object was placed very near the source or the array, the localization error increased significantly. On the other hand, when the unmodeled object was placed away from the source or the array the effect on the localization performance was negligible. Multiple array processing was then used as a method to alleviate the detrimental effect of this uncertainty. The improvement obtained by using two arrays was significant with the localization error decreased from 10 *m* to 2 *m*.

Numerical simulations showed that the fingerprinting method in conjunction with the low fidelity acoustic model (energy method) provided reliable localization capability. This approach was evaluated on two large 3D models. The results also showed a progressive improvement in the source localization capabilities as the number of microphones increased. For example, for the Fort Benning urban model the average and standard deviation of the localization error using four microphones was 23 and 16 *m*, respectively. On the other hand, using 8 microphones they were decreased to 15 and 8.7 *m*, respectively. Finally, it was shown that an extremely accurate model of the environment was not needed to implement fingerprinting method as long as enough microphones were used (i.e. the energy method provided relatively accurate fingerprints). For acoustic fingerprinting with three microphones in the simplified urban model, the localization error using the exact fingerprints was less than 5 *m* while the localization error using the approximate fingerprints were between 6 and 8 *m*, depending on the configuration.

7.2 Recommendations

The implementation of a microphone phased array based localization technique requires an accurate knowledge of the environment to account for all acoustic effects in real life. There are, however, always model mismatches between the numerically modeled and the actual environment. To account for microphone-self noise, uncorrelated noise was added to the model where different noise levels were considered. Uncertainties such as temperature gradient and unmodeled object were also investigated. Still, there are other sources of model errors that have not been investigated for more realistic validation such as wind effects. In fact, wind is likely a very important effect and very complicated to model in a cluttered environment. Therefore, it is suggested to investigate the localization techniques using numerical modeling tools that have the ability to model wind effects.

Both previous and the proposed localization techniques in this dissertation were validated only using numerical simulation. As mentioned earlier, a key requirement for these methods (beamforming and time reversal) is the need for an accurate model of the environment. Hence, the reliability of the model in real environments is a concern. Therefore, experimental validation (e.g. field test or laboratory test) should be considered as future research.

Here uncertainty studies were performed only using beamforming in conjunction with the high fidelity model. The affect of uncertainties on the localization performance of the acoustic fingerprinting in conjunction with the low fidelity model should be investigated. The effect of uncorrelated noise should also be assessed.

In this work, the localization of multiple sources was not considered. In other words, there was no investigation on how to localize more than one source at the same time. Distinct sources can potentially be confused when the sources are not far from each other. Beamformers have the inherent capability to spatially separate sources but it needs to be investigated. Also, other techniques such as time adjusted blocks of data and/or spectral filtering can be used in conjunction with the source localization methods investigated here.

Bibliography

Abraham D. A. and Owsley N., “Beamforming with dominant mode rejection,” in IEEE Oceans 1990 Conference Proceedings, pp. 470–475, 1990.

Akyildiz, I. F., Su W., Sankarasubramaniam Y. and Cayirci E., “Wireless sensor networks: A survey,” in Computer Networks, 38: 393-422. 2002.

Albert D. G., Liu L. and Moran M. L., “Time reversal processing for source location in an urban environment,” J. Acoust. Soc. Am. 118, pp. 616-619, 2005.

Albert D. G., Liu L. and Moran M. L. “The effect of buildings on acoustic pulse propagation in an urban environment.” J. Acoust. Soc. Am. 127, pp. 1335-1346, 2010.

Anderson J.B., Rappaport T. S., Yoshida S., “Propagation measurements and models for wireless communications channels.” IEEE Communications Magazine, vol. 33, no.1, pp. 42 – 49, Jan 1995.

Arora A., Dutta P., Bapat S., Kulathumani V., Zhang H., Naik V., Mittal V., Cao H., Demirbas M., Gouda M., Choi Y., Herman T., Kulkarni S., Arumugam U., Nesterenko M., Vora A., and Miyashita M., “A line in the sand: a wireless sensor network for target detection, classification, and tracking,” Computer Networks, Vol. 46, pp. 605-634, 2004.

Bahl P. and Padmanabhan V. N., “RADAR: an in-building RF-based user location and tracking system”, In Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. 2(2): 775--784, March 2000.

Baggeroer A. B., Kuperman W.A., and Mikhalevsky P.N., “An Overview of Matched Field Methods in Ocean Acoustics,” IEEE Journal of Oceanic Engineering, Vol. 18, No. 4, October 1993.

Bao, L. and Garcia-Luna-Aceves J. J., “Transmission scheduling in ad-hoc networks with directional antennas,” ACM/SIGMOBILE MobiCom Sep. 2002.

Barger J. and Stanley J., “Sensor for Airborne Shooter Localization System”, BBN Technologies, Corp., Cambridge, patent no. US 7,787,331 B2, 2008.

Billon A., Picaut J., Valeau V. Sakout A., “Numerical simulations of the sound propagation in non rectilinear streets.” Euronoise 2009, Edinburgh, Scotland, 2009.

Birtchfield S.T. and Gangishetty R., “Acoustic Localization by Interaural Level Difference,” Proc. ICASSP, 1109-1112, 2005.

Brandstein M., and Ward D., Microphone Arrays, Springer, 1995.

Broadhead M. K., Plug L. A., and Field R. L., “Use of higher order statistics in source signature estimation,” J. Acoust. Soc. Am. 107, 2576-2585, 2000.

Brooks R.R. and Iyengar S.S., “Distributed Sensor Networks” CRC press ‘Inc. pp 1188. Dec 28, 2004.

Bucker H. P., “Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,” J. Acoust. Soc. Amer. 59 (2), 368-373, 1976.

Carevica D., “Tracking target in cluttered environment using multilateral time-delay measurements,” J. Acoust. Soc. Am. 115, 1198-1206, 2001.

Caffrey J., Govindan R., Johnson E., Krishnamachari B., Masri S., and Sukhatme G., “Networked Sensing for Structural Health Monitoring.” In Proceedings of the Fourth International Workshop on Structural Monitoring and Control, June 2004.

Chan Y, Tsui Wing-Yue, Hing-Cheung So and Pak-Chung Ching., “Time-of- Arrival based localization under NLOS conditions.” In IEEE Transactions on Vehicular Technology, 55(1): 17--24, Jan. 2006.

Choi B., Pasareanu S., Burdisso R., and Inman D., “Reduced-Order High Fidelity Models for Signature Propagation: 1st Year Final Report,” VT Report to AVID-LLC, September 25, 2010.

Choi B., Pasareanu S., Burdisso R., and Inman D., “Reduced-Order High Fidelity Models for Signature Propagation: 2nd Year Progress Report,” VT Report to AVID-LLC, January 27, 2011a.

Choi B., Pasareanu S., Burdisso R., Inman D., Johnson M. and Johnson M., “Acoustic Source localization using matched-field processing-based enhanced MVDR beamforming in 3-D complex urban environment,” Noise-Con Proc. 129, 567, 2011b.

Choi B., Pasareanu S. and Burdisso R., “Source Localization with Multiple Dispersed Microphones using Acoustic Fingerprinting Technique in a 3D complex urban environment”, submitted to J. Acoust. Soc. Am, 2012.

Chong C. S. P. Kumar S. P., “Sensor Networks: Evolution, Opportunities, and Challenges,” Proceedings of the IEEE, Vol. 91, No. 8, pp. 1247-1256, August 2003.

Coope I.D., “Reliable computation of the points of intersection of n spheres in < ,” Australian, New Zealand Ind. Appl. Math. J., pt. C, vol. 42, pp. 461–477, 2000.

Corcoran J., Pasareanu S., Choi B., Burdisso R., and Inman D. “Reduced-Order High-Fidelity Models for Signature Propagation: VT Deliverable 1.2 Uncertainty Effects on FDTD Model” // VT Report to AVID-LLC, July 8, 2010.

Cox, H., “Resolving power and sensitivity to mismatch of optimum array processors,” J. Acoust. Soc. Am. 54, 771-785, 1973.

DiBiase J. H., "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," in *Division of Engineering*. vol. PhD Providence, Rhode Island: Brown University, p 112., 2000.

Drake S.P. and Dogancay K., “Geo-location by time difference of Arrival using Hyperbolic Asymptotics”, Proc. ICASSP, 361-364, 2001.

Eberly D, “Finding the intersection of three spheres”, comp.graphics.algorithms newsgroup, 1996.

Estrin D, Girod L., Pottie G., and Srivastava M., “Instrumenting the world with wireless sensor networks,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, Salt Lake City, Utah, USA, May 2001.

Ferris B., Hahnel D. and Fox. D., “Gaussian processes for signal strength-based location estimation,” In Proceedings of Robotics: Science and Systems. August 2006.

Fialkowski L., Collins M., Kuperman W., Perkins J., Kelly L., Larsson A., Fawcett J., and Hall L., "Matched-field processing using measured replica fields," J. Acoust. Soc. Am. 2000.

Fink M., “Time reversal of ultrasonic fields I. Basic principles,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control 39, 555–566, 1992.

Fink M., “Time-reversed acoustics,” Phys. Today 50, 34–40, 1997.

Fizell R. G. and Wales S. C., “Source localization range and depth in an Arctic environment,” J. Acoust. Soc. A. 78 (5), 1985.

Foy Wade H., “Position-location solutions by Taylor-series estimation.” In IEEE Transaction on Aerospace Electronic System, volume AES-12, pages 187--194, March, 1976.

