2010

STATISTICAL MODELS FOR CONSTANT FALSE-ALARM RATE THRESHOLD ESTIMATION IN SOUND SOURCE DETECTION SYSTEMS

Sayed Mahdi Saghaian Nejad Esfahani
University of Kentucky, smsa222@uky.edu

ABSTRACT OF THESIS

STATISTICAL MODELS FOR CONSTANT FALSE-ALARM RATE THRESHOLD ESTIMATION IN SOUND SOURCE DETECTION SYSTEMS

Constant False Alarm Rate (CFAR) Processors are important for applications where thousands of detection tests are made per second, such as in radar. This thesis introduces a new method for CFAR threshold estimation that is particularly applicable to sound source detection with distributed microphone systems. The novel CFAR Processor exploits the near symmetry about 0 for the acoustic pixel values created by steered-response coherent power in conjunction with a partial whitening preprocessor to estimate thresholds for positive values, which represent potential targets.

To remove the low frequency components responsible for degrading CFAR performance, fixed and adaptive high-pass filters are applied. A relation is proposed and it tested the minimum high-pass cut-off frequency and the microphone geometry.

Experimental results for linear, perimeter and planar arrays illustrate that for desired false alarm (FA) probabilities ranging from 10^{-1} and 10^{-6}, a good CFAR performance can be achieved by modeling the coherent power with Chi-square and Weibull distributions and the ratio of desired over experimental FA probabilities can be limited within an order of magnitude.

KEYWORDS: Sound Source Localization, CFAR Processor, High-pass Filter, Chi-square Distribution, Weibull Distribution

Sayed Mahdi Saghaian Nejad Esfahani

April 19th, 2010

STATISTICAL MODELS FOR CONSTANT FALSE-ALARM RATE THRESHOLD ESTIMATION IN SOUND SOURCE DETECTION SYSTEMS

By

Sayed Mahdi Saghaian Nejad Esfahani

Kevin D. Donohue
Director of Thesis

Stephen D. Gedney
Director of Graduate Studies

April 19th, 2010

RULES FOR THE USE OF THESES

Unpublished theses submitted for the Master's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the thesis in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this thesis for use by its patrons is expected to secure the signature of each user.

Name	Date

THESIS

Sayed Mahdi Saghaian Nejad Esfahani

The Graduate School

University of Kentucky

2010

STATISTICAL MODELS FOR CONSTANT FALSE-ALARM RATE THRESHOLD ESTIMATION IN SOUND SOURCE DETECTION SYSTEMS

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in Electrical Engineering in the
College of Engineering at the University of Kentucky

By
Sayed Mahdi Saghaian Nejad Esfahani
Lexington, Kentucky

Director: Dr. Kevin D. Donohue, Professor of Electrical and Computer Engineering
Lexington, Kentucky 2010

To my loving Parents, Wife, Sister and Brothers

ACKNOWLEDGMENTS

I am deeply thankful to my supervisor, Dr. Kevin Donohue, whose encouragement, guidance and support from the initial to the concluding level of the research, enabled me to develop an understanding of the subject. His knowledge and character provided a flourishing environment.

I'd like to cordially thank my parents, my wife and my siblings for their love, inspiration, motivation and support during all aspects of my life, including my college career.

I would like to thank the committee member Dr. Jens Hannemann, Dr. Sen-ching Cheung and Dr. Laurence G. Hassebrook for their valuable time and suggestions for improving this thesis.

Finally, I offer my regards to all of those who supported me in any respect during the completion of the project. In particular, I would like to show my gratitude to my colleges at Center for Visualization and Virtual Environments.

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

Chapter 1 Sound Source Localization Algorithms 1

1.1 Introduction 1
1.2 TDOA-based sound source localization approaches 4
1.3 SRP-based sound source localization strategies 8

Chapter 2 Statistics of Noise-Only SRCP Image Values 13

2.1 Introduction 13
2.2 Noise-only SRCP image values statistics 15

Chapter 3 Constant False Alarm Rate (CFAR) Processor 23

3.1 Introduction 23
3.2 Cell Averaging (CA) CFAR processor 26
3.3 Greatest Of (GO) CFAR processor 27
3.4 Smallest Of (SO) CFAR processor 28
3.5 Order Statistics (OS) CFAR processor 29
3.6 CFAR processor in SSL applications 30
3.6.1 Chi-squared distribution 31
3.6.2 Weibull distribution 34

Chapter 4 Experimental Results 40

4.1 Introduction 40
4.2 Mic-distribution factor 43
4.3 CFAR performance, considering only mic-distribution factor 45
4.3.1 Chi-square distribution 46
4.3.2 Developing scaling methods for the Chi-square distribution application 52
4.3.3 Weibull distribution 67
4.4 Low frequency limit 78
4.5 Adaptive high-pass filters 79
4.6 Combining mic-distribution and noise-path factors 82

Chapter 5 Conclusion 99

5.1 Conclusion 99

Bibliography 101

Vita 104

LIST OF TABLES

Table 4.1: Low frequency limit resulted by noise-path factor..... 85

LIST OF FIGURES

Figure 1.1 A filter-and-sum beamformer structure. 8
Figure 3.1 CA-CFAR processor 25
Figure 3.2 Degradation in CFAR performance caused by error in estimating noise power for different number of samples used in estimation of noise power. σ is set to 1. 34
Figure 4.1 Microphone distributions and FOV (shaded plane) for simulation and experimental recordings with axes in meters. Small filled circles outside the FOV denote a microphone position and the square and star markers in the FOV denote the smallest and largest (respectively) microphone inter-distance standard deviation overall pairs (a) linear (b) perimeter and (c) planar..... 41
Figure 4.2 Normalized histograms for microphone pair differential path lengths at FOV points that generate the minimum and maximum standard deviations for (a) linear geometry (b) perimeter geometry, and (c) planar geometry. 44
Figure 4.3 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 300 Hz. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution (a) variations of PHAT-β parameters using degree of freedom of 1 for Chi-square distribution (b) variations in Chi-square distribution degree of freedom using beta equal to 0.75. ... 47
Figure 4.4 Ratios of specified to empirical (experimental) FA probabilities for linear array. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution. The beta value is set to 0.75 and different Chi-square distribution degree of freedoms are applied using high-pass filter cut-off frequency of (a) 800 Hz (b) 1500 Hz. 49
Figure 4.5 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut-off frequency of 300 Hz. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution (a) variations of PHAT-β parameters using degree of freedom of 1 for Chi-square distribution (b) variations of PHAT-β parameters using degree of freedom of 6 for Chi-square distribution (c) variations in Chi-square distribution degree of freedom using beta equal to 0.75. 50
Figure 4.6 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut-off frequency of 300 Hz. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution (a) variations of PHAT-β parameters using degree of freedom of 1 for Chi-square distribution (b) variations of PHAT-β parameters using degree of freedom of 6 for Chi-square distribution (c) variations in Chi-square distribution degree of freedom using beta equal to 0.75. 51
Figure 4.7: Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85. The performances of modeling three linear combinations of

the coherent power by Chi-square distribution are compared together (a) for degree of freedom of 1 (b) for degree of freedom of 6. 55
Figure 4.8: Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 800 Hz. The PHAT-β parameter is set to 0.85. The performances of modeling three linear combinations of the coherent power by Chi-square distribution are compared together (a) for degree of freedom of 1 (b) for degree of freedom of 2. 56
Figure 4.9: Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 1500 Hz. The PHAT-β parameter is set to 0.85. The performance of modeling three linear combinations of the coherent power by Chi-square distribution are compared together for degree of freedom of(a) 4 (b) 48 (c) 240 (d) 1600. 58
Figure 4.10 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85. The performance of modeling three linear combinations of the coherent power by Chi-square distribution are compared together for degree of freedom of(a) 4 (b) 48 (c) 240 (d) 1600. 60
Figure 4.11 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85. The performance of modeling three linear combinations of the coherent power by Chi-square distribution are compared together for degree of freedom of(a) 3 (b) 4 (c) 10 (d) 32. 62
Figure 4.12 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 1500 Hz. The beta is equal to 0.85. Variations of the neighborhood size using (a) *mean-cx2* approach and degree of freedom of 4 for Chi-square distribution (b) *mean-var-cx2* method and degree of freedom of 240..... 64
Figure 4.13 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut-off frequency of 300 Hz. The beta is equal to 0.85. The *mean-cx2* approach is applied (a) variations of the neighborhood size using Chi-square distribution degree of freedom of 4 (b) for neighborhood size of 7x7 pixels the degree of freedom should be set to 2 while for neighborhood size of 21x21 pixel it can be either 4 or 5 to achieve a reasonable CFAR performance. 65
Figure 4.14 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut-off frequency of 300 Hz. The beta is equal to 0.85. The *mean-cx2* approach is applied (a) variations of the neighborhood size using Chi-square distribution degree of freedom of 3 (b) for neighborhood size of 7x7 pixels a reasonable CFAR performance is not achievable while for neighborhood size of 21x21 pixel the degree of freedom can be either 3 or 4 to achieve a reasonable CFAR performance. 66

Figure 4.15 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. (a) Variations of PHAT-β parameters using shape parameter of 1.26 (b) variations of shape parameters using beta equal to 0.85. 69
Figure 4.16 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut off frequency of 1500 Hz when the coherent power is modeled by Weibull distribution. (a) Variations of PHAT-β parameters using shape parameter of 1.26 (b) variations in shape parameters using beta equal to 0.85. 70
Figure 4.17 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. (a) Variations in PHAT-β parameters using shape parameter of 1.26 (b) variations in shape parameters using beta equal to 0.85. 72
Figure 4.18 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. (a) Variations in PHAT-β parameters using shape parameter of 1.26 (b) variations in PHAT-β parameters, using shape parameter of 1.12. 73
Figure 4.19 CFAR Performance for 3 different neighborhood sizes using Weibull distribution, using partial whitening value of 0.85 for (a) linear array with 1500 Hz cut of frequency and using shape parameter of 1.26 (b) perimeter array with 300 Hz cut of frequency and using shape parameter of 1.26 (c) planar array with 300 Hz cut-off frequency and using shape parameter of 1.12. 75
Figure 4.20 CFAR Performance using Weibull distribution for neighborhood size of 7x7, using partial whitening value of 0.85 for 3 different shape parameters. (a) linear array with 1500 Hz cut-off frequency (b) perimeter array with 300 Hz cut-off frequency (c) planar array with 300 Hz cut-off frequency. 77
Figure 4.21 CFAR performances using adaptive high-pass filters for each FOV point for linear array with beta value equal to 0.85 (a) using *mean-cx2* method and variations in degree of freedom (b) using Weibull distribution and variations in shape parameter. 80
Figure 4.22 CFAR performance using adaptive high-pass filters for each FOV point for beta value equal to 0.85. (a) and (b) for perimeter array. (c) and (d) for planar array. (a) and (c) using *mean-cx2* method and variations in degree of freedom (b) and (d)using Weibull distribution and variations in shape parameter. 81
Figure 4.23 SRCP images. (a) and (c) locating noise source by finding mean value of position of all potential noise source positions. (b) and (d)locating noise source by

finding the position which have maximum value of SRCP among all potential pixels. (a) and (b) for broad side noise. (c) and (d) for endfire noise. 84
Figure 4.24 CFAR performance using 3rd null for both mic-distribution and noise-path factors to design adaptive high-pass filters for each FOV point for linear array with beta value equal to 0.85 (a) using *mean-cx2* method with degree of freedom equal to 2 (b) using Weibull distribution and variations in shape parameter. 86
Figure 4.25 Mic-distribution factor is dominant for no whitening case. CFAR performance for linear array: (a) and (b) using mean-cx2 approach with degree of freedom equal to 2. (c) and (d) exploiting Weibull distribution with shape parameter of 0.95. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation. 88
Figure 4.26 Noise-path factor is dominant for partial whitening case. CFAR performance for linear array: (a) and (b) using mean-cx2 approach with degree of freedom equal to 2. (c) and (d) exploiting Weibull distribution with shape parameter of 0.95. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation. 89
Figure 4.27 CFAR performance using 1st null for both mic-distribution and noise-path factors to design adaptive high-pass filters for each FOV point for perimeter array with beta value equal to 0.85 (a) using *mean-cx2* method with degree of freedom equal to 2 (b) using Weibull distribution and variations in shape parameter. 90
Figure 4.28 CFAR performance using the 1st null for mic-distribution factor and the 2nd null for noise-path factor to design adaptive high-pass filters for each FOV point for perimeter array for no-whitening case (a) using *mean-cx2* method variations in degree of freedom. (b) using Weibull distribution and variations in shape parameter. 91
Figure 4.29 CFAR performance for perimeter array when the 3rd null of the related sinc functions are selected for both low frequency limits resulted by mic-distribution and noise-path factors: (a) and (b) using mean-cx2 approach, variation in degree of freedom. (c) and (d) exploiting Weibull distribution, variation in shape parameter. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation. 93
Figure 4.30 CFAR performance by using the 1st null for both mic-distribution and noise-path factors to design adaptive high-pass filters for each FOV point for planar array with beta value equal to 0.85 (a) using *mean-cx2* method using degree of freedom of 2. (b) using Weibull distribution and variations in shape parameter. 94
Figure 4.31 CFAR performance using the 1st null for mic-distribution factor and the 2nd null for noise-path factor (139 Hz) to design adaptive high-pass filters for each FOV point for planar array for no-whitening case (a) using *mean-cx2* method variations in degree of freedom. (b) using Weibull distribution and variations in shape parameter. 95

Figure 4.32 CFAR performance for planar array when the 2nd null of the related sinc functions are selected for both low frequency limits resulted by mic-distribution and noise-path factors: (a) and (b) using mean-cx2 approach, variation in degree of freedom. (c) and (d) using Weibull distribution, variation in shape parameter. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation. 98

Chapter 1

Sound Source Localization Algorithms

1.1 Introduction

Many applications require or can be enhanced by automatic sound source detection and location, in particular in applications that use microphone arrays such as teleconferencing [1-5], speech recognition [6-12], talker tracking [13], and beamforming for SNR enhancement [14].

Two main approaches for solving the source localization problem are those approaches that use *time-difference of arrival* (TDOA) information, and those based on maximizing the *steered response power* (SRP) of a beamformer [15].

In TDOA based approaches, the goal is to derive the location of the sound source by employing an estimate of TDOA. The most common method to estimate TDOA is exploiting the generalized cross-correlation (GCC) function [16]. The GCC function is the cross-correlation of two filtered versions of received signals by two microphone pairs. The GCC function has a peak at the time corresponding to the TDOA. Once an estimate of TDOA is derived, one can find the source location by various techniques [17, 18]. This method suffers from reverberation, especially in those applications that need a short data segment such as adaptive beamforming and tracking of multiple talkers. There exists a couple of weighting function such as Maximum Likelihood (ML) and Phase Transform (PHAT) weighting function. ML weightings are optimal under free-reverberation conditions but reverberation significantly degrades their performance.

While PHAT weightings are suboptimal in the absence of reverberation, they are more robust against reverberation than ML weightings.

In SRP based methods the goal is to find which point in the space will maximize the SRP of a beamformer. In conventional beamformers, also known as delay-and-sum beamformers, to compensate for the propagation delays in the received signals by microphones, time shifts will be applied to the array signals. Once these signals are time aligned, they are summed together to generate the steered response of the beamformer. In the favorable conditions, the peak of SRP will correspond to the location of the sound source. Filters can be applied to the time aligned signals before adding them together to have a better performance. Recent work shows that the SRP algorithm in conjunction with PHAT (the DFT of the filters are equal to one over the amplitude of DFT of the received signal by each microphone), SRP-PHAT, has one of the most robust performances. Since the SRP can be derived by summing GCC functions of all possible microphone pairs and the autocorrelation terms are independent from the location of the sound source, the autocorrelation terms can be subtracted out which leads to a coherent power. This method is called steered response coherent power (SRCP).

The estimation of location of active voices needs to not only be accurate enough but also should be updated at a high rate with minimal latency. Commercialization of inexpensive and high-speed DSPs made sound source localization feasible by means of microphone arrays in conjunction with adaptive array processing algorithms.

The primary goal of sound source localization is to derive automatically an accurate enough estimation of the location of the sound sources. The first question that comes up in mind is how much the estimation is accurate. One way to describe the term “accurate enough” is using false alarm probability. The idea of how to select automatically threshold for specific desired probability of false alarm is a new idea [19]. They have presented a method for automatically designing threshold using local noise statistics; however, the analysis of the noise field was limited as well as the performance. It only considered experimental studies using a Weibull distribution for

threshold design. This thesis proposes and compares other statistical models for the design of constant false alarm rate thresholds as well using experimental studies.

The key elements about this thesis are that it describes the statistics of noise field using experimental studies, and based on this analysis proposes a statistical model for an adaptive threshold design.

Experiments show that low frequency components are responsible for degradation in threshold estimation. High-pass filters, either a fixed high-pass filter for all FOV points or an adaptive high-pass filter for each FOV point, are exploited to provide a symmetric condition for noise distribution which is a required feature for the novel *Constant False Alarm Rate* (CFAR) processor introduced in this thesis. Based on the experimental result, a relationship for low frequency limit will be derived for all microphone geometries.