Gabor D., G.W. Stroke, Proc. Roy. Soc. London A, 304, 275, 1948.

Girod L. and Estrin D., “Robust range estimation using acoustic and multimodal sensing”, In Proceedings of International Conference on Intelligent Robots and Systems, volume 3, pages 1312--1320, Oct-Nov. 2001.

Gilbert E. N., Morgan S. P., “Optimim design of directive antenna arrays subject to random variations”, Bell System Technical Journal, vol 34., May 1955.

Gourley S. R., “Soldier Armed : Ears/SWATS” , Army Magazine, available at http://www.ausa.org/publications/armymagazine/archive/2009/7/Documents/SA_0709.pdf, 2009.

Gutin G., Yeo A. and Zverovich A., “Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP,” in Discrete Applied Mathematics 117, pp 81-86., 2002.

Hanlon M., “Vehicle-mounted Acoustic Sniper Detection System”, available at http://www.gizmag.com/go/4497/ , 2005.

Helsing S., Fenno C., Ko A., Schuchardt P., Gelhausen P., Inman D., and Burdisso R. “Reduced-Order High-Fidelity Models for Signature Propagation,” STTR Phase II Final Report to the Army, Contract No. W913E5-10-C-0001, December 2012.

Iskander M.F. and Yun Z., “Propagation prediction models for wireless communication systems,” IEEE Transations on microwave theory and techniques, vol. 50, no. 3, pp. 662 – 673, March 2002.

Jing Y. and Xiang, N. "A modified diffusion equation for room-acoustic predication." Journal of Acoustical Society of America 121, no. 6, 3284-3288, 2007.

Jing Y., Larsen E. W., and Xiang N., “One-dimensional transport equation models for sound energy propagation in long spaces: theory,” Journal of the Acoustical Society of America, vol. 127, no. 4, pp. 2312–2322, 2010.

Johnson M, Carneal J., and Gillett P., “Comparison of a diffracting and a non-diffracting circular acoustic array,” ICASSP 2006, Toulouse, France, vol. 4, pp. 1081-1084, 2006.

Julliard E., Pauzin S., Simon F. and Biron D., “Acoustic sources localization in presence of reverberation”, J. Acoust. Soc. Am. Vol. 118, Issue 3, pp. 1886-1886 Sep. 2005.

Kaplan L.M., Le Q., and Molnar P., “Maximum Likelihood Methods for Bearing-Only Target Localization,” Proc. ICASSP, 001-3004, 2001.

Kozick R. J. and Sadler B. M., “Source localization with distributed sensor arrays and partial spatial coherence,” IEEE Transactions on Signal Processing, 52, 601-616, 2004.

Kuperman W. A., Hodgkiss W. S., Song H. C., Akal T., Ferla C., and Jackson D. R., “Phase conjugation in the ocean: Experimental demonstration of a time reversal mirror,” Journal of the Acoustical Society of America, vol. 103, pp. 25-40, 1998.

Ladd, A. N. and Cooper, T. A. “Multiple domains control the subcellular localization and activity of ETR-3, a regulator of nuclear and cytoplasmic RNA processing events”. J Cell Sci 117: 3519-3529. 2004.

LaMarca, A., Hightower J., Smith I and Consolvo S., “Self-mapping in 802.11 location systems,” In International Conference on Ubiquitous Computing (UbiComp), 2005.

Li Dan, Yu Hen Hu, “Energy-Based Collaborative Source Localization Using Acoustic Microsensor Array”, EURASIP Journal on Applied Signal Processing 321-337, 2003.

Liu L., Albert D. G., and Wilson D. K., “The effect of changing scatterer positions on acoustic time-reversal refocusing in a 2D urban environment at low frequencies,” Journal of Geophysical Engineering 4, pp. 276-284, 2007.

Le P., Bardos T.C., "Sound field modeling in a street canyon with partially diffusely reflecting boundaries by the transport theory." Journal of Acoustical Society of America 116, no. 5, pp. 2969-2983, 2004.

Leveque R.J., “Finite Difference Methods for Ordinary and Partial Differential Equations.” , Society for Industrial and Applied Mathematics, 2007.

Mak L.C. and Furukawa T., "A Time-of-Arrival-Based Positioning Technique With Non-Line-of-Sight Mitigation Using Low-Frequency Sound," *Advanced Robotics*, vol. 22, no. 5, pp. 507-526, 2008.

Mak L.C. and Furukawa T., "Non-Line-of-Sight Localization of a Controlled Sound Source," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Singapore, Singapore, Jul. 14-17, pp. 475-480, 2009.

Maroti M., Simon G., Ledeczi A. and Sztipanovits J., “Shooter Localization in Urban Terrain,” Computer, 37, 8, pp. 60-61, August 2004.

Mazerolle L.G., J. Frank, Rogan D., and Watkins C., “Field Evaluation of the ShotSpotter Gunshot Location System: Final Report on the Rewood City Field Trial,” submitted to the U.S. Department of Justice, Document No. 180112, January 2000.

Meesookho, C. and Mitra U., “On Energy-based Acoustic Source Localization for Sensor Networks”, IEEE Trans. On Signal Processing. Vol 56. No 1. Jan 2008.

Mennit D., “Multiarray passive acoustic localization and tracking,” PhD dissertation,Virginia Polytechnic Institute and State University, Blacksburg, Virginia. 2008.

Mennit D. and Johnson M., “Multiarray passive acoustic localization in urban environments,” J. Acoust. Soc. Am. 127(5), 2932-2942. 2010.

Messerschmitt, T. R., and Gramann, R. A., “Evaluation of the dominant mode rejection beamformer using reduced integration times,” IEEE Journal of oceanic engineering, Vol. 22, n 2, pp 385-392. April 1997.

Mulgrew B. and Chen S., “Adaptive minimum-BER decision feedback equalizers for binary signaling,” Signal Process., vol. 81, no. 7, pp. 1479–1489, 2001.

Navarro G. and Makinen V., “Compressed full-text indexes”. ACM Computing Surveys, 39(1), 2009.

Nerguizian C., Despins C. and Affes S., “Geolocation in mines with an impulse response fingerprinting technique and neural networks.” In IEEE Transactions Wireless Communications, 5(3): 603--611, March 2006.

Neskovic A., Neskovic N., and Paunovic G., “Modern approaches in modeling of mobile radio systems propagation environment,” IEEE Communications Surveys and Tutorials, vol. 3, no. 3, 2000.

Nuttall A.H. and Hyde D.W., “Adaptive beamforming at very low frequencies in spatially coherent, cluttered noise environments with low signal-to-noise and finite-averaging times,” J. Acoustic. Soc. Am. 2256-2265, 1969.

Ostashev V.E., Wilson D.K., Liu L., Aldrige D.F., Symons N.P., and Marlin D., “Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation”, J. Acous. Soc. Am. 503-517, 2005.

Omologo. M. and Svaizer P., "Acoustic source location in noisy and reverberant environment using CSP analysis," icassp, vol. 2, pp.921-924, Acoustics, Speech, and Signal Processing, 1996.

Ott. H., “Noise Reduction Techniques in Electronic Systems”, John Wiley & Sons, 1976.

Picaut. J, Simon. L., and Hardy. J. “Sound field in streets with a diffusion equation” J. Acous. Soc. Am. 106(5), 2638-2645, 1999.

Pasareanu S., Remillieux M., and Burdisso R., “Energy-based method for near real-time modeling of sound field in complex urban environments” J. Acus. Soc. Am. 130(4), 2367, 2011.

Paul A., and Wan E., “RSSI-based indoor localization and tracking using sigma-point Kalman smoothers,” in IEEE Journal of Selected Topics in Signal Processing, 3(5), 2009.

Rappaport T.S. “Wireless Communication: Principles and Practice”, Prentice Hall, 1996.

Robinson E.A. and Treitel S., “Geophysical signal analysis”, New York, Prentice Hall. pp. 81-86 , 1967.

Sasaki Y., Kagami S., and Mizoguchi H., "Multiple sound source mapping for a mobile robot by self-motion triangulation," in IEEE/RSJ International Conference on Intelligent Robots, Beijing, China, pp. 380-385. 2006.

Savarese C., Rabaey J. M., and Beutel J., “Locationing in distributed Ad-hoc wireless sensor networks,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, Salt Lake City, Utah, USA, May 2001.

Schwaighofer A., Grigoras M., Tresp V., and Hoffmann C., “GPPS: A Gaussian process positioning system for cellular networks.” In Advances in Neural Information Processing Systems (NIPS), 2003.

Seidel S. and Rappaport T, “914 MHz path loss prediction models for indoor wireless communications in multifloored buildings”, In IEEE Transactions on Antennas and Propagation, 40(2): 207--217, Feb 1992.

Sheng X. and Yu-Hen Hu, “Maximum Likelihood Multiple-Source Localization Using Acoustic Energy Measurements with Wireless Sensor Networks”, IEEE Trans. On Signal Processing, Vol 53. No 1. Jan 2005.

Stephenne A. and Champagne B., “Cepstral pre-filtering for time delay estimation in reverberant environments”, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP, pp. 3081-3084, 1999.

Svaizer P., Matassoni M. and Omologo M., “Acoustic Source Location in a Three-Dimensional Space Using Crosspower Spectrum Phase”, Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97) -Volume 1, p.231, April 21-24, 1997.