As a matter of fact, the accuracy of estimation of sound source position as well as CFAR threshold performance is dependent on the array's geometry. In this thesis, the three microphone distribution used to investigate the performance of CFAR threshold are linear, perimeter and planar distributions.

Chapter 1 provides a detailed analysis on the main sound source localization algorithms. Statistics of noise-only distribution are derived in chapter 2. In chapter 3, different CFAR processors commonly used in radar application as well as the novel CFAR processor for sound source localization issues are introduced. Finally, the experimental results are presented in chapter 4.

1.2 TDOA-based sound source localization approaches

In TDOA-based approaches, the sound sources will be located using a 2-step procedure. In the first step, TDOA's between all possible microphone pairs are estimated. In the second step, these TDOA's along with prior knowledge of microphone positions are exploited to generate hyperbolic curves. These hyperbolic curves intersect and the collection of intersection points is used to estimate the location of a sound source.

As it can be figured out from the name of this approach, it is vital to estimate TDOA's with high accuracy in order to have an accurate estimation of the location of the sound source.

One of the robust methods to compute TDOA'S between microphone pairs is generalized cross-correlation (GCC) function. The GCC function is defined as the cross correlation of two filtered versions of two microphone signals.

Let us denote the sound source signal located at $\boldsymbol{r_i}$ in the space by $u_i(t)$. Then the received signal by the p^{th} microphone located at position $\boldsymbol{r_p}$ in the space will be:

$$v_p(t) = u_i(t) * h_{ip}(t) + \sum_{k=1}^{K} n_k(t) * h_{kp}(t) \qquad (1.1)$$

where K denotes the total number of noise sources and $n_k(t)$ is the k^{th} noise source that could be even non-target speakers, as well as the ambient room noises. Also, $h_{ip}(t)$ is the effective impulse response of the microphone and propagation path from $\boldsymbol{r_i}$ to $\boldsymbol{r_p}$. Also, '*' operator represents the convolution operator.

One of the sources of signal degradation in acoustic conditions such as small rooms is reverberation. Therefore, in addition to the direct path, reflected paths should be considered in the impulse response:

$$h_{ip}(t) = a_{ip,0}\left(t - \tau_{ip,0}\right) + \sum_{n=1}^{\infty} a_{ip,n}(t - \tau_{ip,n}) \qquad (1.2)$$

where $a_{ip,n}(t)$ represents the n^{th} path of the effective impulse response between the sound source located at $\boldsymbol{r_i}$ and p^{th} microphone with corresponding delay $\tau_{ip,n}$ and n represents the n^{th} reflected path. As n increases, the amplitude of the reflected path decreases. Therefore, only a few reflected paths (effective paths) will be considered in estimating the location of the sound source.

The p^{th} microphone signal can be represented in frequency domain by taking Fourier transform of equation (1.1):

$$V_p(\omega) = U_i(\omega)\left[A_{ip,0}(\omega)e^{-j\omega\tau_{ip,0}} + \sum_{n=1}^{E} A_{ip,n}(\omega)\, e^{-j\omega\tau_{ip,n}}\right] + \sum_{k=1}^{K} N_k(\omega)\left[\sum_{n=0}^{F} A_{kp,n}(\omega)\, e^{-j\omega\tau_{kp,n}}\right] \tag{1.3}$$

where E is the number of effective paths (excluding the direct path) of the sound source at $\boldsymbol{r_i}$ and F is the number of effective paths (including the direct path) of the noise sources which contribute to the signal segment used in the estimation.

Similar to the received signal by p^{th} microphone, denote the received signal by q^{th} microphone by $v_q(t)$. The *cross correlation* of two signals $v_p(t)$ and $v_q(t)$, $c_{pq}(\tau)$ is defined as:

$$c_{pq}(\tau) = \int_{-\infty}^{\infty} v_p(t)\, v_q(t+\tau)dt \tag{1.4}$$

And the Fourier transform of cross correlation is called *cross spectrum*, $C_{pq}(\omega)$:

$$C_{pq}(\omega) = \mathcal{F}\{c_{pq}(\tau)\} = \int_{-\infty}^{\infty} c_{pq}(\tau)\, e^{-j\omega\tau}\, d\tau \tag{1.5}$$

And by taking inverse Fourier transform, cross spectrum can be converted to cross correlation:

$$c_{pq}(\tau) = \mathcal{F}^{-1}\{C_{pq}(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} C_{pq}(\omega)e^{j\omega\tau}\,d\omega \quad (1.6)$$

Since microphone *p* and *q* are spatially distributed and $\boldsymbol{r_p}$ and $\boldsymbol{r_q}$ are different, $u_i(t)$ will be received to each microphone with deferent propagation delays. In the case that no reverberation or attenuation exists, $v_p(t)$ and $v_q(t)$ are only two shifted versions of the emerged signal, $u_i(t)$ located at $\boldsymbol{r_i}$. Consequently, the cross correlation of $v_p(t)$ and $v_q(t)$ peaks at the time related to TDOA between received signals by microphone *p* and *q*.

Transforming time domain to frequency domain reduces the required computation for estimating TDOA. Applying the Fourier transform to equation (1.4) and exploiting the convolution property of Fourier transform results in:

$$C_{pq}(\omega) = V_p(\omega)V_q^*(\omega) \quad (1.7)$$

where '*' denotes complex conjugate operator.

As it was defined before, the GCC is the cross correlation of two filtered versions of the two microphone signals. Let $y_p(t)$ and $y_q(t)$ represent filtered versions of signals $v_p(t)$ and $v_q(t)$ respectively and denote Fourier transform of the filtered signals by $Y_p(\omega)$ and $Y_q(\omega)$ respectively.

$$Y_p(\omega) = H_p(\omega)V_p(\omega) \quad and \quad Y_q(\omega) = H_q(\omega)V_q(\omega) \quad (1.8)$$

where $H_p(\omega)$ and $H_q(\omega)$ are Fourier transform of impulse responses of those filters.

Therefore, the GCC of received signals by p^{th} and q^{th} microphones is denoted by $R_{pq}(\tau)$ and can be computed by:

$$R_{pq}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} Y_p(\omega)Y_q^*(\omega)e^{j\omega\tau}\,d\omega \quad (1.9)$$

By applying equation (1.8), equation (1.9) can be rewritten as:

$$R_{pq}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi_{pq}(\omega) V_p(\omega) V_q^*(\omega) e^{j\omega\tau}\, d\omega \tag{1.10}$$

where $\Psi_{pq}(\omega)$ is called *weighting function* and is defined as:

$$\Psi_{pq}(\omega) = H_p(\omega) H_q^*(\omega) \tag{1.11}$$

The TDOA between microphone *p* and *q* can be estimated by finding the maximum value of GCC function. In general, the GCC function has more than one maximum. A couple of weighting functions are exploited to emphasize the local maximum corresponding to the actual value of TDOA and to deemphasize other local maximum values. *ML* (Maximum Likelihood) weighting is one example that is optimal in the no reverberation conditions. However, the performance of *ML* weightings degrades dramatically by increasing the amount of reverberation.

Another weighting function that has received significant attention, not only in TDOA-based methods, but also in SRP-based approaches, is *PHAT* (Phase Transform). Although *PHAT* weightings are suboptimal, they are more robust than *ML* weightings in the presence of reverberation. In *PHAT*, the weighting function is defined as:

$$\Psi_{pq}(\omega) = \frac{1}{|V_p(\omega) V_q^*(\omega)|} \tag{1.12}$$

Later in this chapter more details will be provided about PHAT weighting function.

Today, many sound source localization systems are exploiting TDOA-based techniques to locate the unknown position of sound sources due to the practicality of required computation in this approach. However, this method suffers acutely from reverberation and the performance degrades severely as reverberation increases. As a matter of fact, reverberation, which is a common condition in the small rooms, makes TDOA-based approaches unreliable. Another significant limitation of TDOA-based approaches is their incapability to locate multiple sound sources simultaneously.

1.3 SRP-based sound source localization strategies

The idea in SRP-based approaches is to locate sound sources by creating an array of signals by beamforming or focusing on a set of points in the field-of-view. A single focused signal is generated by time aligning and weighting all the microphone signals and summing the results together, instead of pair wise processing the received signals by microphones and looking for a single intersection point, which is the key technique in the TDOA-based methods.

Steered response is referred to the output of a beamformer in the situation where the beamformer is used to locate the unknown position of sound sources by steering over the field of view (FOV) (a region in the space that most likely contains the sound source). The steered response power (SRP) will show a peak at the position corresponding to the position of the interested sound source. Fig 1.1 illustrates the structure of a *filter-and-sum* beamformer.

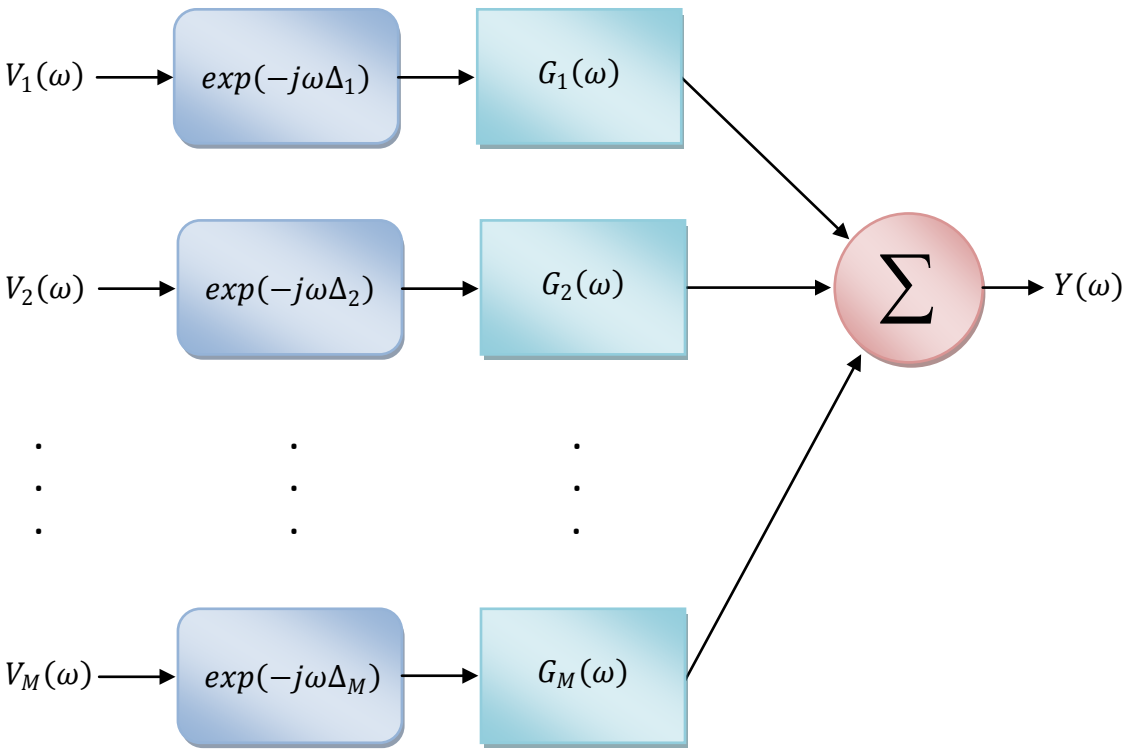

Figure 1.1 A filter-and-sum beamformer structure.

In filter-and-sum beamformers, first, the microphone signals are time aligned to compensate for propagation delays. This time alignment is performed by time-shifting the microphone signals by related time delays denoted by Δ_m. The time delays are selected such that all time-shifts are causal which is essential for real-time systems. One way to make the system be causal is to select time-shifts delays, Δ_m, equal to the largest direct path propagation delay to any of the microphones minus the direct path of propagation delay to the related microphone.

$$\Delta_m = \tau_{max,0} - \tau_{m,0} \tag{1.13}$$

where $\tau_{m.0}$ is the direct path of propagation delay between the sound source located at $\boldsymbol{r_i}$ and m^{th} microphone.

After time aligning the received signals by microphones and before summing them together, some filter processing is performed on each time aligned signal in order to boost the power of the interested sound source while attenuating power of other sources (noise sources as well as non-interested sound sources).

If $y(t)$ represents the steered response, it can be formulated as:

$$y(t) = \sum_{m=1}^{M} g_m(t) * v_m(t - \Delta_m) \tag{1.14}$$

where *M* is the total number of microphones in the array and $g_m(t)$ is the impulse response of the applied filter on the received signal by m^{th} microphone. Also, '*' denotes the convolution operator.

Equation (1.14) can be converted to frequency domain by taking Fourier transform of both sides:

$$Y(\omega) = \sum_{m=1}^{M} G_m(\omega) V_m(\omega) e^{-j\omega \Delta_m} \tag{1.15}$$

where $G_m(\omega)$ is the Fourier transform of the impulse response of the applied filter on the received signal by m^{th} microphone.

SRP can be obtained by exploiting the Parseval theorem:

$$S = \int_{-\infty}^{\infty} Y(\omega)\, Y^*(\omega) d\omega$$
$$= \int_{-\infty}^{\infty} \left(\sum_{p=1}^{M} G_p(\omega) V_p(\omega) e^{-j\omega\, \Delta_p} \right) \left(\sum_{q=1}^{M} G_q^*(\omega) V_q^*(\omega) e^{j\omega\, \Delta_q} \right) d\omega \quad (1.16)$$

where *S* denotes the steered response power.

By applying some simple algebra and exploiting equation (1.13), the SRP can be computed from:

$$S = \sum_{p=1}^{M} \sum_{q=1}^{M} \int_{-\infty}^{\infty} \Psi_{pq}(\omega) V_p(\omega) V_q^*(\omega) e^{j\omega\,(\tau_{p,0} - \tau_{q,0})} d\omega \quad (1.17)$$

where similar to TDOA-based approach, $\Psi_{pq}(\omega)$ is called *weighting function* and is defined as:

$$\Psi_{pq}(\omega) = G_p(\omega) G_q^*(\omega) \quad (1.18)$$

By comparing equations (1.17) and (1.10) it can be concluded that in fact, the SRP is the sum of all possible GCC functions computed at the time corresponding to the TDOA of the pair wised microphones. Therefore, SRP and GCC function are related to each other by:

$$S = 2\pi \sum_{p=1}^{M} \sum_{q=1}^{M} R_{pq}(\tau_{p,0} - \tau_{q,0}) \quad (1.19)$$

Several different weighting functions have been introduced to enhance and improve performance of SRP-based techniques. Recent research showed that SRP-PHAT is robust among all other sound source localization methods. In PHAT, the weighting function, $\Psi_{pq}(\omega)$, is defined as:

$$\Psi_{pq}(\omega) = \frac{1}{|V_p(\omega)||V_q^*(\omega)|} \tag{1.20}$$

By comparing equations (1.18) and (1.20) and applying some simple algebra, it can be demonstrated that the Fourier transform of the impulse response of the applied filter on the received signal by m^{th} microphone is equal to one over the magnitude of the received signal by m^{th} microphone:

$$G_m(\omega) = \frac{1}{|V_m(\omega)|} \quad \text{for } m = 1, \dots, M \tag{1.21}$$

It should be noted that in PHAT weighting, only the Fourier magnitudes of time aligned signals will be affected and the phase will remain unaltered. In fact, PHAT weighting preserves only the spectral phase.

Although previous research showed that PHAT weighting considerably improves the performance of sound source localization by reducing the impact of reverberation effectively, further research demonstrated that superior performance is achieved by exploiting partial PHAT weighting instead of applying PHAT weighting (total weighting) [20,21]. In partial PHAT weighting, also known as PHAT-β, the Fourier transform of the applied filter to the received signal by m^{th} microphone is equal to one over magnitude of the m^{th} microphone signal power to partial weighting parameter, β. The partial weighting parameter is a real number between 0 and 1:

$$G_m(\omega) = \frac{1}{|V_m(\omega)|^{\beta}} \quad \text{for } m = 1, \dots, M \tag{1.22}$$

where β equals to zero denotes no-weighting while β equals to one indicates total weighting. As a matter of fact, it has been shown that better target detection can be achieved with partial PHAT weighting with partial weighting parameter values close to one, for instance 0.85 [20, 21].

One of the advantages of SRP method is that more than one sound source can be detected and located simultaneously. In such a case, the SRP will show multiple peaks corresponding to the positions of those sound sources. However, SRP-based approaches require a large amount of computation to locate the sound sources.

It is worth noting that the SRP is computed from summation of all possible paired-wise filtered versions of time aligned signals receive by the microphone array. On the other hand, the autocorrelation terms in this summation only act as a bias to keep the power a positive value. These product-pairs consisting of the same channel (autocorrelation terms) are independent of spatial position $\boldsymbol{r_i}$. Therefore, the autocorrelation terms can be ignored in the computation of power of steered response. This subtraction out of autocorrelation terms provides a coherent power given by:

$$S_c = \sum_{p=1}^{M} \sum_{q \neq p}^{M} \int_{-\infty}^{\infty} \Psi_{pq}(\omega) V_p(\omega) V_q^*(\omega) e^{j\omega(\tau_{p,0} - \tau_{q,0})} d\omega \qquad (1.23)$$

where S_c represents the steered response coherent power.

This method to locate sound sources is known as *steered-response coherent power* (SRCP).