Szewczyk R., Mainwaring A., Polastre J., Anderson J., and Culler D., "An Analysis of a Large Scale Habitat Monitoring Application," SenSys "04, Baltimore, Maryland, USA, November 2004.

Tanter M., Bercoff J., Athanasiou A., Deffieux T., Gennisson J. L., Montaldo G., Muller M., Tardivon A., and Fink M., "Quantitative Assessment of Breast Lesion Viscoelasticity: Initial Clinical Results Using Supersonic Shear Imaging", Ultrasound in Medicine & Biology, Vol 34, pp 1373-1386, Sep. 2008.

Tiemann A. M., Thode J. Straley, O'Connell V., and Folkert K., "Three-dimensional localization of sperm whales using a single hydrophone," Journal of the Acoustical Society of America, Vol. 120, No. 4, pp. 2355-2365, Oct. 2006.

Tolan J. G. and Schneider J. B., "Locally conformal method for acoustic finite-difference time-domain modeling of rigid surface," Journal of the Acoustical Society of America, vol. 115, no. 5, pp. 2575-2581, 2003.

Thomas J. L., Wu F., Fink M, "Time reversal focusing applied to lithotripsy", Ultrason Imaging, Vol 18, pp 106-21, 1996.

Tolstoy A., "Matched Field Processing for Underwater Acoustics", World Scientific Publishing, Singapore, 1993.

Valin J.M., Michaud F., and Rouat J., "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," Robotics and Autonomous Systems, vol. 55, pp. 216-228, 2007.

Van Veen B.D., and Buckley K. M., "Beamforming: A versatile approach to spatial filtering", IEEE ASSP Mag., pp. 4-24, Apr. 1988.

Wang. X., Wang Z., and O'Dea B., "A TOA-based location algorithm reducing the errors due to nonline-of-sight (NLOS) propagation," *IEEE Transactions on Vehicular Technology,* vol. 52, pp. 112-116, 2003.

Wang W., Yu J., Xiong and Zhu Z. L., "A new NLOS mitigation algorithm in location estimation." In IEEE Transaction on Vehicular Technology, 54(6): 2048--2053, Nov 2005.

Whipps, G. T., Kaplan L. M., and Damarla R, "Analysis of sniper localization for mobile asynchronous sensor," in Proc. Of SPIE, vol. 7336. 2009.

Yee K.,"Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, " IEEE Trans. Antennas Propag. 14, 302-307, 1996.

Yon S., Tanter M., and Fink M., "Sound focusing in rooms. II. The spatio-temporal inverse filter," J. Acoust. Soc. Am. Vol. 114, pp. 3044–3052, 2003.

Yuan X., D. Borup, J. W. Wiskin, M. Berggren, R. Eidens, and S. B. Johnson, "Formulation and validation of Berenger's PML absorbing boundary for the FDTD simulation of acoustic scattering," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 44, no. 4, pp. 816-822, 1997.

Appendix A Mathematical Derivation of the *WNG*

A.1. Derivation of *WNG*

Assuming stationary signals, the cross spectral matrix can be may be modeled as,

$$\boldsymbol{\Phi}_{RR} = \sigma_s{}^2 \boldsymbol{d}\boldsymbol{d}^H + \sigma_n{}^2 I_N \tag{A.1.1}$$

where $\sigma_s{}^2$ and $\sigma_n{}^2$ are the signal and noise powers, respectively and $\boldsymbol{d}$ is a vector whose elements are the transfer function between the source and the microphones in the array. This vector is normalized so it has a unit length.

The input SNR (in dB) is then,

$$SNR_{in} = 10\log_{10}\frac{\sigma_s{}^2}{\sigma_n{}^2} \tag{A.1.2}$$

Similarly, the beamformer output due the array may be written as,

$$\boldsymbol{b}\text{=}\boldsymbol{w}^H \boldsymbol{\Phi}_{RR}\boldsymbol{w} = \sigma_s{}^2 \boldsymbol{w}^H \boldsymbol{d}\boldsymbol{d}^H \boldsymbol{w}\text{+}\sigma_n{}^2 \boldsymbol{w}^H \boldsymbol{w} \tag{A.1.3}$$

where w is the weighting vector of the corresponding beamformer.

The output SNR (in dB) is then given by,

$$SNR_{out} = 10\log_{10}\frac{\sigma_s{}^2 \left|\boldsymbol{w}^H \boldsymbol{d}\right|^2}{\sigma_n{}^2 \boldsymbol{w}^H \boldsymbol{w}} \tag{A.1.4}$$

The *WNG* (in dB) is a value that shows the improvement of the SNR due to the beamforming. It is defined as,

$$WNG = SNR_{output} - SNR_{input} = 10\log_{10}\frac{\left|\boldsymbol{w}^{H}\boldsymbol{d}\right|^{2}}{\boldsymbol{w}^{H}\boldsymbol{w}} \tag{A.1.5}$$

A.2. Derivation of *WNG* for Mennitt-MVDR beamformer

The weighting vector of the EMVDR beamformer is,

$$\boldsymbol{w} = \frac{\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}{\sqrt{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}} \tag{A.2.1}$$

Substituting eq. (A.2.1) into (A.1.5) gives,

$$WNG = 10\log_{10}\frac{\left|\left(\dfrac{\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}{\sqrt{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}}\right)^{H} d\right|^{2}}{\left(\dfrac{\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}{\sqrt{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}}\right)^{H}\left(\dfrac{\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}{\sqrt{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}}\right)} \tag{A.2.2}$$

Now, eq. (A.2.2) can be reduced to,

$$WNG = 10\log_{10}\frac{\dfrac{\left(\boldsymbol{d}^{H}\left(\boldsymbol{\Phi}_{VV}{}^{-1}\right)^{H}\boldsymbol{d}\right)^{2}}{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}}{\dfrac{\boldsymbol{d}^{H}\left(\boldsymbol{\Phi}_{VV}{}^{-1}\right)^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}{\boldsymbol{d}^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}}} = 10\log_{10}\frac{\left(\boldsymbol{d}^{H}\left(\boldsymbol{\Phi}_{VV}{}^{-1}\right)^{H}\boldsymbol{d}\right)^{2}}{\boldsymbol{d}^{H}\left(\boldsymbol{\Phi}_{VV}{}^{-1}\right)^{H}\boldsymbol{\Phi}_{VV}{}^{-1}\boldsymbol{d}} \tag{A.2.3}$$

Appendix B Mathematical Derivation of EMVDR beamformer

B.1. Derivation of EMVDR beamformer

The array cross spectral matrix (CSM) with spatially uncorrelated noise has the following form

$$\boldsymbol{\Phi}_{RR} = E[\boldsymbol{p}\boldsymbol{p}^H] + {\sigma_n}^2 I_N \tag{B.1.1}$$

where $\boldsymbol{p}$ is the vector of the Fourier Transform of the microphone signals due to the source, ${\sigma_n}^2$ is the uncorrelated noise variance (same for all microphone), I_N is the N-by-N identity matrix, and N is the number of microphones in the array, e.g. 4 for the 2D case.

The eigen-decomposition of the CSM due to the source signal only in (4.19) is

$$\boldsymbol{\Phi}_{RR} = \sum_{i=1}^{N} \lambda_i \boldsymbol{m}_i \boldsymbol{m}_i^H \tag{B.1.2}$$

where λ_i are the eigenvalues and $\boldsymbol{m}_i$ are the eigenvectors. Ordering the eigenvalues from high to low $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, it assumed that the first D eigenvalues are due to sound sources, and the rest due to weak sources and spatially uncorrelated noises. For the 2D simulations performed, there is only one source in the environment and the array has 4 microphones, $N = 4$. Thus, this is a single dominant eigenvalue, λ_1. The enhanced MVDR (EMVDR) beamformer was proposed by Abraham and Owsley (1990) to avoid the need to invert the CSM in the MVDR algorithm. To this end, the CSM is approximated using only the dominant mode as follows

$$\boldsymbol{\Phi}_{RR} \approx \hat{\boldsymbol{\Phi}}_{RR}(e) = e\,\psi_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H + \sigma_n^2 I_N \tag{B.1.3}$$

where $\psi_i = \lambda_i - \sigma_n^{\ 2}$ and e is a scalar enhancement factor. Note that the method requires knowing the uncorrelated noise variance $\sigma_n^{\ 2}$.

To implement the MVDR using the approximate CSM in (4.21), we need to find its inverse. To this end we use the ***Sherman-Morrison Formula*** (Golub and Van Loan, 1996). Let u and v be two arbitrary vectors and A an invertible square matrix. Then Sherman-Morrison Formula states that

$$\left(A + uv^H\right)^{-1} = A^{-1} - \frac{A^{-1}uv^H A^{-1}}{1 + v^H A^{-1} u} \tag{B.1.4}$$

Then, taken $A = \sigma_n^{\ 2} I_N$, $u = \boldsymbol{m}_1$, and $v = \boldsymbol{m}_1^H$ results

$$\hat{\boldsymbol{\Phi}}_{RR}(e)^{-1} = \left(\sigma_n^{\ 2} I_N + e\,\psi_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H\right)^{-1} \tag{B.1.5}$$

or

$$= \sigma_n^{\ -2} I - \frac{\sigma_n^{\ -2} I\left(e\,\psi_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H\right)\sigma_n^{\ -2} I}{1 + \boldsymbol{m}_1^H\left(\sigma_n^{\ -2} I\right) e\,\psi_1 \boldsymbol{m}_1} = \sigma_n^{\ -2}\left(I - \frac{e\,\psi_1}{\sigma_n^{\ 2} + e\psi_1}\boldsymbol{m}_1 \boldsymbol{m}_1^H\right) = \sigma_n^{\ -2}\left(I - \beta_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H\right) \tag{B.1.6}$$

where $\beta_1 = \dfrac{e\psi_1}{\sigma_n^{\ 2} + e\psi_1}$

The weighting vector is now obtained by placing the approximated inverse from (6) into the conventional MVDR weight vector formula,

$$\boldsymbol{w}_j = \frac{\hat{\boldsymbol{\Phi}}_{RR}^{\ -1} \boldsymbol{g}_j}{\boldsymbol{g}_j^H \hat{\boldsymbol{\Phi}}_{RR}^{\ -1} \boldsymbol{g}_j} \tag{B.1.7}$$

where $\boldsymbol{g}_j$ is the normalized steering vector at the j^{th} grid point, e.g. $\boldsymbol{g}_j^H \boldsymbol{g}_j = 1$.

Thus,

$$\boldsymbol{w}_j = \frac{\hat{\boldsymbol{\Phi}}_{RR}^{\ -1} \boldsymbol{g}_j}{\boldsymbol{g}_j^H \hat{\boldsymbol{\Phi}}_{RR}^{\ -1} \boldsymbol{g}_j} = \frac{\left(I_N - \beta_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H\right)\boldsymbol{g}_j}{\boldsymbol{g}_j^H\left(I_N - \beta_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H\right)\boldsymbol{g}_j} = \frac{\boldsymbol{g}_j - \beta_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{g}_j}{1 - \beta_1\left|\boldsymbol{m}_1^H \boldsymbol{g}_j\right|^2} \tag{B.1.8}$$

Note that $\beta = 1$ when the dominant mode is maximized (i.e. $e \to \infty$).

B.2 Derivation of *WNG* for EMVDR beamformer

The weighting vector of the EMVDR beamformer is,

$$\boldsymbol{w} = \frac{\boldsymbol{g} - \beta_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{g}}{1 - \beta_1 \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2} \tag{B.2.1}$$

Substituting eq. (B.2.1) into (A.1.5) gives,

$$WNG = 10\log_{10} \frac{\left(\dfrac{\boldsymbol{g}^H \boldsymbol{d} - \beta_1 \boldsymbol{g}^H \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{d}}{1 - \beta_1 \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2}\right)\left(\dfrac{\boldsymbol{g}^H \boldsymbol{d} - \beta_1 \boldsymbol{g}^H \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{d}}{1 - \beta_1 \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2}\right)}{\left(\dfrac{\boldsymbol{g}^H - \beta_1 \boldsymbol{g}^H \boldsymbol{m}_1 \boldsymbol{m}_1^H}{1 - \beta_1 \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2}\right)\left(\dfrac{\boldsymbol{g} - \beta_1 \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{g}}{1 - \beta_1 \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2}\right)} \tag{B.2.2}$$

Since, $\boldsymbol{g}^H \boldsymbol{d} = \sqrt{M}$ and $\boldsymbol{g}^H \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{d} = \dfrac{\left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2}{\sqrt{M}}$, eq. (B.2.2) can be reduced to,

$$WNG = 10\log_{10} \frac{\left(\sqrt{M} - \beta \dfrac{\left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2}{\sqrt{M}}\right)^2}{\left(\boldsymbol{g}^H - \beta \boldsymbol{g}^H \boldsymbol{m}_1 \boldsymbol{m}_1^H\right)\left(\boldsymbol{g} - \beta \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{g}\right)} = 10\log_{10} \frac{\dfrac{1}{M}\left(M - \beta \left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2\right)^2}{\boldsymbol{g}^H \boldsymbol{g} - 2\beta \boldsymbol{g}^H \boldsymbol{m}_1 \boldsymbol{m}_1^H \boldsymbol{g} + \left(\beta \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|\right)^2}$$

$$= 10\log_{10} \frac{\left(M - \beta \left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2\right)^2}{M\left(1 - 2\beta \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2 + \beta^2 \left|\boldsymbol{m}_1^H \boldsymbol{g}\right|^2\right)} \tag{B.2.3}$$

$$\xrightarrow{\beta=1} 10\log_{10} \frac{\left(M - \left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2\right)^2}{M - \left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2} = 10\log_{10}\left(M - \left|\boldsymbol{m}_1^H \boldsymbol{d}\right|^2\right) \tag{B.2.4}$$

When $\beta = 0$ which is the case for the conventional DS beamformer, the *WNG* reaches to $10\log_{10}M$ (the optimal) as expected.

Appendix C Additional Multi-Array Localization Maps

This appendix shows the localization map (Top view) for the Fort Benning model using the EMVDR beamformer for multiple source locations and multiple arrays. Here, the results for two source locations are presented in conjunction with the number of arrays ranging from a single one to four arrays. Each source can be localized with one of 15 array combinations, i.e. 4 using single arrays, 6 using two arrays, 4 using 3 arrays, and 1 using all four arrays.

C.1. Source Location 1

The performance of the multi-array EMVDR beamformer with source location at (5, 72, 2) *m* is shown in Figures C.1.1 through C.1.4. Figure C.1.1 shows results for the 4 single array cases. This figure clearly shows that the localization capability depends on the amount of clutter between the source and the array. The array nearest to the source (array 2) can effectively locate the source as well as outperforms the others. The array 2 is also capable to locate the source but with higher degree of ambiguity. Given the increase tortuosity of the noise path towards arrays 3 and 4 as well as the inferior SNR, they perform poorly.

Figure C.1.2 shows the results for the possible 6 array pair cases. In all cases, the performance improves dramatically. Moreover, virtually all array pair provides a very reliable estimate of the source location, except for the pair 3 and 4, as expected. Figure C.1.3 presents the 4 three-arrays cases which show that any three-array system locates the source very effectively. Finally, the case of using all four arrays is shown in Figure C.1.4. As expected, no significant improvement is observed over the three-array results in Figure C.1.3. It is important to note that in many of the cases using multiple arrays that included array 2, the beamforming maps resembles very much the case of the single array 2. This suggests that the beamforming map is strongly dominated by the array 2 results.

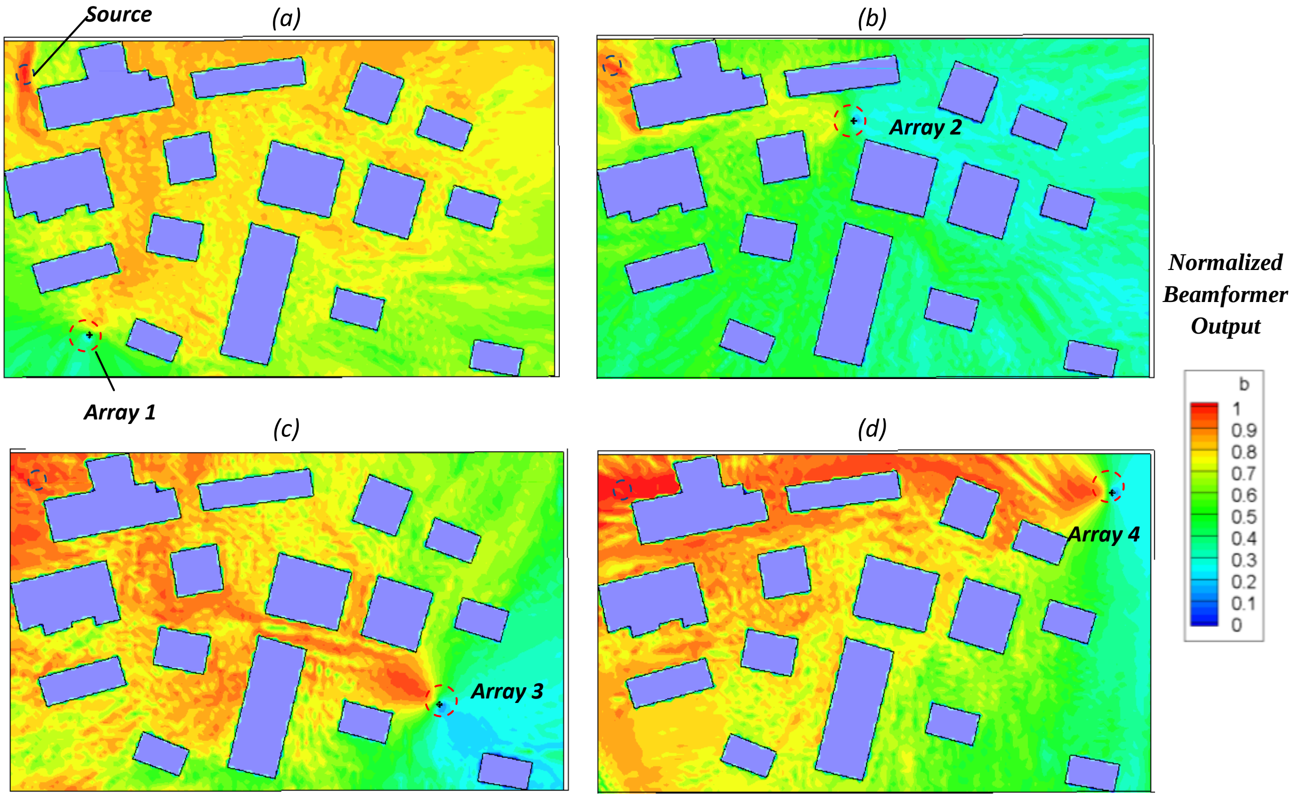


Figure C.1.1: Localization map (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for single array cases whose SNRs are approximately (a) 5dB, (b) 8dB, (c) 2dB and (d) 0dB, respectively. Source is at (5, 72, 2) m.