In the next chapter, after computing statistics of SRCP image values, the advantage of making power coherent will be illuminated. It will be shown that making power coherent gives a researcher the capability to use the ideas currently used in radar application. One of these ideas is *Constant False Alarm Rate* (CFAR) to enhance target detection. In this thesis, a novel CFAR threshold technique will be introduced. This novel CFAR processor results in a novel approach to enhance detection and location of sound sources by microphone arrays.

Chapter 2

Statistics of Noise-Only SRCP Image Values

2.1 Introduction

In chapter 1, sound source localization algorithms were introduced. It was stated that recent works showed that SRCP-PHAT-β is robust among all other different sound source localization approaches in terms of better target detection. Detailed analyses to derive the SRCP image values were also provided.

As it was presented in chapter 1, the SRCP value is computed by:

$$S_c = \sum_{p=1}^{M} \sum_{q \neq p}^{M} \int_{-\infty}^{\infty} \Psi_{pq}(\omega) V_p(\omega) V_q^*(\omega) e^{j\omega(\tau_{p,0} - \tau_{q,0})} d\omega$$

where $\Psi_{pq}(\omega)$ is weighting function which for PHAT-β weighting function is equal to:

$$\Psi_{pq}(\omega) = \frac{1}{|V_p(\omega)|^{\beta} |V_q^*(\omega)|^{\beta}}$$

where β is partial weighting parameter and is a real number between 0 and 1 whereas $V_p(\omega)$ is the received signal by the p^{th} microphone and is given by:

$$V_p(\omega) = U_i(\omega)\left[A_{ip,0}(\omega)e^{-j\omega\tau_{ip,0}} + \sum_{n=1}^{E} A_{ip,n}(\omega)\, e^{-j\omega\tau_{ip,n}}\right] + \sum_{k=1}^{K} N_k(\omega)\left[\sum_{n=0}^{F} A_{kp,n}(\omega)\, e^{-j\omega\tau_{kp,n}}\right]$$

where $U_i(\omega)$ and $N_k(\omega)$ are Fourier transform of sound source located at $\boldsymbol{r_i}$ and k^{th} noise source respectively. $A_{ip,n}(t)$ represents the Fourier transform of the n^{th} path of the effective impulse response between the sound source located at $\boldsymbol{r_i}$ and p^{th} microphone with corresponding delay $\tau_{ip,n}$. *K* is the total number of noise sources. Also, *E* is the total reflected paths (excluding the direct path) of the sound source at $\boldsymbol{r_i}$ and *F* is the total reflected paths (including the direct path) of the noise sources which will be used in the estimation of sound source location.

The main goal of this thesis is to achieve a good CFAR threshold performance for sound source localization applications. In order to achieve a good CFAR threshold performance, it is required to estimate accurate adaptive thresholds sufficiently. On the other hand, it is essential that good CFAR threshold performances are obtained for noise only situations. Therefore, in this thesis all sound sources are set to zero and all the results are for noise-only distribution.

If the sound source located at $\boldsymbol{r_i}$ is set to zero, then $V_p(\omega)$ can be rewritten as:

$$V_p(\omega) = \sum_{k=1}^{K} N_k(\omega)\left[\sum_{n=0}^{F} A_{kp,n}(\omega)\, e^{-j\omega\tau_{kp,n}}\right] \qquad (2.1)$$

In this chapter, statistics of noise-only SRCP image values are derived and it will be demonstrated how microphone geometry in conjunction with the position of noise sources affect the performance of localization of sound sources as well as the performance of CFAR threshold processors. Later in this chapter, it will be demonstrated that the primary sources of degradation in CFAR threshold performance are the low frequency components (relative to inter-path distances between field of view (FOV)

points and microphone pairs as well as inter-path distances between noise sources and microphone pairs). Some signal processing will be exploited to alleviate degradation in CFAR threshold performance due to these low frequency components.

2.2 Noise-only SRCP image values statistics

By taking the expected value from both sides of equation (1.23) over all microphone pairs, the expected value of the SRCP pixels is computed from:

$$E\{S_c\} = \int_{-\infty}^{\infty} \sum_{p=1}^{M} \sum_{q \neq p}^{M} E\left\{\Psi_{pq}(\omega) V_p(\omega) V_q^*(\omega) e^{j\omega(\tau_{p,0}-\tau_{q,0})}\right\} d\omega \qquad (2.2)$$

And since the expected value is taken over all possible microphone pairs, $E\{\Psi_{pq}(\omega)V_p(\omega)V_q^*(\omega)e^{j\omega(\tau_{p,0}-\tau_{q,0})}\}$ is a constant value for all microphone pairs and can be brought out of the summations. Therefore:

$$\begin{aligned} &E\{S_c\} \\ &= (M^2 - M) \int_{-\infty}^{\infty} E\left\{\Psi_{pq}(\omega) V_p(\omega) V_q^*(\omega) e^{j\omega(\tau_{p,0}-\tau_{q,0})}\right\} d\omega \end{aligned} \qquad (2.3)$$

By setting all sound sources to zero as in equation (2.1), the expected value for noise-only distribution case over all microphone pairs in the integrand of equation (2.3) with the assumption that different sources are uncorrelated is computed by:

$$\begin{aligned} &E\{\Psi_{pq}(\omega)V_p(\omega)V_q^*(\omega)e^{j\omega(\tau_{p,0}-\tau_{q,0})}\} \\ &= E\left\{\frac{1}{|V_p(\omega)|^{\beta}|V_q^*(\omega)|^{\beta}}\left(\sum_{k=1}^{K} N_k(\omega) \sum_{n=0}^{F} A_{kp,n}(\omega)\, e^{-j\omega\tau_{kp,n}}\right)\right. \\ &\left.*\left(\sum_{k=1}^{K} N_k^*(\omega) \sum_{t=0}^{F} A_{kq,t}^*(\omega)\, e^{j\omega\tau_{kq,t}}\right) e^{j\omega(\tau_{p,0}-\tau_{q,0})}\right\} \end{aligned} \qquad (2.4)$$

Note that $|V_q(\omega)| = |V_q^*(\omega)|$.

Suppose *X* and *Y* be two uncorrelated random variables whereas *a* and *b* are two scale values. Since expected value is a linear operator, it can be stated that:

$$E\{aX + bY\} = aE\{X\} + bE\{Y\} \tag{2.5}$$

And since random variables *X* and *Y* are uncorrelated, it can be written:

$$E\{XY\} = E\{X\}E\{Y\} \tag{2.6}$$

By exploiting equations (2.5) and (2.6) and applying some algebra, equation (2.4) is rewritten as:

$$\begin{aligned} & E\left\{\Psi_{pq}(\omega)V_p(\omega)V_q^*(\omega)exp\left(j\omega\left(\tau_{p,0}-\tau_{q,0}\right)\right)\right\} \\ & = \frac{1}{|V_p(\omega)|^\beta |V_q(\omega)|^\beta} E\left\{e^{j\omega\left(\tau_{p,0}-\tau_{q,0}\right)}\right\} \\ & * \sum_{k=1}^{K} E\{|N_k(\omega)|^2\} \sum_{n=0}^{F} \sum_{t=0}^{F} \left[A_{kp,n}(\omega)A_{kq,t}^*(\omega)\, E\left\{e^{j\omega\left(\tau_{kq,t}-\tau_{kp,n}\right)}\right\}\right] \end{aligned} \tag{2.7}$$

Clearly, one can convert angular frequency, ω, to frequency, *f*, by $\omega = 2\pi f$. The inter-path distance and propagation delay are related to each other by:

$$\tau_{p,0} = \frac{d_{ip}}{c} \tag{2.8}$$

where *c* is the speed of sound and d_{ip} is the spatial distance between FOV point corresponding to spatial position $\boldsymbol{r}_i$ and the p^{th} microphone.

Finally, in order to investigate the statistics of noise-only SRCP image values in terms of microphone geometry, the frequency, *f*, is converted to wavelength, λ, by:

$$f = \frac{c}{\lambda} \tag{2.9}$$

In consequence, equation (2.7) is expressed in terms of microphone geometry by:

$$
\begin{aligned}
&E\left\{\Psi_{pq}(\omega)V_p(\omega)V_q^*(\omega)\exp\left(j\omega\left(\tau_{p,0}-\tau_{q,0}\right)\right)\right\}\\
&=\frac{1}{|V_p(\omega)|^\beta|V_q(\omega)|^\beta}E\left\{\exp\left(j2\pi\left(\frac{d_{ip}-d_{iq}}{\lambda}\right)\right)\right\}\\
&*\sum_{k=1}^{K}E\{|N_k(\omega)|^2\}\sum_{n=0}^{F}\sum_{t=0}^{F}\left[A_{kp,n}(\omega)A_{kq,t}^*(\omega)\;E\left\{e^{j2\pi\left(\frac{d_{kq,t}-d_{kp,n}}{\lambda}\right)}\right\}\right]
\end{aligned}
\tag{2.10}
$$

Equation (2.10) consists of two exponential terms which are sources of incoherency or decorrelation; In fact, the expected value of noise-only SRCP image values may take a value other than zero because of these exponential functions. Note that in the ideal situations, the exponential arguments span uniformly from $-\pi$ to π over all microphone pairs and result in a zero expected value for noise-only SRCP image values.

The exponential term which is factored out of the summation of equation (2.10), contains inter-path distance between microphone pairs and FOV point as of its argument. This term depends on the microphone geometry and is referred to as the *mic-distribution factor* in this thesis. The other exponential term which is inside the summation is due to inter-path distance between noise sources and microphone pairs. This term depends on the position of noise sources and will be referred to as the *noise-path factor* throughout this thesis.

Under the assumptions that the inter-path distance random variable has either a zero-mean Gaussian distribution or a zero-mean uniform distribution, equation (2.10) can be represented as a closed-form equation. Let us denote the inter-path distance (between a focal point at $\boldsymbol{r_i}$ and array's microphone) random variable by Δ_{pq}. In the case that Δ_{pq} is a zero-mean Gaussian random variable with standard deviation of σ_Δ, moment generating function can be exploited to compute the expected value. By definition, the expected value of the random variable e^{tX} is called moment generating function and is denoted by $M_X(t)$:

$$
M_X(t)=E\{e^{tX}\},\qquad t\in\mathbb{R} \tag{2.11}
$$

If random variable X has Gaussian distribution with mean and variance of μ and σ^2 respectively then the moment generating function is computed from:

$$M_X(t) = E\{e^{tX}\} = \exp\left[\mu t + \frac{1}{2}\sigma^2 t^2\right] \tag{2.12}$$

Therefore, for the random variable Δ_{pq} which has a zero-mean Gaussian distribution with standard deviation of σ_Δ by substituting t with $-j\frac{2\pi}{\lambda}$, the expected value is derived by:

$$E\left\{\exp\left[-j2\pi\left(\frac{\Delta_{pq}}{\lambda}\right)\right]\right\} = \exp\left[-2\left(\pi\frac{\sigma_\Delta}{\lambda}\right)^2\right] \tag{2.13}$$

A reasonable assumption is to assume Δ_{pq} has a zero-mean uniform distribution with standard deviation of σ_Δ which spans from $-\pi$ to π. Applying the definition of expected value results in:

$$E\left\{\exp\left[-j2\pi\left(\frac{\Delta_{pq}}{\lambda}\right)\right]\right\} = \int_{-\pi}^{\pi}\exp\left[-j2\pi\left(\frac{\delta}{\lambda}\right)\right]f_{\Delta_{pq}}(\delta)d\delta \tag{2.14}$$

where $f_{\Delta_{pq}}(\delta)$ is probability density function (pdf) which for uniform distribution is equal to one over 2π. Consequently:

$$E\left\{\exp\left[-j2\pi\left(\frac{\Delta_{pq}}{\lambda}\right)\right]\right\} = \frac{\lambda}{(2\pi)^2 j}\left[\exp\left(j\frac{2\pi^2}{\lambda}\right) - \exp\left(-j\frac{2\pi^2}{\lambda}\right)\right] \tag{2.15}$$

A useful equation that relates subtraction of two complex conjugate exponential functions to a sinusoidal function is:

$$\sin x = \frac{\exp(jx) - \exp(-jx)}{2j} \tag{2.16}$$

Also, variance of a zero-mean uniform random variable which takes values in the interval of $[-\pi, \pi]$ is computed from:

$$\sigma^2{}_{\Delta} = \frac{4\pi^2}{\sqrt{12}} \tag{2.17}$$

By substituting equations (2.16) and (2.17) into equation (2.15), the expected value is computed in terms of wavelength and standard deviation of inter-path distances:

$$E\left\{\exp\left[-j2\pi\left(\frac{\Delta_{pq}}{\lambda}\right)\right]\right\} = \operatorname{sinc}\left(\pi\frac{\sqrt{12}\,\sigma_{\Delta}}{\lambda}\right) \tag{2.18}$$

where the sinc function is defined as:

$$\operatorname{sinc}(x) = \frac{\sin(x)}{x} \tag{2.19}$$

Note that the exponential function will diminish with a much higher rate than the sinc function. Therefore, the worst scenario is to assume that the inter-path distances have uniform distribution.

The key feature of the novel CFAR processor that will be introduced in the next chapter is to assume the noise-only distribution is symmetric, so one can set the statistics (shape parameter, scalar parameter and etc.) of the positive coherent power values equal to the statistics of the negative values. However, equations (2.13) and (2.18) show that the expected value can never be identically zero over a range of frequencies and therefore the symmetric condition is violated. Furthermore, equations (2.13) and (2.18) indicate that if the standard deviation of the inter-path distances relative to source wavelength is increased sufficiently, the zero-mean condition can be a reasonable assumption, despite the fact that the expected value is not identically zero over a range of frequencies. In consequence, the microphone distribution which has the greatest variance of inter-path distances between microphone pairs and FOV should

perform best among all other microphone arrays in terms of better CFAR performance. On the other hand, if the low frequency components of microphone signals are filtered out, the zero-mean condition will be bolstered as well. Therefore, high-pass filters can be exploited to alleviate the inability of specific microphone distribution to decorrelate low frequency components.

It should be noted that in addition to the zero-mean condition, the distribution should not have either positive or negative skewness to be a symmetric distribution. The skewness, γ, is defined as the ratio of the third central moment, μ_3, and the third power of standard deviation:

$$\gamma = \frac{\mu_3}{\sigma^3} \tag{2.20}$$

Due to the complexity in the computation of the third moment, only the zero-mean condition is examined in this thesis while the skewness will be examined directly from the histograms of negative and positive coherent power values.

In this thesis, two approaches are used to filter out the low frequency components; the simplest approach is to design only one high-pass filter for all FOV points. The mic-distribution factor scales all noise components. In addition, it only depends on the microphone geometry as well as FOV. These features provide a convenient point for the purpose of designing a high-pass filter. In this method, only one high-pass filter with the cut-off frequency corresponding to the mic-distribution factor is applied to all FOV points in order to make the noise-only distribution near-symmetric.

A more sophisticated approach to make the noise-only distribution near-symmetric by filtering out the low frequencies is to apply high-pass filters to each point of FOV adaptively. The cut-off frequencies of these adaptive high-pass filters are designed based on the combinations of mic-distribution and noise-path factors for each specific point of FOV.

Since the exponential function will diminish with a much higher rate than the sinc function, the worst scenario is to assume that the inter-path distances have uniform

distribution and therefore, equation (2.18) is exploited to derive a proper relationship for the high-pass filter cut-off frequency (or also, called as the low frequency limit throughout this thesis). Based on experimental results a proper cut-off frequency for the high-pass filter will be designed.

Experimental results, which will be represented later, showed that if only the mic-distribution is considered to determine the low frequency limit, the frequencies larger than the third null of the sinc function of equation (2.18), which are limited by a -20 dB or less from the maximum, will lead to a good CFAR performance for all microphone distributions examined in this thesis. Therefore, based on the third null of the sinc function related to mic-distribution factor, the high-pass filter cut-off frequency is computed from:

$$f_L = \frac{3c}{\sigma_\Delta \sqrt{12}} \tag{2.21}$$

where σ_Δ is the smallest standard deviation of inter-path distance over the FOV . Note that the standard deviation of inter-path distance changes as the point of interest in the FOV changes.

It is worth noting that if the mic-distribution and noise-path factors are combined with each other to determine the low frequency limit, even a smaller value of the sinc function null can be selected as the high-pass filters cut-off frequency. In the situation where the mic-distribution and noise-path factors are combined together, the adaptive high-pass filter cut-off frequencies are set to the maximums of the low frequency limits resulted by mic-distribution and noise-path factors. The low frequency limit resulted by noise-path factor can be computed by equation (2.21) where in this case, σ_Δ is the standard deviation of inter-path distances between the position of noise source and microphone pairs. Furthermore, the low frequency limits resulting from the mic-distribution factor can be computed adaptively, based on the position of each FOV point. In this case, for all microphone geometries, the adaptive high-pass filter cut-off frequencies can be computed from:

$$f_L(x,y) = \frac{3c}{\sqrt{12}\sigma_\Delta(x,y)} \tag{2.22}$$

where $\sigma_\Delta(x,y)$ denotes the standard deviation of inter-path distances between the point inside the FOV plane, located at position of (*x,y*) and the microphone pairs. Later, detailed results will be presented in the experimental results chapter.