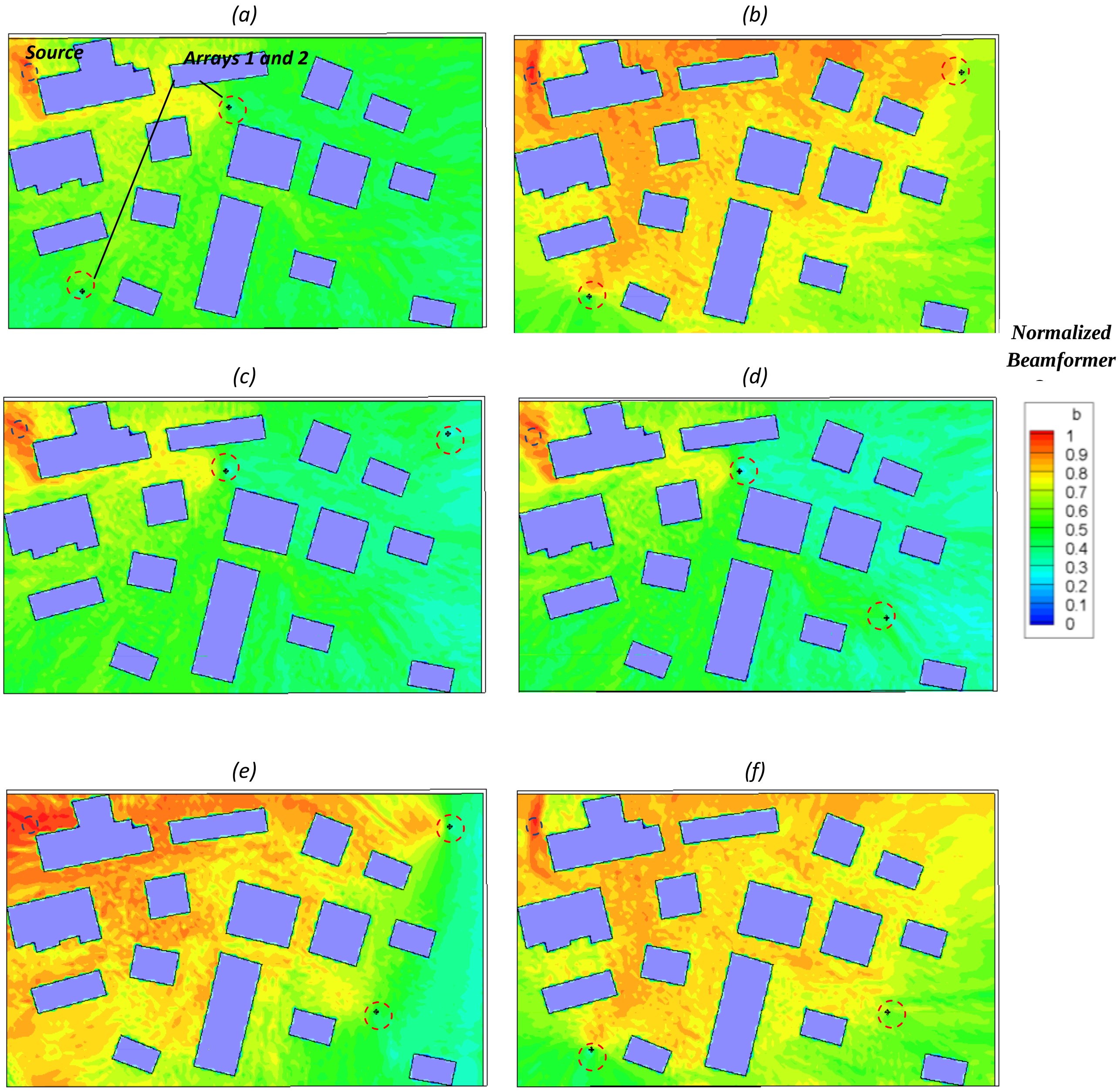


Figure C.1.2: Localization map (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for two-array cases whose SNRs are the same as in Figure C.1.1. Source is at (5, 72, 2) m.

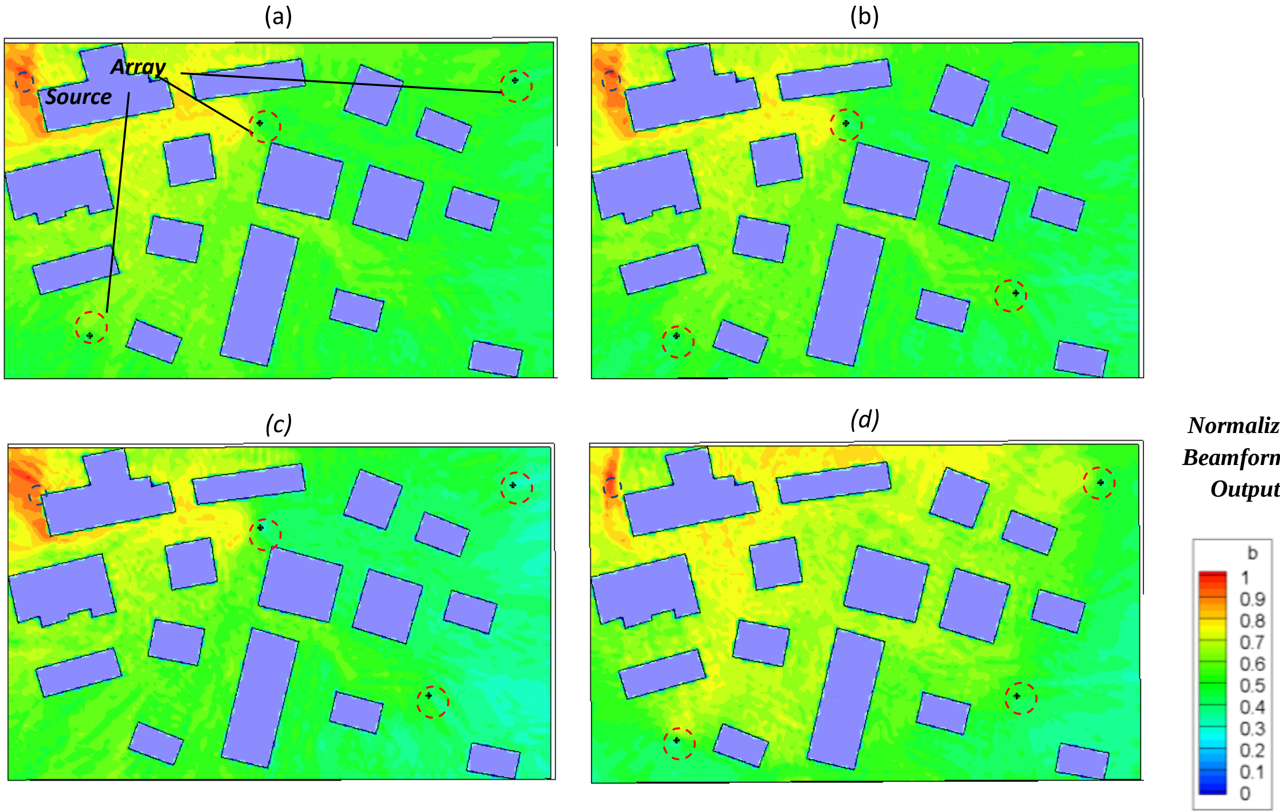


Figure C.1.3: Localization maps (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for three-array cases whose SNRs are the same as in Figure C.1.1. Source is at (5, 72, 2) m.

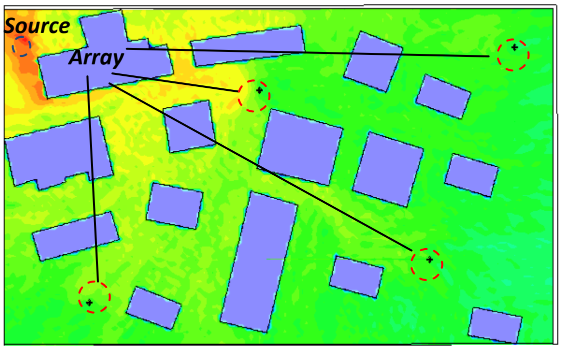


Figure C.1.4: Localization map (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for four-array case whose SNRs are the same as in Figure C.1.1. Source is at (5, 72, 2) m.

C.2. Source Location 2

The performance of the multi-array EMVDR beamformer with source location at (50, 20, 2) *m* is shown in Figures C.2.1 through C.2.4. Figure C.2.1 shows results for the 4 single array cases. The array 2 effectively locates the source as well as outperforms the others since it has direct LoS from the source. Given the increase tortuosity of the noise path towards arrays 3 and 4 as well as the inferior SNR, they perform poorly.

Figure C.2.2 shows the results for the possible 6 array pair cases. The performance improves dramatically when the array 2 is included. Figure C.2.3 presents the 4 three-arrays cases which show that any three-array system locates the source very effectively expect the case where the array 2 is not included. Finally, the case of using all four arrays is shown in Figure C.2.4. As expected, no significant improvement is observed over the three-array results in Figure C.2.3. It is important to note that in many of the cases using multiple arrays that included array 2, the beamforming maps resembles very much the case of the single array 2. This suggests that the beamforming map is strongly dominated by the array 2 results due to the direct LoS from the source.

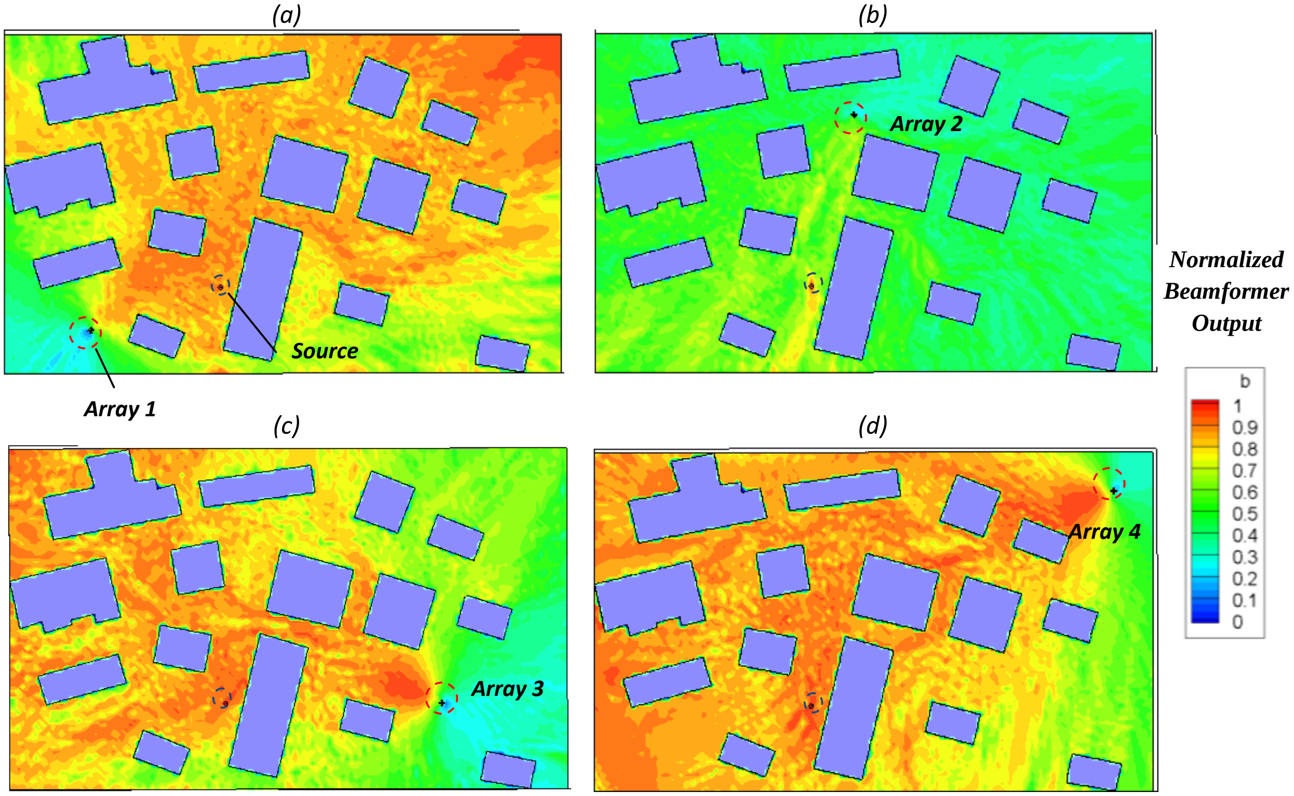


Figure C.2.1: Localization map (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for single array cases whose SNRs are approximately (a) 3dB, (b) 10dB, (c) 2dB and (d) 0dB, respectively. Source is at (50, 20, 2) m.

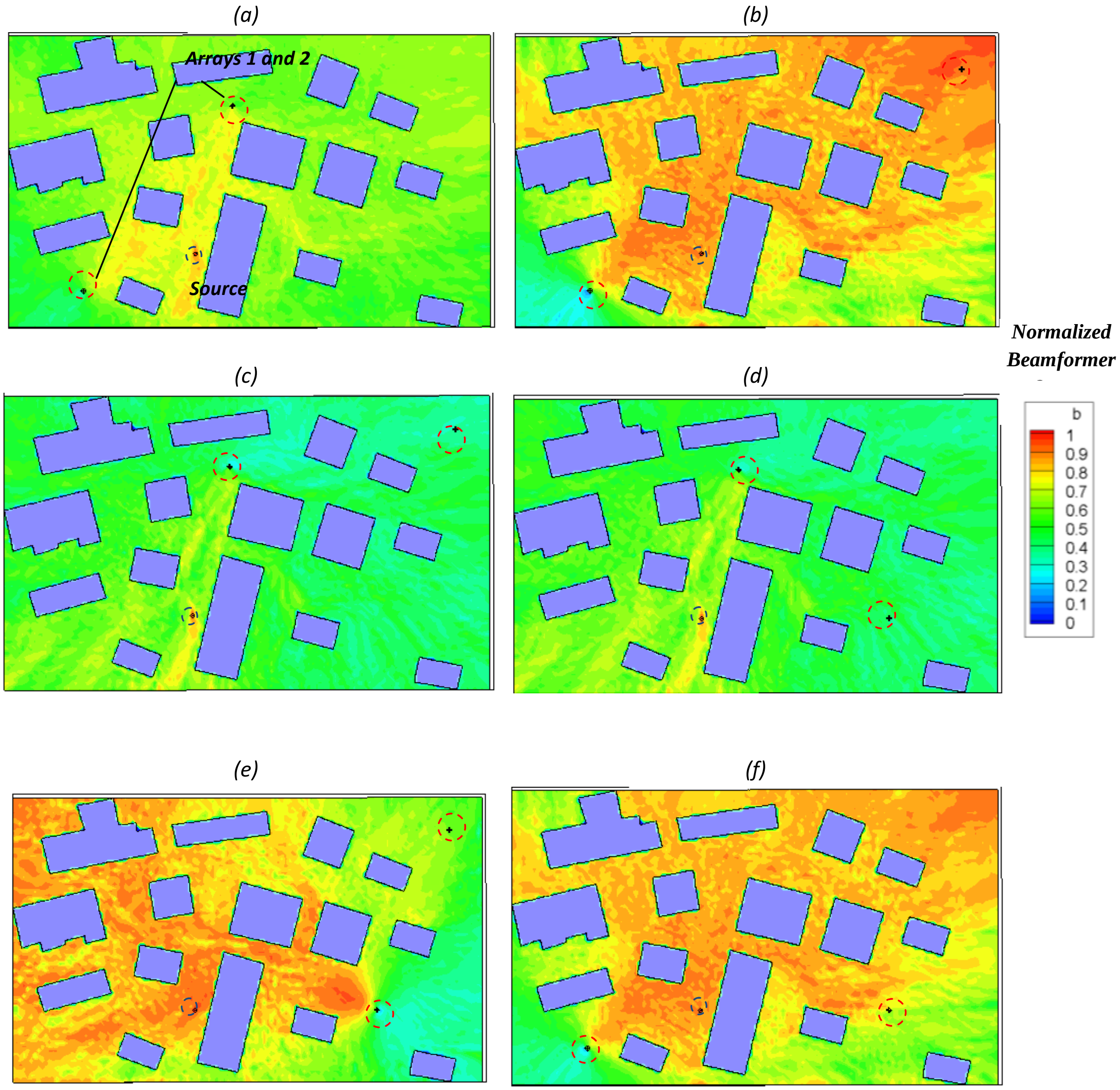


Figure C.2.2: Localization map (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for two-array cases whose SNRs are the same as in Figure C.2.1. Source is at (50, 20, 2) m.