CFAR processor is a technique widely used in radar applications. In fact, this technique provides a facilitating tool for target detection. Analogous to radar applications, the primary goal of sound source localization applications is to detect and locate sound sources with a reasonable accuracy. This fact brings up the idea of exploiting CFAR processors in sound source localization applications as well. This chapter as well as the first chapter supplies the required analysis and materials for applying CFAR processor into sound source localization applications as well as enhancing the CFAR performance. The next chapter introduces different CFAR processors. After delineating CFAR processors used in radar applications, a novel CFAR method feasible in sound source localization applications will be introduced.

Chapter 3

Constant False Alarm Rate (CFAR) Processor

3.1 Introduction

The main goal of many applications such as radio detection and ranging (radar) or Sound Source Localization (SSL) is to detect targets or sound sources. It is not practical to perform target detection processing by human beings due to the large amount of information being presented from the data. Therefore, algorithms are developed to reject automatically data with a low likelihood of being related to a target of interest and only present to an observer (or other intelligent process) information with a low likelihood of being noise. This procedure is known as "*automatic*" target detection. The detection process involves comparing the amplitudes or coherent power of received signals with a threshold. A simple assumption is to assume the noise is stationary. In such a case, the received signal can be compared with a fixed threshold over all time/space. If the signal exceeds this threshold, it will be counted as a target in radar applications or sound source in SSL applications. However, in practice, clutter and noise signals are non stationary, and adaptive thresholds are more appropriate for target detection process. Therefore adaptive signal processing should be exploited such that for each local neighborhood a threshold is selected adaptively.

Constant False Alarm Rate (CFAR) processing techniques are widely used to facilitate target detection, especially in the non-stationary environments. The existing CFAR approached are based on sliding window technique. The test cell or pixel is associated with data within a reference window, which is assumed to contain noise samples similar

to that of the test pixel. Based on the statistics of the noise in the reference window, a threshold is calculated. CFAR detectors are common in radar processing whereas it is a novel approach in SSL applications.

In order to limit error in the adaptive threshold due to the leakage of the target's energy to the neighborhood cells, the two cells directly adjacent to the cell under testing will not be used in the estimation of the clutter power. These adjacent cells are called *guard* cells.

In this chapter, different CFAR methods are introduced. In general, there are two main CFAR approaches, Cell Averaged CFAR (CA-CFAR) and Order Statistic CFAR (OS-CFAR). Fig 3.1 shows a schematic for CA-CFAR processors.

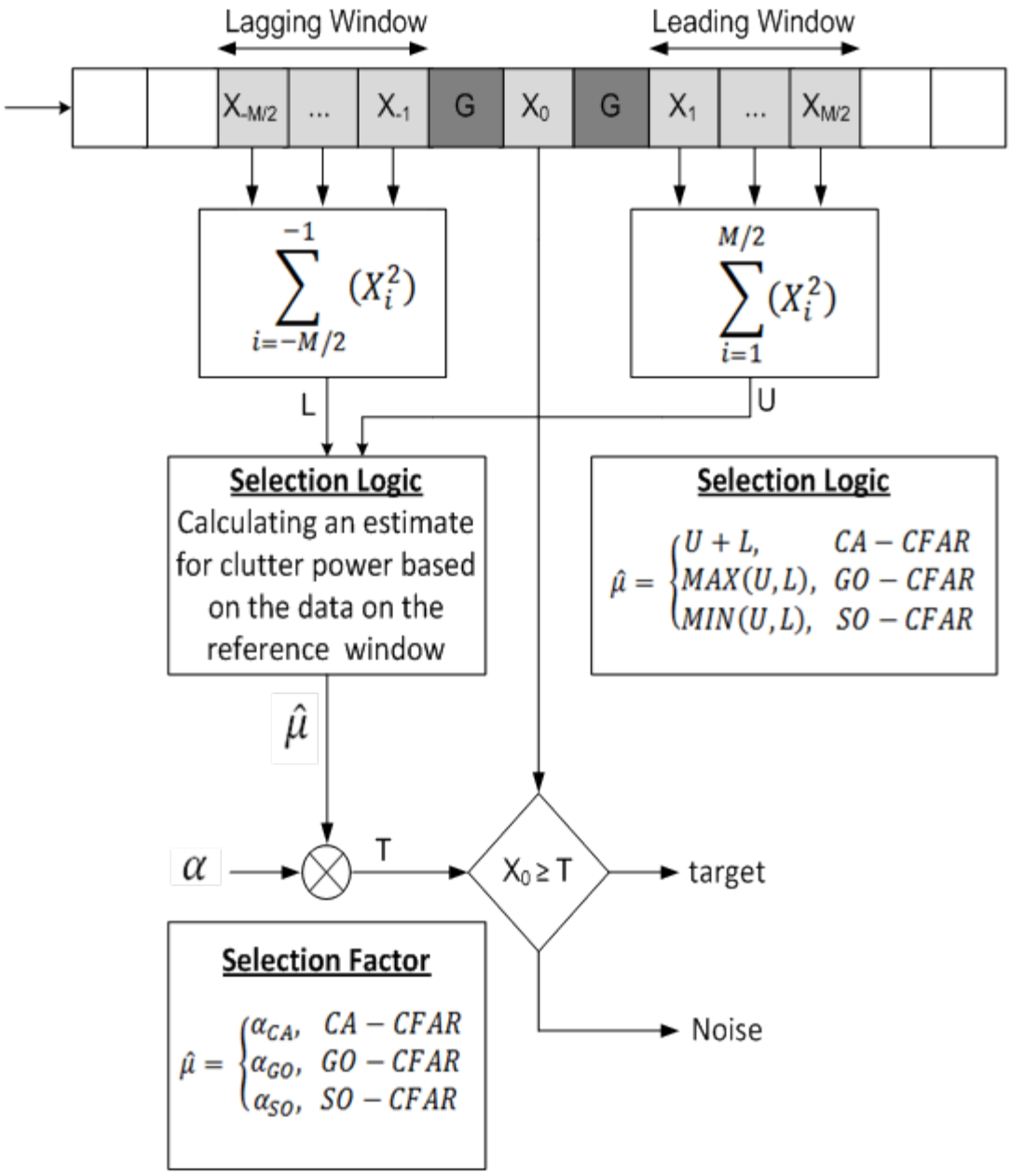

Figure 3.1 CA-CFAR processor

3.2 Cell Averaging (CA) CFAR processor

In the Cell Averaging CFAR (CA-CFAR) detectors [22], the adaptive threshold is computed in two steps. In the first step, which makes different CA-CFAR methods distinguished, the clutter power level is estimated by taking the average of all cells in the reference window, denoted by $\hat{\mu}$. In the second step, the estimated clutter power from CA-CFAR approach, $\hat{\mu}_{cA}$, is multiplied by a scaling factor (α) to obtain the adaptive threshold. The scaling factor is mainly dependent on the CFAR method and also, the required false alarm probability. So, the scaling factor in CA-CFAR method is denoted by α_{CA}. Mathematically speaking, to examine if the test cell is target or not the following procedure needs to be done. Let $\mathbf{H_0}$ denote the condition or hypothesis that no target is present and $\mathbf{H_1}$ denote the hypothesis that the target plus noise is present. The threshold decision is simply:

$$\begin{cases} \text{Decide } \mathbf{H_1}, & \text{if } X_0 \geq T \\ \text{Decide } \mathbf{H_0}, & \text{if } X_0 < T \end{cases} \tag{3.1}$$

where X_0 is the cell under test and *T* is the adaptive threshold.

The threshold is obtained by finding an estimate for average clutter power:

$$\hat{\mu}_{cA} = \frac{1}{M} \sum_{i=-M/2}^{M/2} X_i^2 \tag{3.2}$$

where *M* is the total number of cells in the reference window and X_i are the samples (amplitude) of the received data. The estimated power is then scaled to achieve the threshold for a specific false alarm rate:

$$T = \alpha_{CA} \hat{\mu}_{cA} \tag{3.3}$$

The CA-CFAR detectors works well in situations where the statistics of the cell under test is the same as the statistics of each reference cells. In other words, CA-CFAR processors performs well in situations where only a single target presents in the reference windows and also the clutter is homogeneous in the local neighborhoods.

The basic CA-CFAR detectors suffer from situations where the clutter is not uniformly distributed in the reference cells. A common situation where this phenomenon happens is where the reference window crosses the clutter edge. If the test cell is immersed in the clutter but some of the reference cells are in the clear region, then the threshold is decreased and consequently the probability of false alarm is increased intolerably [22]. On the other hand, existence of another target in the reference window of the target in question (primary target) causes the threshold to be increased and hence, detection probability will be decreased. These problems with conventional CA-CFAR processors call for modified versions of CA-CFAR processors.

3.3 Greatest Of (GO) CFAR processor

The basic technique in modified CA-CFAR detectors is to estimate clutter power independently from the leading and lagging reference cells and use either the estimate from leading or lagging windows as an estimate for power of clutter (see Fig 3.1).

In the Greatest Of CFAR (GO-CFAR) approach, in order to alleviate the excessive false alarm rate caused by clutter edges (non-uniformity of clutter power within the reference window), two estimations for power of clutter are derived from leading and lagging reference cells independently and the greatest of these clutter power estimations is selected to compute the threshold adaptively.

$$\hat{\mu}_{GO} = \frac{2}{M} \max\left(\sum_{i=-M/2}^{-1} (X_i^2), \sum_{i=1}^{M/2} (X_i^2) \right) \tag{3.4}$$

Although exploiting GO-CFAR processors reduces the problems caused by the existence of edge in the clutter, it is obvious that the suppression in target detection introduced by the presence of an interfering target is more severe in the GO-CFAR detectors rather than in the conventional CA-CFAR processors.

3.4 Smallest Of (SO) CFAR processor

Another situation where the uniformity of clutter in the reference cells is violated is where an interfering target lies within the reference cells of the primary target. As a matter of fact, the presence of another target in the reference cells of the primary target will cause the threshold to be increased intolerably. Therefore, detection probability along with false alarm probability will decrease. Clearly, this degradation in target detection is more acute in the GO-CFAR detectors.

One way to prevent suppression in target detection due to the presence of an interfering target is to use Smallest Of CFAR (SO-CFAR) detectors. In SO-CFAR processors, the smallest of the mean of either leading or lagging reference cells is selected to be used in the threshold computation (Fig 3.1).

$$\hat{\mu}_{SO} = \frac{2}{M} \min\left(\sum_{i=-M/2}^{-1} (X_i^2), \sum_{i=1}^{M/2} (X_i^2) \right) \tag{3.5}$$

Although SO-CFAR detectors alleviate the degradation in target detection caused by the presence of an interfering target, they suffer from an excessive number of false alarms due to the decrease of the adaptive threshold.

To prevent the degradation in sensitivity in SO-CFAR detectors caused by an excessive number of false alarms, the number of cells in the reference window should be sufficiently large. However, increasing the length of the reference window endangers the assumption that the noise is stationary in the local neighborhoods.

3.5 Order Statistics (OS) CFAR processor

The suppression in target detection in the CA-CFAR processors and their modifications caused by the presence of interfering target, calls for another approach that can resist the abovementioned suppression in target detection.

Order Statistics CFAR (OS-CFAR) detectors [23] are introduced to overcome the problem caused by an interfering target. The idea is to use order statistics of the reference cells to compute the adaptive threshold. The procedure of selecting threshold adaptively by exploiting OS-CFAR processors is as follows. First of all, the M samples are ranked in an increasing order:

$$X_1 \leq X_2 \leq \cdots \leq X_M \tag{3.6}$$

Then two ranked samples, X_i and X_j are selected from the reference window such that:

$$X_1 \leq X_i \leq X_j \leq X_M \tag{3.7}$$

The adaptive threshold, denoted by *T*, depends on the CFAR processor method and the specific false alarm rate. For OS-CFAR detectors the adaptive threshold, T_{OS}, can be derived by

$$T_{OS} = X_i^{1-\beta} X_j^{\beta} \tag{3.8}$$

where

$$\beta = \frac{\ln \alpha_i}{\ln[-\ln(1-p_j)] - \ln[-\ln(1-p_i)]} \tag{3.9}$$

where

$$p_k = \frac{k}{M+1} \tag{3.10}$$

and α_i is related to the specific probability of false alarm by

$$P_{FA} = \frac{M!\,(\alpha_i + M - i)!}{(M-i)!\,(\alpha_i + M)!} \tag{3.11}$$

It has been demonstrated [24, 25] that the optimum choice of *i* and *j* are given by

$$p_i = 0.1673 \quad \text{and} \quad p_j = 0.9737 \tag{3.12}$$

3.6 CFAR processor in SSL applications

The idea of exploiting CFAR processors in Sound Source Localization (SSL) is a novel idea [19]. In this approach, the SRP-PHAT algorithm is used to locate a sound source. Subtracting out the auto correlation terms from the power of the beamformer in the filter-and-sum process, results in creating a coherent power value. The negative pixels in the SRCP image are results from only noise whereas the positive pixels are a result of either an existing noise source or a sound source. Therefore, each positive pixel which has the maximum value in the neighborhood surrounding it, has the potential to be a sound source. The size of the neighborhood should be selected such that it can be assumed that the pixels in the neighborhood have the same statistics. In other words, it can be assumed that noise is stationary in the local neighborhoods. A significant assumption in this novel CFAR processor is to assume a symmetric distribution for noise-only regions. This assumption requires a zero mean distribution for noise-only distribution which can be appeased by applying the high-pass filters introduced in chapter 2. In fact, only the statistics of negative coherent power values (pure noise region) are exploited to model the noise-only distribution and it will be assumed that positive coherent power values have the same statistics as have the negative coherent

power values. In other words, the negative coherent power values mirror the positive coherent power values.

In this thesis, two different distributions are examined to model the coherent power distribution for $\mathbf{H_0}$ (the no-target hypothesis). The first distribution is the *Chi-square distribution*, which is the theoretical model for the sum of squared Gaussian values. *Weibull distribution* is the other distribution, which is considered for its ability to model potential skewness in the coherent power distribution.

3.6.1 Chi-squared distribution

Under the conditions in which the original noise is locally stationary and has Gaussian distribution, it will be a good assumption to model the coherent power by the Chi-square distribution. Therefore, one of the distributions which can be exploited to model the coherent power is the Chi-square distribution. Chi-squared distribution can be derived by Normal (also known as Gaussian) distribution. If X_i are k independent and identically distributed (*iid*) random variables which have Normal distributions with mean μ_i and standard deviation σ_i, then Q will have Chi-square distribution with k degrees of freedom where

$$X_i \sim Normal\,(\mu_i, \sigma_i) \quad \text{for } i = 1, \dots, k \tag{3.13}$$

$$Q = \sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \tag{3.14}$$

or

$$Q \sim \chi_k^2 \tag{3.15}$$

To investigate the CFAR degradation resulting from an error in estimating power of noise, a simulation was run using the *Monte Carlo* simulation technique. The adaptive

thresholds, *T*, are computed for 6 specific false alarm probabilities ranged from 10^{-1} to 10^{-6}. Since the probability of false alarm compliments $(1 - P_{FA})$ is equal to cumulative distribution function (*cdf*) calculated at threshold value, the adaptive thresholds are computed by finding inverse of cdf at probability of false alarm compliments:

$$T = cdf^{-1}(1 - P_{FA}) \tag{3.16}$$

In this simulation, *N* independent Normal random variables with mean 0 and standard deviation σ are created.

$$X_i \sim Normal\ (0, \sigma^2) \quad \text{for } i = 1, \dots, N \tag{3.17}$$

An unbiased estimation for variance of *N samples,* x_i for $i = 1, \dots, N$, can be computed by

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{3.18}$$

where $\bar{x}$ is the sample mean.

Since these *N* Normal random variables have a mean of 0, the variance can be estimated by squaring them and taking the average of the squared values:

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i)^2 \tag{3.19}$$

In the next step, another zero mean Normal random variable, denoted by *D*, is created. This Normal random variable has a variance equal to the estimated variance from the former *N* random variables. Therefore,

$$D \sim Normal\ (0, \hat{\sigma}^2) \tag{3.20}$$

Applying equation 3.14 to the Normal random variable *D* leads to the creation of a Chi-squared random variable with degree of freedom 1:

$$\frac{D^2}{\hat{\sigma}^2} \sim \chi_1^2 \tag{3.21}$$

Finally, this Chi-squared random variable is compared with the threshold and if it exceeds the threshold, then it will be counted as a target.

This simulation is run $3\text{x}10^7$ times. On can calculate the experimental probability of false alarm by dividing the number of times the created Chi-square random variable exceeds the threshold, by the total number of runs ($3\text{x}10^7$). Fig 3.2 illustrates the ratio of experimental to desired FA probability versus desired probability of FA for six specific desired FA probabilities, ranged from 10^{-1} and 10^{-6}. In addition to CFAR performance for the specific desired false alarm probabilities, the effect of different number of samples which are used for estimating the power of noise on CFAR performance is investigated in this figure as well. The broken line represents ratios of one and implies perfect agreement between experimental and desired FA probabilities. A ratio less than one implies the experimental threshold was too low whereas ratios greater than one mean the experimental thresholds were too high.