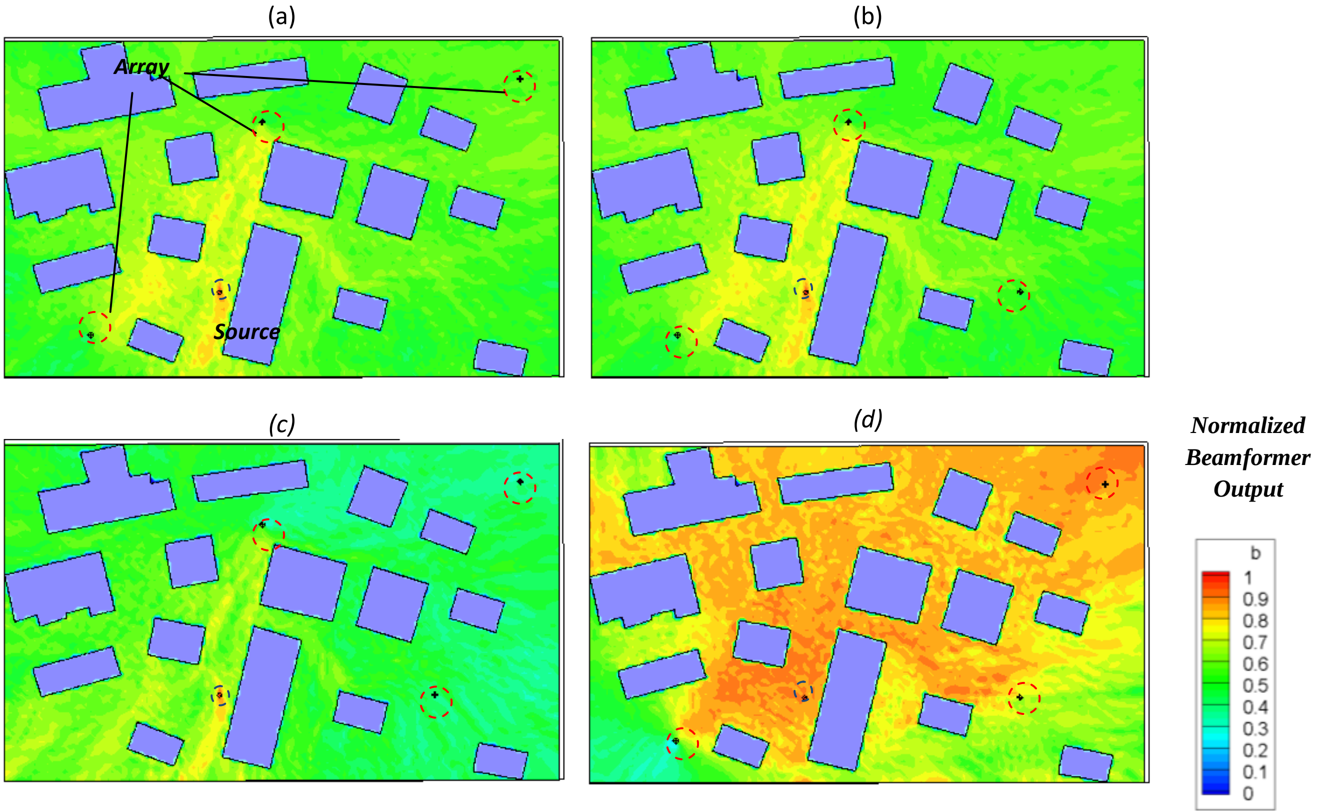


Figure C.2.3: Localization maps (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for three-array cases whose SNRs are the same as in Figure C.2.1. Source is at (50, 20, 2) m.

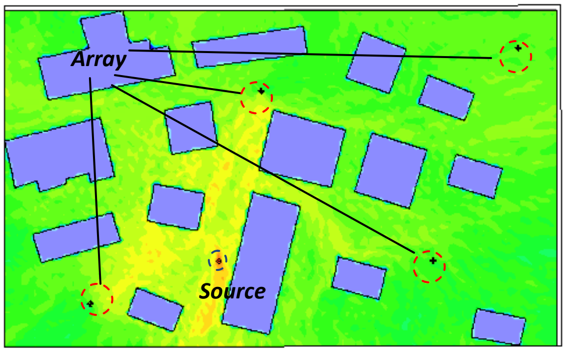


Figure C.2.4: Localization map (top-view) for the Fort Benning urban model using EMVDR (β= 0.2) beamformer for four-array case whose SNRs are the same as in Figure C.2.1. Source is at (50, 20, 2) m.

Appendix D Localization Maps using Acoustic Fingerprinting for Fort Benning Model

This appendix shows the localization map (top view) for the Fort Benning model using acoustic fingerprinting in conjunction with the energy method for several combinations whose number of microphones ranges from five to nine. Source location is at (52, 3, 2) *m*.

5 Microphones

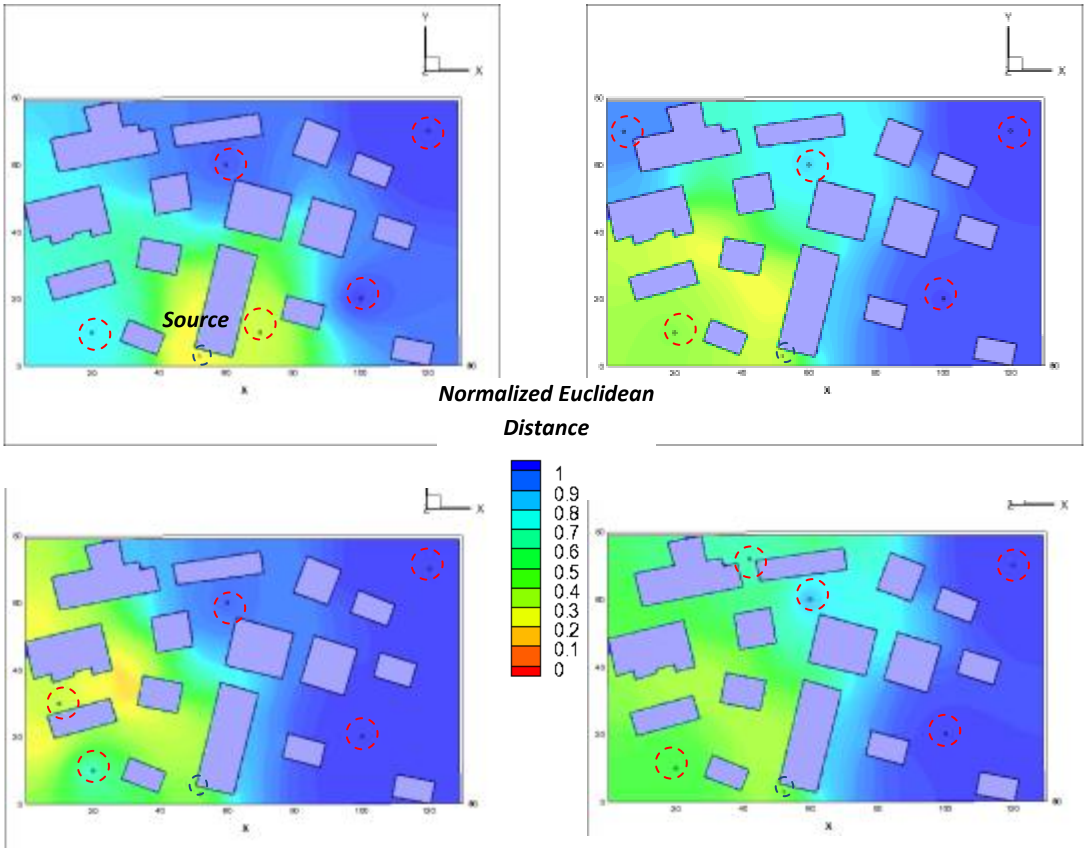


Figure D.1: Localization map (Top view) for the Fort Benning urban model with five microphones using method 1. Source is at (52, 3, 2) *m*.

6 Microphones

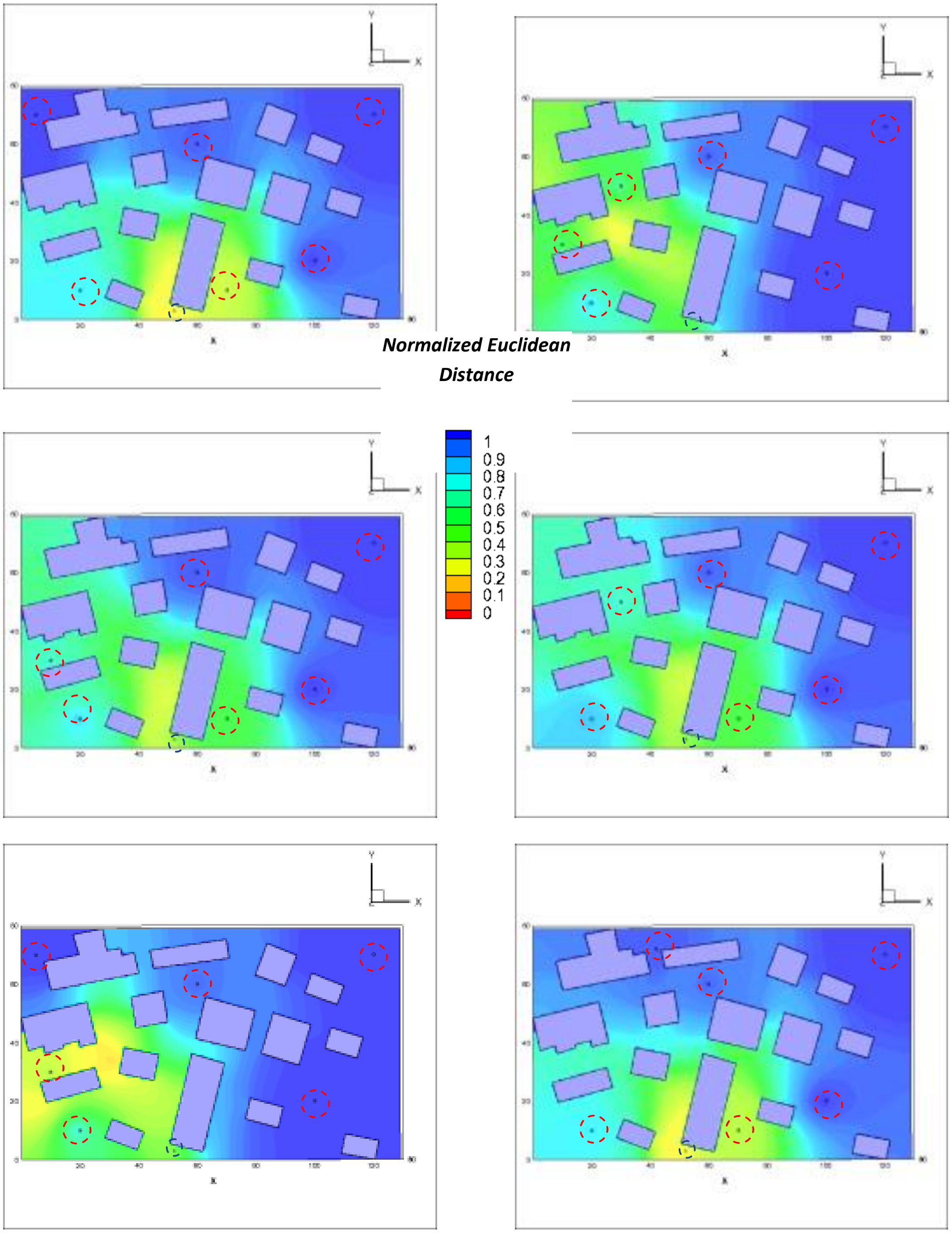


Figure D.2: Localization map (Top view) for the Fort Benning urban model with six microphones using method 1. Source is at (52, 3, 2) *m*.