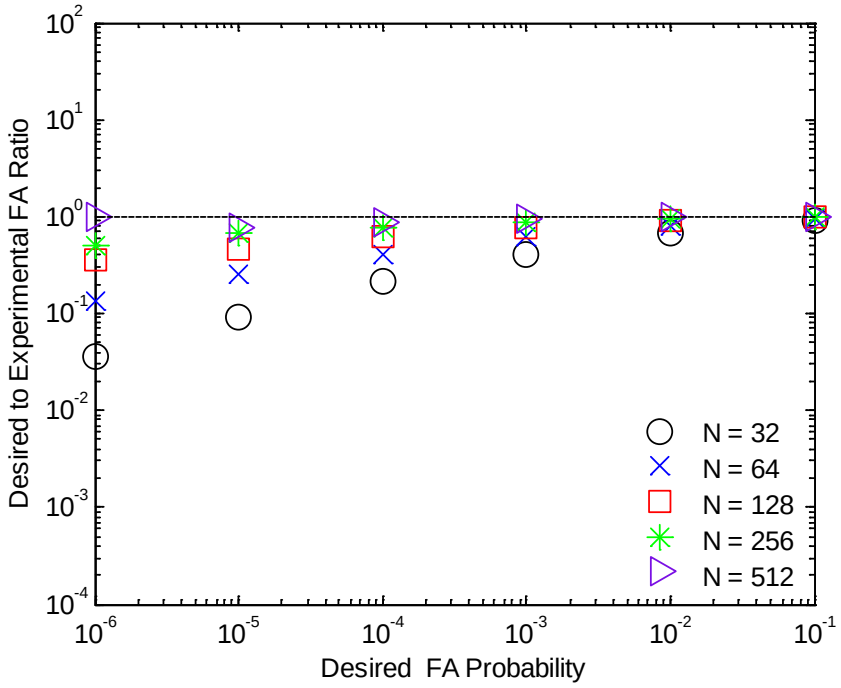

Figure 3.2 Degradation in CFAR performance caused by error in estimating noise power for different number of samples used in estimation of noise power. σ is set to 1.

As Fig 3.2 shows, CFAR performance can be degraded by an error in estimating the power of noise and the degradation in CFAR performance is increased as desired false alarm rate is decreased. The error in estimation of noise power is caused by insufficient number of samples exploited in the estimation. Exploiting higher number of samples which noise power is estimated from improves the estimation of noise power. Consequently, improvement in estimation of noise power will lead to a good CFAR performance.

3.6.2 Weibull distribution

The tail of the distribution has an important role in automatically estimating the probability of false alarm. In particular, good CFAR performance for low false alarm

probabilities requires very accurate modeling of the noise at the tail of the distribution. Since the skewness affecting the tail of the Weibull distribution can be parametrically adjusted, the Weibull is chosen to model the coherent power distribution. Also, it should be noted that the Rayleigh and Exponential distributions, two powerful and widely used distributions are two special cases of the Weibull distribution (if *b* is equal to 1 the Weibull distribution reduces to Exponential distribution, and for *b* equal to 2, Rayleigh distribution is produced). The tail can be adjusted through the shape parameter and a closed form expression exists to compute the threshold from the desired false alarm rate, shape parameter, and estimate of the power or scale parameter.

Assume *X* has a Weibull distribution with scale parameter of *a* and shape parameter of *b*:

$$X \sim Wei\ (a, b) \tag{3.22}$$

Since the cumulative distribution function (*cdf*) of Weibull distribution is continuous function, Weibull distribution is categorized in the continuous probability distribution category. Therefore, if *X* is a Weibull random variable then,

$$\Pr[X = x] = 0 \qquad \text{for } \forall x \in \mathcal{R} \tag{3.23}$$

The probability density function (*pdf*) for a Weibull random variable *X* is derived by

$$f_X\ (x; a, b)\ = \begin{cases} \left(\frac{b}{a}\right)\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3.24}$$

where $a > 0$ is the scale parameter and $b > 0$ is the shape parameter of the distribution. Also, the cdf of Weibull random variable *X* is computed by

$$F_X\ (x; a, b)\ = \begin{cases} 1 - \ e^{-\left(\frac{x}{a}\right)^b} & x \geq 0 \\ \\ 0 & x < 0 \end{cases} \tag{3.25}$$

The probability of false alarm is equal to 1 minus *cdf* computed at the threshold value. Therefore, when the distribution is Weibull it can be expressed as:

$$P_{FA} = \exp\left(-\left(\frac{T}{a}\right)^b\right) \tag{3.26}$$

In addition to have a specific false alarm rate, it is required to have the scale and shape parameters of Weibull distribution in order to find the threshold. Once the parameters of Weibull distribution are derived, the threshold can be easily computed for specific probability of false alarms by finding inverse value of equation (3.26).

The parameters of Weibull distribution can be estimated from samples data (for this thesis, coherent power created by SRCP algorithm) by any parametric estimation techniques such as *Maximum Likelihood Estimation* (*MLE*) or Minimum Mean Squared Error (*MMSE*). However, it is more practical to fix the shape parameter value and estimate scale parameter by *Maximum Likelihood Estimation* (*MLE*) approach.

Assume random variable *X* has Weibull distribution and *N* independent samples of this random variable are available, $X_1, \dots, X_N$. So, these samples are independent and identically distributed (*iid)*. Let us fix the shape parameter of the Weibull distribution. Hence, *pdfs* of these *N* samples are not functions of shape parameter and only depend on possible measured values (x_i) for i^{th} random variable (X_i) and scale parameter (a), $f_{X_i}(x_i; a)$. Since these measurements are independent, the joint probability distribution function, $f_X(x, a)$, is computed from multiplication of the *pdf*s of the *N* sample random variables.

$$f_X(x, a) = \ f_X(x_1, \dots, x_N; a) = \prod_{i=1}^{N} f_{X_i}\ (x_i; a) \tag{3.27}$$

If the values of random variables $X_1, \dots, X_N$ are considered to be fixed values and a is an unknown parameter, then $f_X(x_1, \dots, x_N)$ is called *likelihood function*. The likelihood function is denoted by

$$L(a) = f_X(x_1, \dots, x_N; a) = \prod_{i=1}^{N} f_{X_i}(x_i; a) \tag{3.28}$$

By definition, the *maximum likelihood estimator* $(\hat{a})$ of a is the value that maximizes the likelihood function.

$$L(\hat{a}) \geq L(a), \qquad \text{for } \forall a \tag{3.29}$$

Many of the probability density functions are in the form of exponential terms. Therefore, instead of using likelihood function, natural logarithm of the likelihood function is used in order to simplify required computation of maximum likelihood estimation. The natural logarithm of the likelihood function is called *log-likelihood function*. It is obvious that since the natural logarithm is a strictly increasing function, the log-likelihood function and the likelihood function have the same extremum values.

Returning to our problem, we have *N iid* samples which have Weibull distribution $X_1, \dots, X_N$. Also, the shape parameter is fixed and the goal is to estimate scale parameter from *N* samples by *MLE* method. Let us denote the scale parameter by $\hat{a}$. The likelihood function is

$$L(a) = \prod_{i=1}^{N} f_X(x_i; a) = \prod_{i=1}^{N} \left(\frac{b}{a}\right)\left(\frac{x_i}{a}\right)^{b-1} e^{-\left(\frac{x_i}{a}\right)^b} \tag{3.30}$$

And the log-likelihood function is

$$\ln\big(L\ (a)\big) = \sum_{i=1}^{N}[\ln(b) + (b-1)\ln(x_i)] - Nb\ln(a) - \frac{1}{a^b}\sum_{i=1}^{N}(x_i)^b \tag{3.31}$$

One can find the maximum likelihood estimation of a, the scale parameter, by setting the derivative of equation (3.31) to 0:

$$\hat{a} = \left(\frac{1}{N}\sum_{i=1}^{N}(x_i)^b\right)^{\frac{1}{b}} \tag{3.32}$$

An alternative approach to estimate the scale parameter for a known shape parameter is to exploit the expected value of the Weibull distribution. The expected value of Weibull distribution is computed from:

$$E\{X\} = a\Gamma\left(1 + \frac{1}{b}\right) \tag{3.33}$$

where $\Gamma(.)$ denotes gamma function.

The expected value can be estimated by taking the average of all samples:

$$E\{X\} = \frac{1}{N}\sum_{i=1}^{N}x_i \tag{3.34}$$

Once the expected value is estimated, the scale parameter can be computed from:

$$\hat{a} = \frac{E\{X\}}{\Gamma\left(1 + \frac{1}{b}\right)} \tag{3.35}$$

It should be note that for the sound source localization issue, because the negative pixels are representing pure noise and the goal is to model noise by Weibull distribution, only negative values in the local neighborhoods are used to estimate the scale parameter.

Now that the scale parameter is estimated, the threshold value can be obtained for specific false alarm rates by finding the inverse value of the equation (3.26):

$$T = \hat{a} \ln\left(\frac{1}{P_{FA}}\right)^{\frac{1}{\hat{b}}} \tag{3.36}$$

Each pixel in the SRCP image has the potential to represent position of a sound source. If the test pixel exceeds the threshold, then it will be counted as a target (false alarm), whereas if it does not exceed the threshold, it will be implied as noise (correct rejection).

The next chapter presents the experimental results. Over 46.4 million pixels are used to estimate empirical probability of false alarms corresponding to 6 desired constant false alarm probabilities ranged from 10^{-1} and 10^{-6}.

Chapter 4

Experimental Results

4.1 Introduction

Chapter 1 provided the required theoretical details and analyzes for computation of SRCP-PHAT, the robust approach among all SLL algorithms. Partial whitening with the PHAT-β is applied in order to achieve a better target detection [20, 21]. Therefore, the parameter β should be selected such that its value is close to 1. Consequently, the beta value is set mainly equal to 0.75 or 0.85 in this thesis. However, other beta values will be examined to investigate the effect of beta value on the performance of the CFAR processor.

The statistics of noise only distribution were presented in chapter 2. It was shown that under the condition that the variance of inter-path distances is sufficiently large relative to source wavelength, the noise distribution is effectively symmetrical, and this symmetrical condition is the key feature on which the novel CFAR processor operates. Therefore, filtering or changes in array geometry that increase inter-path distance variance should improve the performance of the CFAR processor. To illustrate this hypothesis, 3 different microphone geometries are examined. Furthermore, different high-pass filter cut-off frequencies are exploited. Extensive research is performed to determine a relationship for low frequency limit relevant to standard deviation of inter-path distances for each geometry array.

The data is collected in a 6x6.7x2.2 meters typical room with a carpet floor, acoustic ceiling tiles, plasterboard walls, and windows on one side. The natural noise sources included vents, florescent lights, computer and traffic noise through the windows.

The three microphone arrays used in this thesis were linear, perimeter and planar arrays. Fig 4.1 shows the three geometries. The filled circles represent the microphone positions while the square and the star markers denote the positions inside the FOV where the standard deviation of the inter-path distance between microphones and focal point are the smallest and largest one respectively.

Each array contains 16 Behringer ECM 8000 omnidirectional microphones and an aluminum struts cage is used to hold the microphones in place. The FOV is selected to be a 3x3 meters plane inside the cage and 1.57 m above the floor. The schematic of the linear array is shown in Fig 4.1a. The microphones were symmetrically placed along the y-axis relative to the FOV, 1.52

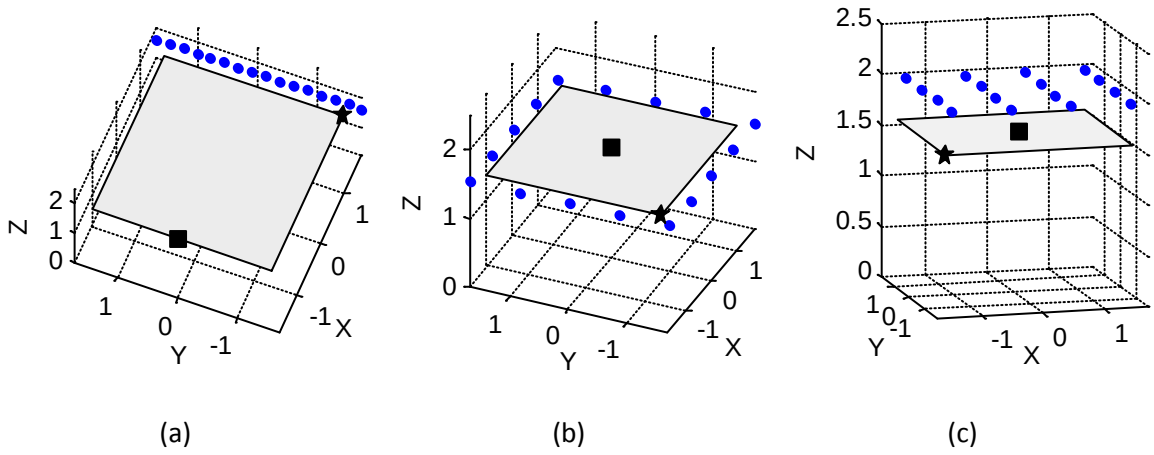

Figure 4.1 Microphone distributions and FOV (shaded plane) for simulation and experimental recordings with axes in meters. Small filled circles outside the FOV denote a microphone position and the square and star markers in the FOV denote the smallest and largest (respectively) microphone inter-distance standard deviation overall pairs (a) linear (b) perimeter and (c) planar.

meters above the floor, and 0.5 m away from the FOV edge. The space between microphones was selected to be a constant value equal to 0.23 m such that the entire array was placed inside the cage. Fig 4.1b presents the perimeter geometry in which

microphones were placed 1.52 meters above the floor, 0.5 m away from the FOV plane, and the microphone spacing was 0.85 m along the perimeter. The planar distribution is exhibited in Fig 4.1c. The microphones were placed in a plane 1.98 m above the floor and in a rectangular grid starting in a corner directly above the FOV, and the microphone spacing was 1 m in the X and Y directions.

The FOV plane was spatially sampled at 4 cm in X and Y directions and the SRCP images were created for the pixels inside the FOV plane, which results in images sized 76 by 76 pixels. Before processing the microphone signals, they are first amplified by M-Audio Buddy preamplifiers. The amplified signals are then digitized through two 8 channels of an M-Audio Delta 1010 Digitizers at 44.1 KHz sampling rate. Finally, the digitized signals are downsampled to 16 KHz for processing. The speed of sound needs to be measured on the day of each recording, which for linear array was 347 m/s and 346 m/s for both perimeter and planar geometries.

Two Yamaha NS-E60 speakers were placed outside the FOV and approximately 1.5 m away from the FOV to create two white noise sources. The noise sources were placed beyond the negative X and negative Y axes relative to the coordinates shown in Fig 4.1. Later in this chapter, two approaches will be applied to locate the accurate position of noise sources and by the knowledge of the noise source positions, the noise-path factor in conjunction with mic-distribution factor will be exploited to determine a low frequency limit which results in a good CFAR performance.

The noise is non-stationary over the FOV; the closer the pixel to the noise source, the higher power the noise has at that pixel. For each microphone geometry, the white noise was played through the speakers and five separate recordings of 25 seconds were captured while the white noise was varied for each separate recording. To create the SRCP image based on equation (1.23), first signals were partitioned into 20 ms segments and incremented every 10 ms because of the non-stationary nature of the data. Then, a specific high-pass filter was applied to the partitioned signals to remove the components that the specific microphone array cannot decorrelate effectively. The high-pass filter can be either a simple high-pass filter for all pixels inside the FOV plane, or,

based on the inter-path distance variance of the specific pixel inside the FOV, adaptive high-pass filters can be exploited. In the former approach, the worst scenario should be considered and the high-pass filter cut-off frequency should be designed based on the minimum standard deviation of the inter-path distances between focal points and microphone pairs. In the next step, partial whiting PHAT-β was performed on the high-pass filtered signals. Once the SRCP images are created, the CFAR processor will be applied. It will be shown later in this chapter that for all microphone geometries, the optimal neighborhood size around the pixel under test is 15x15 pixels, which consequently results in about 46.5 million detection tests for estimating the probability of false alarms.

4.2 Mic-distribution factor

In chapter 2 mic-distribution and noise-path factors were introduced based on equation (2.10). As can be seen from equation (2.10), mic-distribution includes the inter-path distances between all microphone pairs to the focal point. This factor depends only on the microphone geometry and scales all noise components as well. As a result, it provides a convenient point for designing the high-pass filters mentioned earlier.

Histograms can be exploited to illustrate the nature of the microphone differential path length distribution. Since 16 microphones were used, the total number of differential path lengths is equal to the number of permutation for 16 taken 2 at a time, or 240. The histograms of all 240 differential lengths are plotted for two points inside the FOV plane; one point corresponds to the maximum standard deviation of inter-path distances while the other focal point corresponds to the minimum inter-path distances standard deviation. The maximum standard deviation is 1.42, 1.88 and 1.48 for linear, perimeter and planar geometries respectively, whereas the minimum standard deviation is 0.21, 0.38 and 0.67 for linear, perimeter and planar arrays respectively. Fig 4.2 shows the normalized histograms for the three microphone arrays.

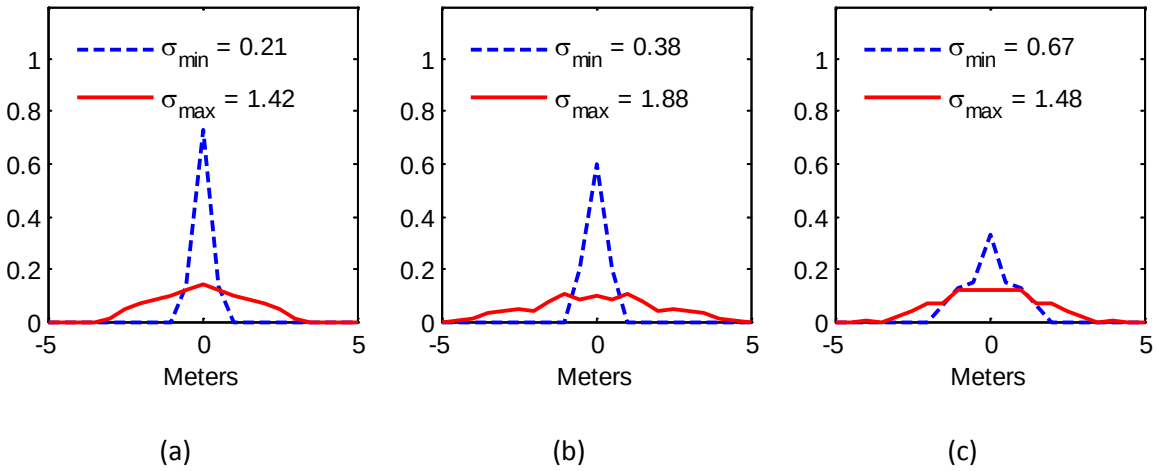

Figure 4.2 Normalized histograms for microphone pair differential path lengths at FOV points that generate the minimum and maximum standard deviations for (a) linear geometry (b) perimeter geometry, and (c) planar geometry.

Fig 4.2 suggests that the distribution of inter-path distances between focal point and microphone pairs is more similar to Gaussian distribution for the least noise distribution symmetry situation (minimum standard deviation). On the other hand, the distribution of inter-path distances becomes more similar to Uniform distribution for the limited support case (maximum standard deviation). By comparing equations (2.13) and (2.18) it can be concluded that the mean offset diminishes faster with increasing the standard deviation, if Gaussian distribution is used to model the distribution of inter-path distances rather than modeling them by Uniform distribution. Therefore, Uniform distribution is selected to model the distribution of inter-path distances as the worst-case limitation. Consequently, based on equation (2.18) as well as experimental results, an empirical relationship will be founded to determine the frequency range for all microphone geometries that lead to a good CFAR performance.

In this thesis, two general methods are applied to reduce the impact of low frequency components, which the microphone geometry cannot decorrelate effectively; one method is high-pass filtering and the other is whitening using PHAT- β.

4.3 CFAR performance, considering only mic-distribution factor

In this section, only the mic-distribution factor is exploited to achieve a good CFAR performance. As it was mentioned in chapter 3, Chi-square and Weibull distributions are used to model the steered-response coherent power distribution in this thesis.

The idea of modeling coherent power by Chi-square distribution came up from the assumption that the noise distribution was truly Gaussian; if the noise distribution is modeled by Gaussian distribution, then the coherent power should have ideally Chi-square distribution with the degree of freedom of 1. On the other hand, Weibull distribution is used because of its ability to adjust the skewness via its shape parameter. In addition, some important distributions such as Rayleigh and Exponential distributions are special cases of Weibull distribution.

After creating SRCP images, adaptive thresholds were computed based on the distribution of the coherent power, statistics of the pixels in the neighborhood, and the specific desired false alarm probabilities. The test pixels are those pixels that have the potential to represent the position of a target. The potential pixels which are capable of representing positions of the sound source (target) have a positive SRCP value as well as being the local maximum value in a neighborhood surrounding them. The pixel under test is compared with the threshold and if it is greater than the threshold, it is counted as the position of the target. Note that there were no sound sources in the experimental recording. Thus, this target detection was a false detection (false alarm). Finally, the total number of pixels which were counted as the positions of sound sources is divided by the total number of pixels to compute the experimental false alarm probability.

4.3.1 Chi-square distribution

For the no-target case, the steered response coherent power (test pixel) was modeled by Chi-square distribution:

$$S_0 \sim \chi^2(d) \tag{4.1}$$

where S_0 is the coherent power (test pixel) and $\chi^2(d)$ represents a Chi-square distribution with *d* degree of freedoms.

Modeling the test pixel by Chi-square distribution resulted in a very poor CFAR performance. In fact, the experimental FA probabilities were always zero (threshold estimates were always too high).

In the next attempt to apply the Chi-square model, the coherent power (test pixel) was squared and then normalized by an estimate of the second moment of the negative pixels. The second moment is estimated by computing the mean value of the squared negative coherent power pixels in the local neighborhood:

$$e_s = \frac{1}{N} \sum_{S_i \in N_0^-} S_i^2 \tag{4.2}$$

where e_s is the estimated second moment, *N* is the total number of negative pixels in the neighborhood and N_0^- denotes the set of negative pixels in the local neighborhood.

In this approach, the test pixel is first squared and then normalized by second moment of the negative pixels. The resulting random variable is modeled by Chi-square distribution with *d* degree of freedoms.

$$\frac{S_0^2}{e_s} \sim \chi^2(d) \tag{4.3}$$

Similar to the first method, the CFAR performance was very poor and the experimental FA probabilities were zero as in the previous case.

With the third attempt, better CFAR performance was achieved by normalizing the pixel under test with the estimated second moment of the negative pixels, and then applying Chi-square distribution with *d* degree of freedoms to model the resulting random variable.

$$\frac{S_0}{e_s} \sim \chi^2(d) \tag{4.4}$$

The second moment is used as the normalizing factor. As it was mentioned in chapter 3, the negative coherent power values are a result of pure noise while the positive values are due to both sound source and noise. The second moment of the test pixel is estimated by using only negative coherent power values in the neighborhood surrounding the test pixel.

The threshold is computed by finding the inverse value of cdf of Chi-square distribution with *d* degree of freedoms computed at the compliment of desired false alarm probability (1-PFA). It is worth noting that the compliment of desired false alarm probability is equal to the cdf computed at the threshold value.

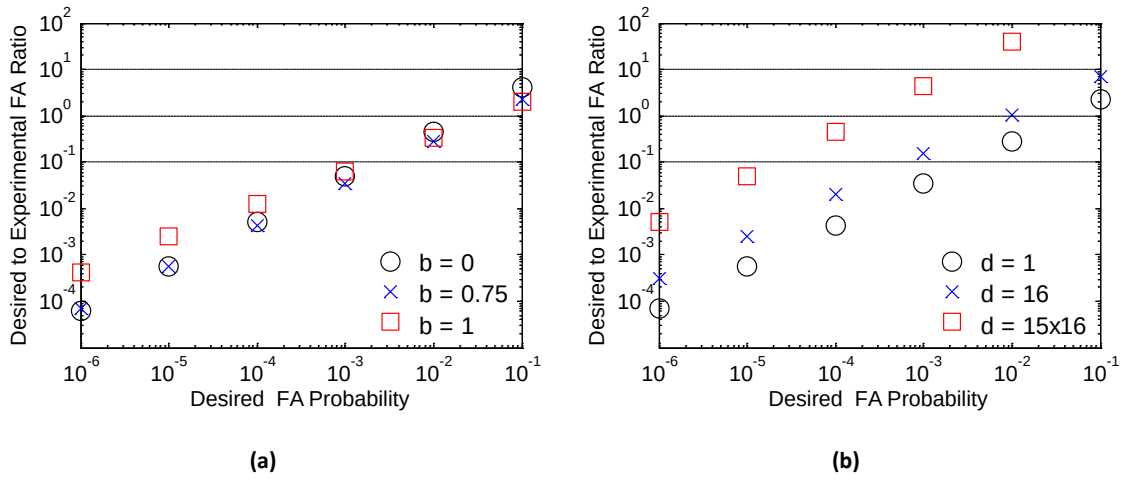

Figure 4.3 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 300 Hz. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution (a) variations of PHAT-β parameters using degree of freedom of 1 for Chi-square distribution (b) variations in Chi-square distribution degree of freedom using beta equal to 0.75.

The high-pass filter cut-off frequency was set to 300 Hz. Figure 4.3 illustrates the ratio of desired to experimental FA (false alarm) probability versus desired probability of FA for six specified desired FA probabilities, ranged from 10^{-1} and 10^{-6} for linear array when the coherent power is modeled by Chi-square distribution with different degree of freedom and the test pixel is normalized by its variance for the no-target situation. The broken line represents ratios of one and implies perfect agreement between experimental and desired FA probabilities. Also, the area between dotted lines is where the ratio of desired over experimental FA probability is at least within 1 order of magnitude. A ratio less than one implies that the experimental threshold was too low, whereas ratios greater than one mean the experimental threshold were too high.

As can be seen from Fig 4.3, whitening tends to improve the CFAR performance but a reasonable CFAR performance cannot be achieved either by applying whitening or by exploiting different degree of freedoms for Chi-square distribution in the linear array when the high-pass filter cut-off frequency is equal to 300 Hz. In fact, the FA probability was under-estimated for high FA probabilities while it was over-estimated for low FA probabilities. Note that the estimated threshold increases by increasing the Chi-square distribution degree of freedom. Consequently, the experimental FA probability decreases by increasing the Chi-square distribution degree of freedom.

Experimental results show that a reasonable CFAR performance cannot be achieved in the linear geometry by modeling the coherent power with Chi-square distribution and normalizing the test pixels by their estimated variance from only negative pixels in the neighborhood even if 1500 Hz is used as the high-pass filter cut-off frequency. Fig 4.4 presents the CFAR performance in the linear array with high-passed filtered signals by cut-off frequencies of 800 Hz and 1500 Hz when the coherent power is modeled by Chi-square distribution and test pixels are normalized by their estimated variance. The PHAT-β parameter is set to 0.75 and different degrees of freedom are examined for the Chi-square distribution. However, none of them led to a reasonable CFAR performance.

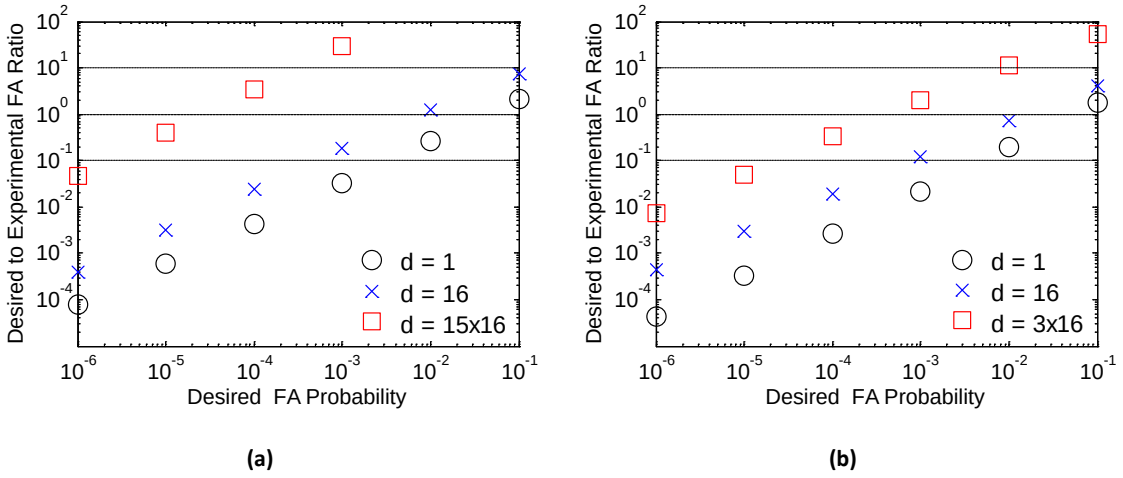

Figure 4.4 Ratios of specified to empirical (experimental) FA probabilities for linear array. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution. The beta value is set to 0.75 and different Chi-square distribution degree of freedoms are applied using high-pass filter cut-off frequency of (a) 800 Hz (b) 1500 Hz.

Perimeter array is the next geometry which performance of CFAR processor is examined in the situation where the coherent power is modeled by Chi-square distribution and the test pixels are normalized by estimated 2nd moment of the negative pixels in the neighborhood. Fig 4.5 shows the performance of the CFAR processor in perimeter array for high-pass filtered signals with a cut-off frequency of 300 Hz. Modeling coherent power by Chi-square distribution with a degree of freedom of 1 did not lead to a reasonable CFAR performance even by examining different values for PHAT-β parameter (Fig 4.5.a). Although a reasonable CFAR performance was achieved for the total whitening case when the degree of freedom was increased to 6 (Fig 4.5.b), for partial whitening cases a reasonable CFAR performance could not be achieved (Fig 4.5.c). It should be noted that the main goal is to achieve a good CFAR performance for partial whitening case with PHAT-β parameter close to 1 as the previous work showed that better correct detection can be achieved by exploiting partial whitening.

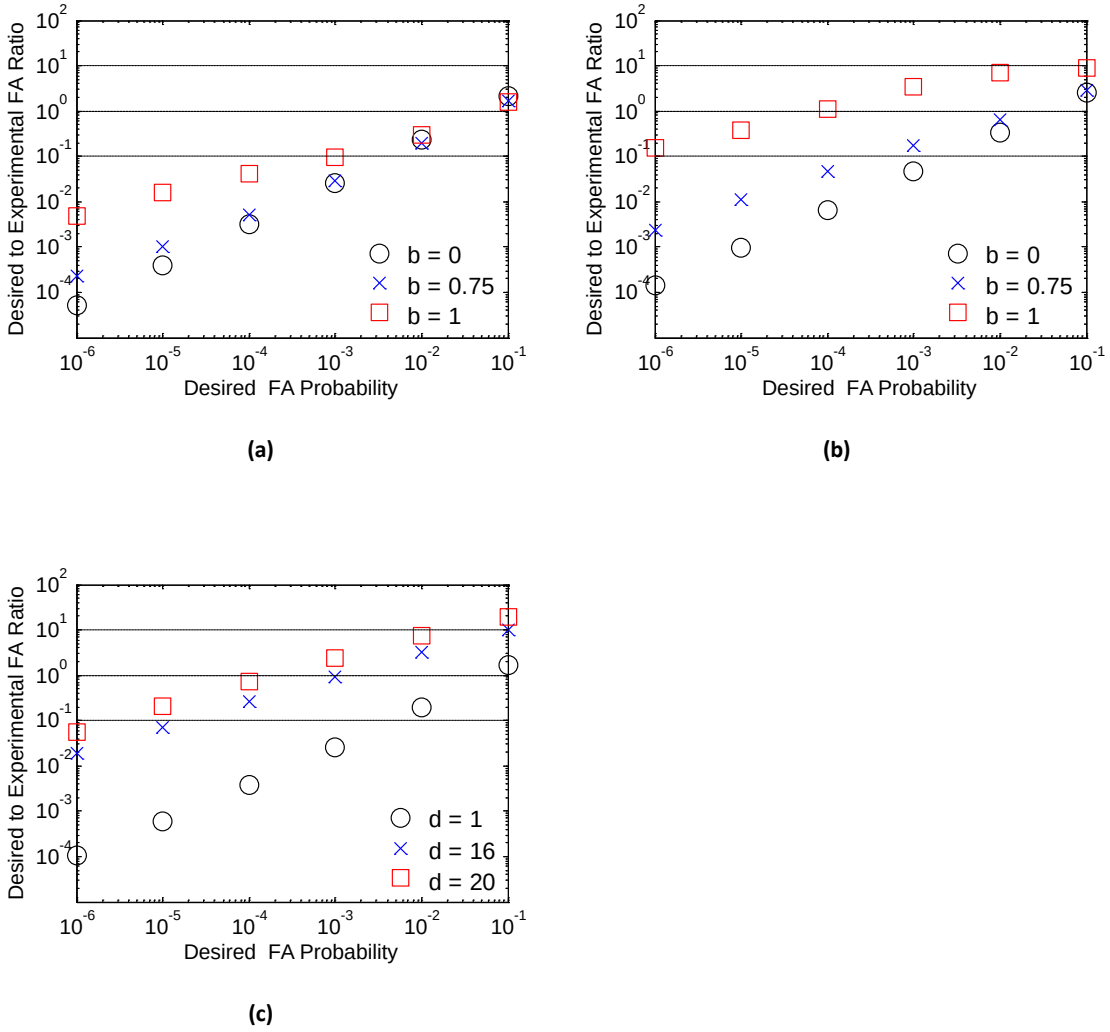

Figure 4.5 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut-off frequency of 300 Hz. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution (a) variations of PHAT-β parameters using degree of freedom of 1 for Chi-square distribution (b) variations of PHAT-β parameters using degree of freedom of 6 for Chi-square distribution (c) variations in Chi-square distribution degree of freedom using beta equal to 0.75.

Finally, the CFAR performance is investigated for the planar geometry when the coherent power is modeled by Chi-square distribution and the test pixels are normalized by their estimated variance. Although whitening tends to improve the CFAR

performance in the planar array, a reasonable CFAR performance was not achieved either by exploiting whitening or by applying different degree of freedoms for the Chi-square distribution. Fig 4.6 presents the CFAR performance in the planar array.