7 Microphones

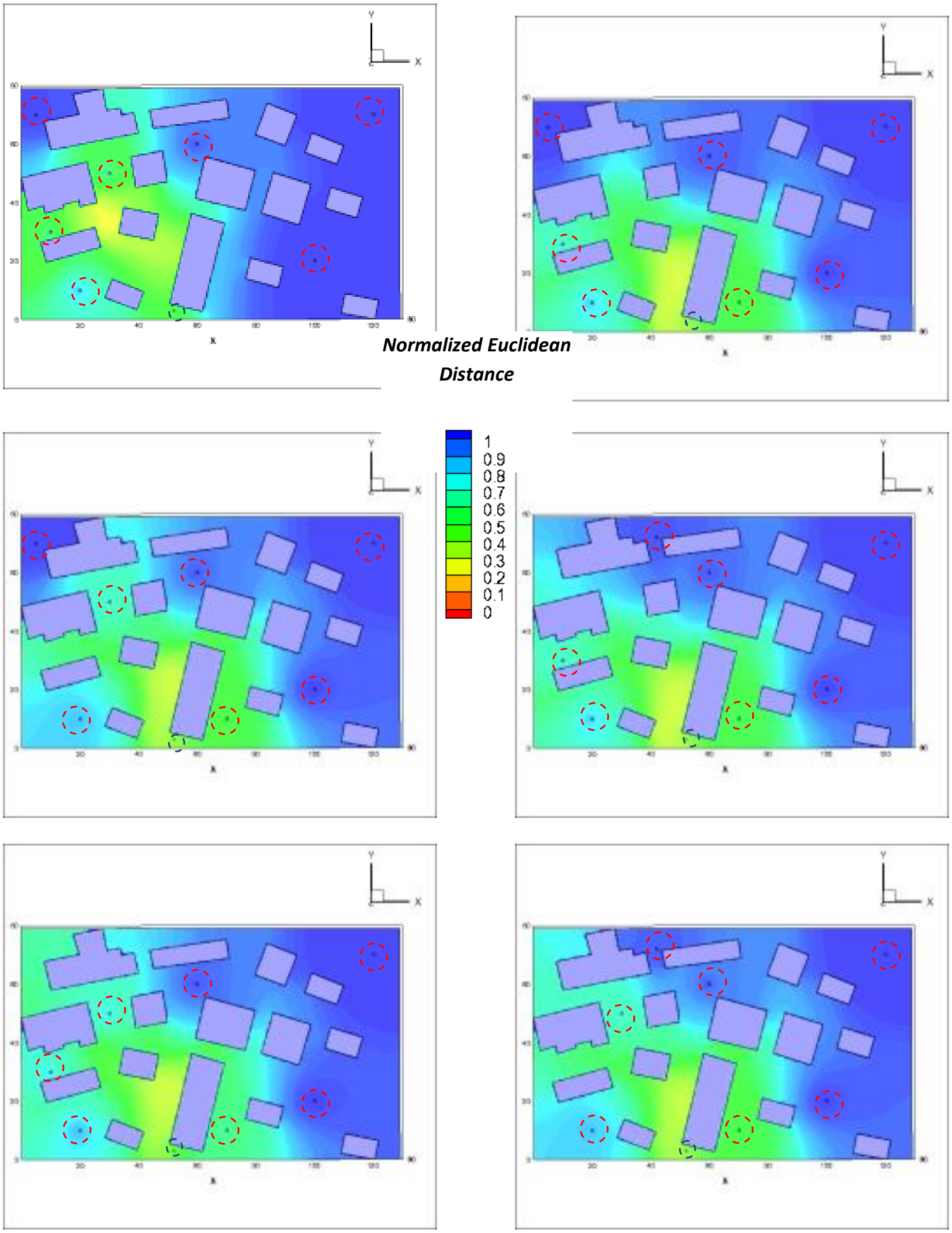


Figure D.3: Localization map (Top view) for the Fort Benning urban model with seven microphones using method 1. Source is at (52, 3, 2) *m*.

8 Microphones

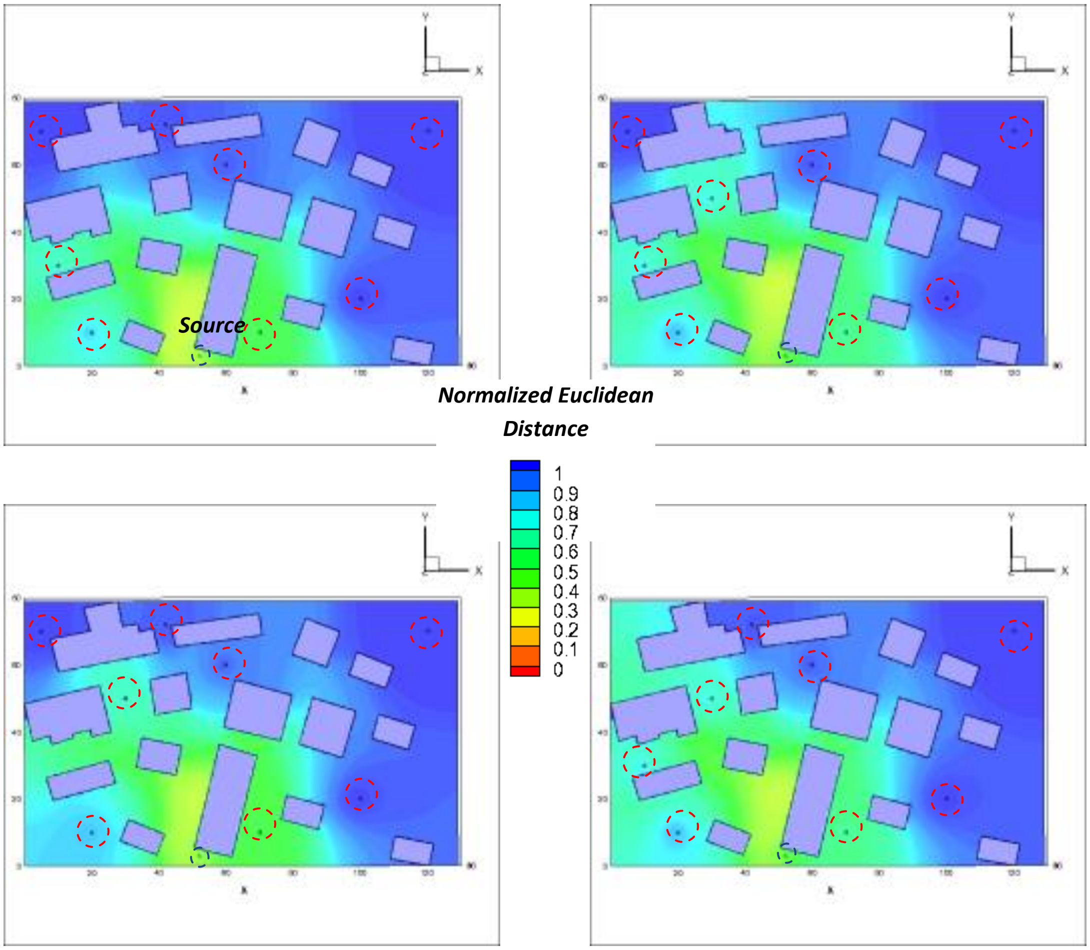


Figure D.4: Localization map (Top view) for the Fort Benning urban model with eight microphones using method 1. Source is at (52, 3, 2) *m*.

9 Microphones

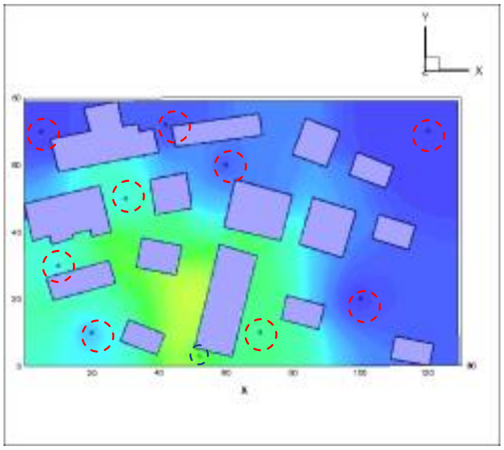


Figure D.5: Localization map (Top view) for the Fort Benning urban model with nine microphones using method 1. Source is at (52, 3, 2) *m*.