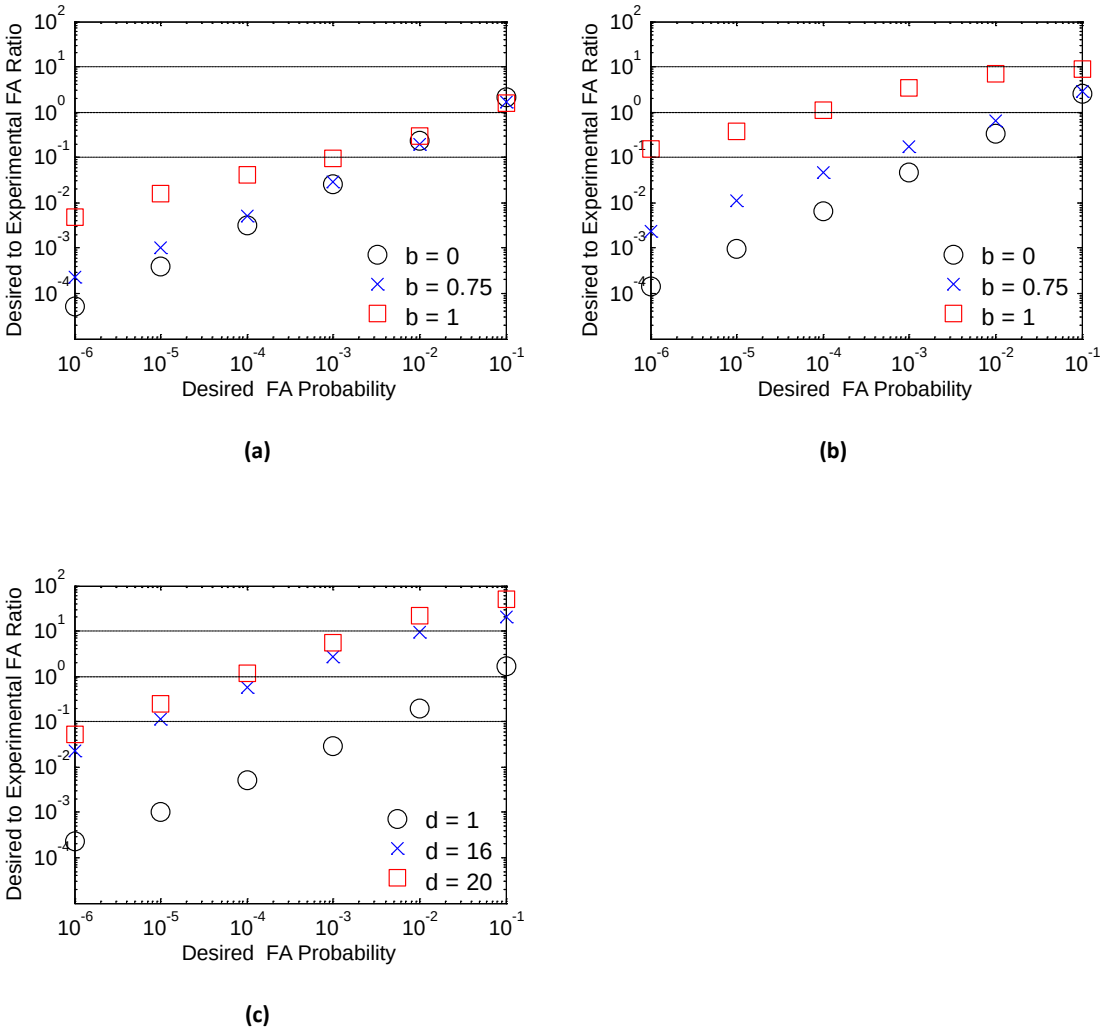

Figure 4.6 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut-off frequency of 300 Hz. The normalized test pixel by the 2nd moments is modeled by Chi-square distribution (a) variations of PHAT-β parameters using degree of freedom of 1 for Chi-square distribution (b) variations of PHAT-β parameters using degree of freedom of 6 for Chi-square distribution (c) variations in Chi-square distribution degree of freedom using beta equal to 0.75.

4.3.2 Developing scaling methods for the Chi-square distribution application

The very poor results of normalizing the test pixels by their estimated variance and modeling the normalized steered response coherent power by Chi-square distribution called for a technique to develop the idea of exploiting Chi-square distribution. The technique that results in a good CFAR performance is to model a linear combination of the steered response coherent power by Chi-square distribution. This approach was motivated by observing the near linear patterns of the FA rates over the probability axis. Therefore, three different linear combinations are applied to the coherent power.

The first approach in developing the idea of exploiting Chi-square distribution is to consider Chi-square distribution as the distribution of the coherent power over a constant value named a_1 and take the expected value from the resultant random variable. It should be noted that the expected value of a Chi-square distribution with degree of freedom *d* is equal to *d*.

$$\frac{S_0}{a_1} \sim \chi^2(d) \tag{4.5}$$

where S_0 is the coherent power and $\chi^2(d)$ represents a Chi-square distribution with *d* degree of freedoms.

Taking expected value from equation (4.5) results in:

$$E\left\{\frac{S_0}{a_1}\right\} = E\{\chi^2(d)\} \tag{4.6}$$

$$\frac{1}{a_1} E\{S_0\} = d \tag{4.7}$$

$$a_1 = \frac{E\{S_0\}}{d} \tag{4.8}$$

The expected value of the coherent power (test pixel) is estimated from only negative pixels in the neighborhood. The absolute value of the negative pixels average is used as the estimated mean value of the pixel under test.

Once a_1 is computed, the test pixel is divided by a_1 and the result is compared with the threshold. The threshold is computed by taking inverse value of the cdf of Chi-square distribution at the compliment of the desired FA probability. This approach will be referred to as *mean-cx2* approach throughout this thesis.

The next approach is called *var-cx2*. *Var-cx2* is similar to the *mean-cx2* method with the difference that instead of exploiting the expected value of Chi-square distribution, the variance of Chi-square distribution is being used in the *var-cx2* approach. It should be noted that the variance of a Chi-square distribution with degree of freedom *d* is equal to *2d*. In the *var-cx2* approach, the constant value which the coherent power is divided by is denoted by a_2. The random variable of the coherent power over a_2 is modeled by Chi-square distribution.

$$\frac{S_0}{a_2} \sim \chi^2(d) \tag{4.9}$$

By taking variance from both sides of the equation (4.9), the constant value a_2 is calculated.

$$Var\left\{\frac{S_0}{a_2}\right\} = Var\{\chi^2(d)\} \tag{4.10}$$

$$\frac{1}{a_2^2} Var\{S_0\} = 2d \tag{4.11}$$

$$a_2 = \sqrt{\frac{Var\{S_0\}}{2d}} \tag{4.12}$$

The variance of the coherent power (pixel under test) is estimated from only negative pixels in the neighborhood. The negative pixels are squared and then averaged to estimate the variance of the test pixel. Once a_2 is computed the test pixel is divided by a_2 and the resultant random variable is compared with the threshold. Similar to the *mean-cx2* approach, the thresholds are computed by taking the inverse value of the cdf of Chi-square distribution at the compliment of the desired FA probability.

In the last approach for developing the idea of exploiting Chi-square distribution, the coherent power (test pixel) is divided by a constant value called *a* and the result is added with another constant value denoted by *b*. The resultant random variable is then modeled by Chi-square distribution.

$$\left(\frac{S_0}{a} + b\right) \sim \chi^2(d) \tag{4.13}$$

This approach is named *mean-var-cx2* in this thesis. In the *mean-var-cx2* method, both expected value and variance are used. If the expected value is taken from equation (4.13), then:

$$E\left\{\frac{S_0}{a} + b\right\} = d \tag{4.14}$$

or

$$\frac{1}{a} E\{S_0\} + b = d \tag{4.15}$$

And taking the variance from equation (4.13) will result in:

$$Var\left\{\frac{S_0}{a} + b\right\} = 2d \tag{4.16}$$

or

$$\frac{1}{a^2} Var\{S_0\} = 2d \tag{4.17}$$

Therefore, the constant value of *a* is computed from:

$$a = \sqrt{\frac{Var\{S_0\}}{2d}} \tag{4.18}$$

By projecting the *a* value into equation (4.15) the constant value b is calculated:

$$b = d - \frac{E\{S_0\}}{a} \tag{4.19}$$

The expected value of the coherent power (test pixel) is computed similarly as it was computed in the *mean-cx2* approach. Also, the variance of the pixel under test is estimated by using the same technique in the *var-cx2* approach.

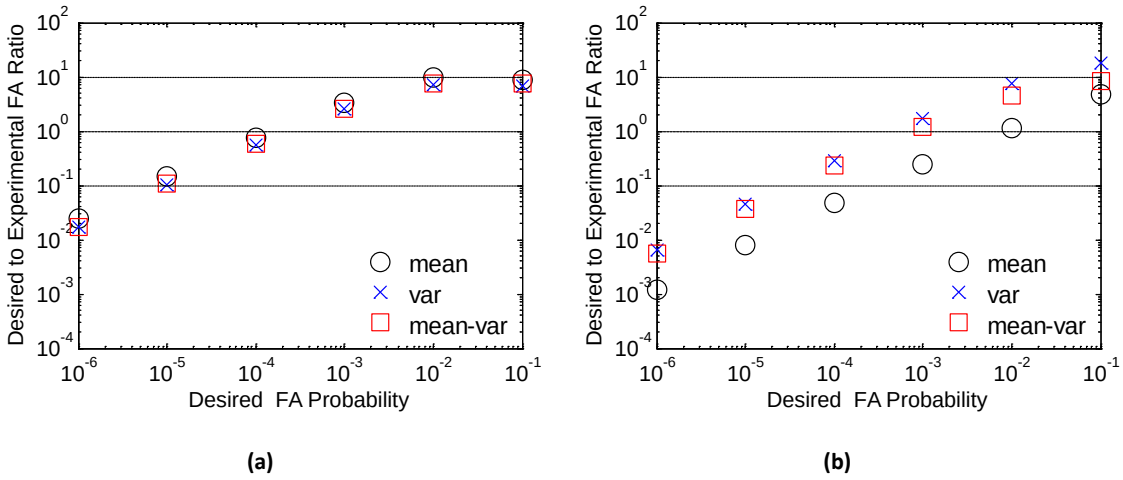

Figure 4.7: Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85. The performances of modeling three linear combinations of the coherent power by Chi-square distribution are compared together (a) for degree of freedom of 1 (b) for degree of freedom of 6.

Now that constant values *a* and *b* are computed, the coherent power is divided by *a* and *b* is added to the result. Finally, the created linear combination of the coherent power is compared with the threshold.

To investigate the performance of the CFAR processor by using the developed Chi-square distribution approaches, the PHAT-β parameter is set to 0.85 and the high-pass

cut-off frequency is set to 300 Hz. Fig 4.7 shows the performance of the developed Chi-square distribution approaches in the linear array.

For the high FA probabilities, the experimental FA probabilities were too low while for the low FA probabilities the empirical FA probabilities were too high in all three approaches. Note that even by increasing the degree of freedom of Chi-square distribution, a reasonable CFAR performance was not achieved. In fact, increasing the Chi-square distribution degree of freedom increases the empirical FA probabilities and hence the CFAR performance even degrades for low FA probabilities. Consequently, the high-pass filter cut-off frequency needs to be increased.

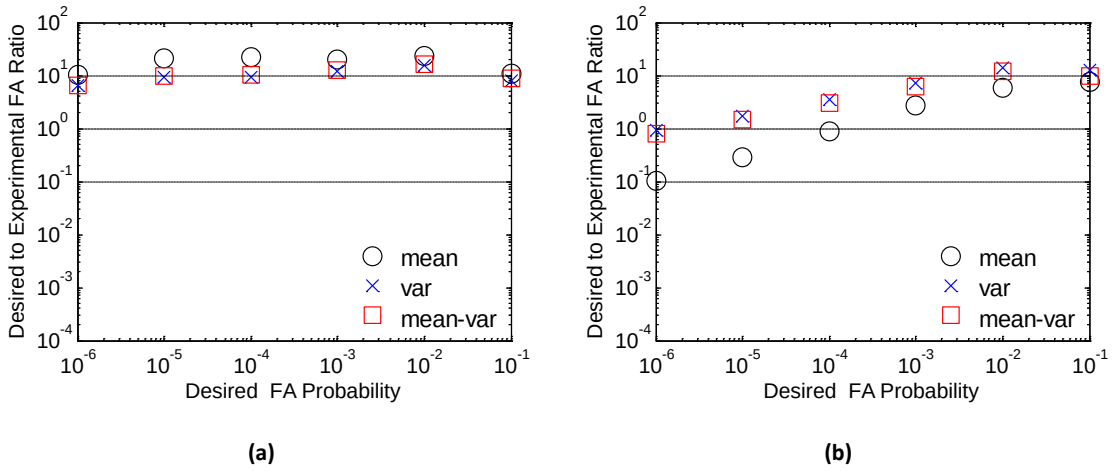

Figure 4.8: Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 800 Hz. The PHAT-β parameter is set to 0.85. The performances of modeling three linear combinations of the coherent power by Chi-square distribution are compared together (a) for degree of freedom of 1 (b) for degree of freedom of 2.

The results of increasing the cut-off frequency to 800 Hz are presented in Fig 4.8. It is interesting to note that the CFAR performance for all three approaches are close together when the Chi-square distribution degree of freedom is equal to 1 and by increasing the degree of freedom, the experimental FA probability increases. The *mean-cx2* method is the most sensitive approach to the Chi-square degree of freedom relevant to other approaches. As can be seen from Fig 4.8 the FA probabilities are under-estimated for all approaches. However, if the degree of freedom is increased from 1 to 2, then the FA probabilities are over-estimated for low desired FA probabilities. The CFAR performance is improved by increasing the cut-off frequency from 300Hz to 800Hz, especially when the degree of freedom is set to 2. However, the cut-off frequency still needs to be increased more in order to achieve a good CFAR performance in linear geometry. Fig 4.9 shows the CFAR performance for linear array using a cut-off frequency of 1500 Hz. The PHAT-β parameter is set to 0.85 and three different combinations of the coherent power are modeled by Chi-square distribution.

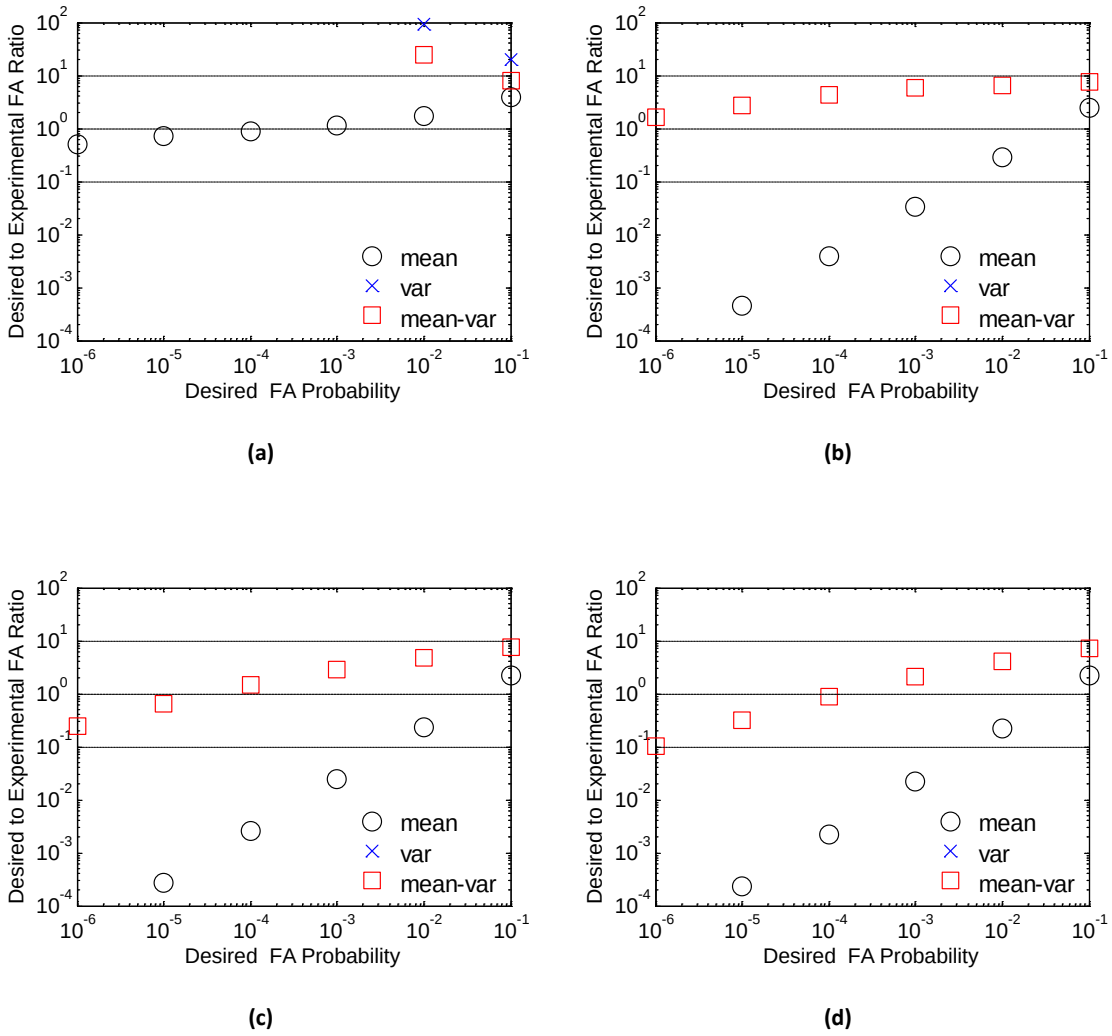

Figure 4.9: Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 1500 Hz. The PHAT-β parameter is set to 0.85. The performance of modeling three linear combinations of the coherent power by Chi-square distribution are compared together for degree of freedom of(a) 4 (b) 48 (c) 240 (d) 1600.

As Fig 4.9a demonstrates, a very good CFAR performance for linear array is achieved in the situation where the *mean-cx2* approach is exploited and the degree of freedom is set to 4 using 0.85 as the PHAT-β parameter and high-pass filtering signals with cut-off frequency of 1500 Hz. The experimental FA probability is increased, if the degree of freedom of Chi-square distribution is increased, and in consequence, the ratios of

desired over empirical FA probabilities are decreased. The ratio was less than 10^{-4} for desired FA probability equal to 10^{-6} in Figs 4.9b-d and therefore was not shown in these figures.

An interesting phenomenon is the very wide range for Chi-square distribution degree of freedom in which the ratios of desired over empirical FA probabilities are within 1 order of magnitude (a reasonable CFAR performance) when the *mean-var-cx2* approach is exploited. The range for degree of freedom of Chi-square distribution which resulted in a reasonable CFAR performance using the *mean-var-cx2* method is 48 to 1600. On the other hand, a reasonable CFAR performance is not achieved for linear array in the situation where the *var-cx2* approach is used. In this situation, the FA probability is always under-estimated. The ratios of experimental to desired FA probabilities were beyond 10^2 and therefore were not shown in the figures.

The performance of modeling the linear combinations of the coherent power by Chi-square distribution in the perimeter array with high-passed filtered signals with a cut-off frequency of 300 Hz is identical to the performance of linear geometry with high-passed filtered signals with a cut-off frequency of 1500 Hz. Identical to linear array using cut-off frequency of 1500 Hz, a good CFAR performance is achieved by exploiting the *mean-cx2* approach in the perimeter array with 300 Hz as the high-pass filter cut-off frequency while the Chi-square distribution degree of freedom is set to 4. Also, a reasonable CFAR performance is achieved by exploiting the *mean-var-cx2* method for Chi-square distribution degree of freedom ranging from 48 to 1600. Furthermore, similar to linear geometry, a reasonable CFAR performance is not achieved by using the *var-cx2* approach in perimeter array. Fig 4.10 demonstrates the ratios of experimental to desired FA probabilities in perimeter array with high-passed filtered signals with a cut-off frequency of 300 Hz and using PHAT-β parameter of 0.85 when a linear combination of the coherent power is modeled by Chi-square distribution.

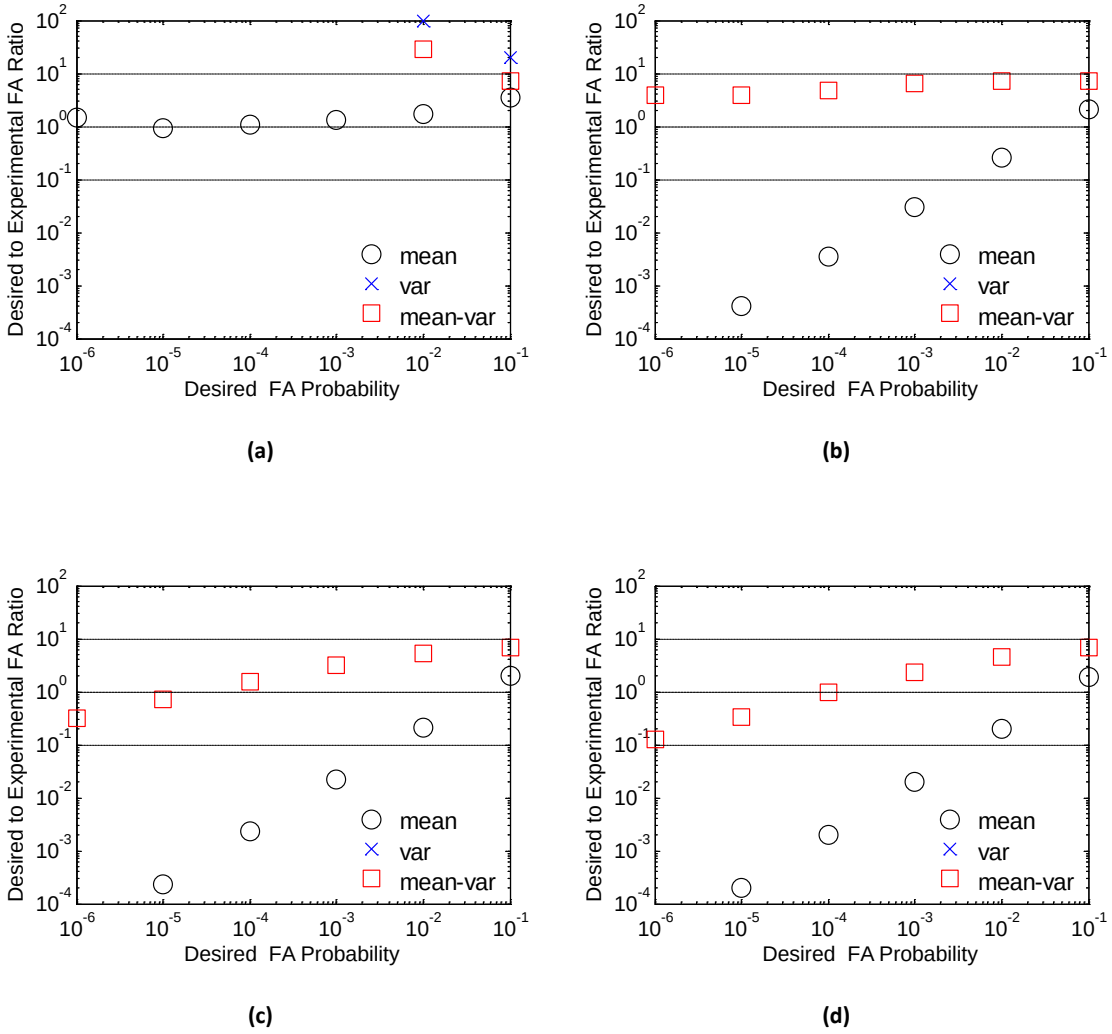

Figure 4.10 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85. The performance of modeling three linear combinations of the coherent power by Chi-square distribution are compared together for degree of freedom of(a) 4 (b) 48 (c) 240 (d) 1600.

If the planar array is used and a linear combination of the coherent power is modeled by Chi-square distribution, a good CFAR performance can be achieved by high-pass filtering signals with a cut-off frequency of 300 Hz. In this situation, the performance of planar geometry is similar to the linear array using 1500 Hz as the cut-off frequency and the perimeter array using cut-off frequency of 300 Hz; however, they have some

differences. As it was presented above, by exploiting the *mean-cx2* approach and setting the Chi-square distribution degree of freedom to 4, a good CFAR performance can be achieved in the linear array as well as in the perimeter array. However, if the planar geometry is used, the degree of freedom should be set to 3 in order to result in a good CFAR performance using the *mean-cx2* method.

On the other hand, the range for Chi-square distribution degree of freedom which resulted in a reasonable CFAR performance in linear and perimeter geometries using the *mean-var-cx2* approach is much wider than the range in the planar array. In the planar array, a reasonable CFAR performance is achieved by exploiting the *mean-var-cx2* method for a Chi-square distribution degree of freedom ranging from 10 to 32.

Furthermore, similar to linear and perimeter geometries, a reasonable CFAR performance was not achieved for the var-cx2 method in the planar array. The FA probabilities are under-estimated for all degrees of freedom, if the *var-cx2* method is used. Fig 4.11 presents the CFAR performance for the planar geometry using a cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85 and the performance of the *mean-cx2*, *var-cx2* and *mean-var-cx2* approaches are presented and compared together.

By comparing the CFAR performance in all of the three microphone geometries when using the *mean-cx2*, *var-cx2* and *mean-var-cx2* approaches to model the steered response coherent power, it can be concluded that the *mean-cx2* method outperforms in terms of higher agreement between experimental and desired FA probabilities while a reasonable CFAR performance cannot be achieved by using the *var-cx2* method. Furthermore, in the situation where the *mean-cx2* approach is exploited to achieve a perfect CFAR performance, the Chi-square distribution degree of freedom needs to be set to 4 in the linear and perimeter arrays, whereas a good CFAR performance is achieved in the planar array if the degree of freedom is equal to 3.

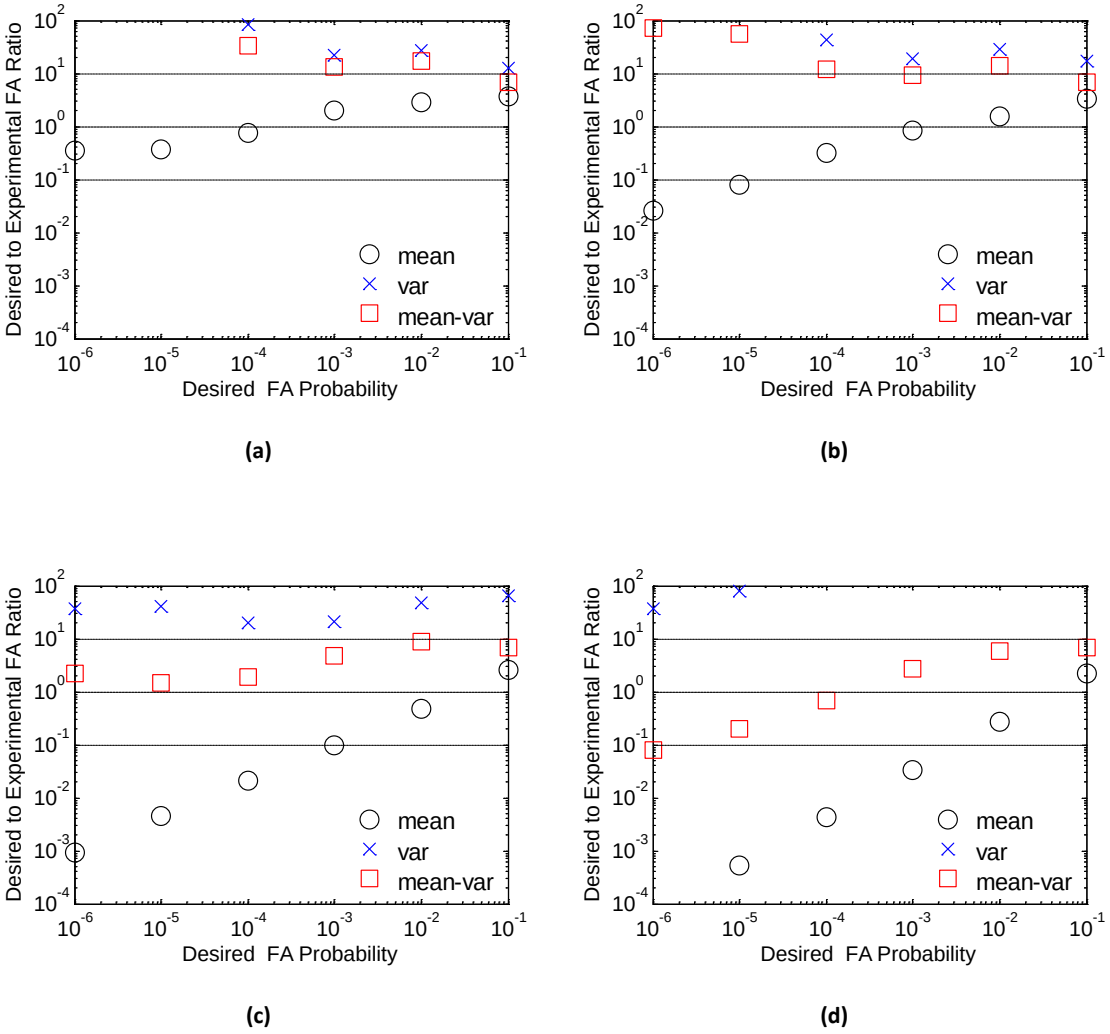

Figure 4.11 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut-off frequency of 300 Hz. The PHAT-β parameter is set to 0.85. The performance of modeling three linear combinations of the coherent power by Chi-square distribution are compared together for degree of freedom of(a) 3 (b) 4 (c) 10 (d) 32.

All of the above performances were resulted by exploiting the neighborhood size of 15x15 pixels (a radius of 7 around the test pixel). The neighborhood size has an important impact on the performance of the CFAR processor. In the Chi-square approaches, the mean-value and the second moment of the test pixel are estimated from the negative pixels in the neighborhood. The accuracy of the estimates depends on

the total number of samples used for the estimations. The estimated values are more accurate if more samples are used or a larger neighborhood size is exploited.

On the other hand, because of the non-stationary nature of the data, in order to consider the same distribution for the pixels in the neighborhood, it is required to keep the neighborhood size small enough.

In this study, three different neighborhood sizes are examined in order to investigate the effect of the neighborhood size on the performance of the CFAR processor. These three neighborhood sizes are 7x7, 15x15 and 21x21 pixels. Experimental results demonstrate that the neighborhood size of 15x15 pixels outperforms in terms of better CFAR performance for all microphone geometries. Fig 4.12 compares the performance of different neighborhood sizes using the *mean-cx2* and *mean-var-cx2* approaches for linear array with high-passed filtered signals with a cut-off frequency of 1500 Hz and exploiting 0.85 for the PHAT-β parameter. The Chi-square distribution is set to the value which resulted in a good CFAR performance for the neighborhood size of 15x15 pixels.

By increasing the neighborhood size, the ratio of experimental over desired FA probability increases. Equivalently, the experimental FA probability decreases if the neighborhood size increases for both the *mean-cx2* and *mean-var-cx2* methods. A reasonable CFAR performance cannot be achieved for the linear geometry if the neighborhood size is set to 7x7 pixels. On the other hand, with the neighborhood size equal to 21x21 pixels, a reasonable CFAR performance is achievable but the Chi-square distribution degree of freedom needs to be increased (e.g. the degree of freedom should be increased from 4 to 5 for the *mean-cx2* approach in the linear array). Furthermore, if the neighborhood size is equal to 21x21 pixels and the *mean-var-cx2* method is used the range for degree of freedom which results in a reasonable CFAR performance will be increased such that even a high degree of freedom, such as 5600 for the linear geometry, leads to a reasonable CFAR performance.

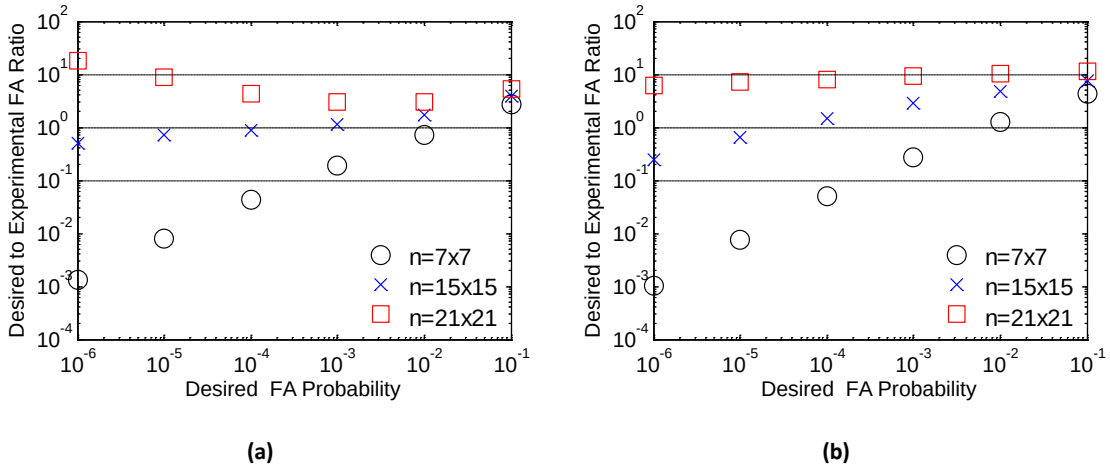

Figure 4.12 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut-off frequency of 1500 Hz. The beta is equal to 0.85. Variations of the neighborhood size using (a) *mean-cx2* approach and degree of freedom of 4 for Chi-square distribution (b) *mean-var-cx2* method and degree of freedom of 240.

Similar to the linear array, increasing the Chi-square distribution degree of freedom decreases the experimental FA probability for the perimeter geometry for both the *mean-cx2* and *mean-var-cx2* methods. The degree of freedom equal to 2 will lead to a reasonable CFAR performance for the perimeter array if the neighborhood size is equal to 7x7 pixels using the *mean-cx2* method. Although a reasonable CFAR performance is achieved for the perimeter array for all three neighborhood sizes, the CFAR performance is the best if the neighborhood size is equal to 15x15 pixels. The higher the neighborhood size, the higher the degree of freedom needs to be set to achieve a reasonable CFAR performance. Furthermore, the range for the degree of freedom is increased by increasing the neighborhood size if the *mean-var-cx2* approach is used. Fig 4.13 presents the CFAR performance in the perimeter array using different neighborhood sizes for high-passed filtered signals with a cut-off frequency of 300 Hz and exploiting 0.85 for the PHAT-β parameter.

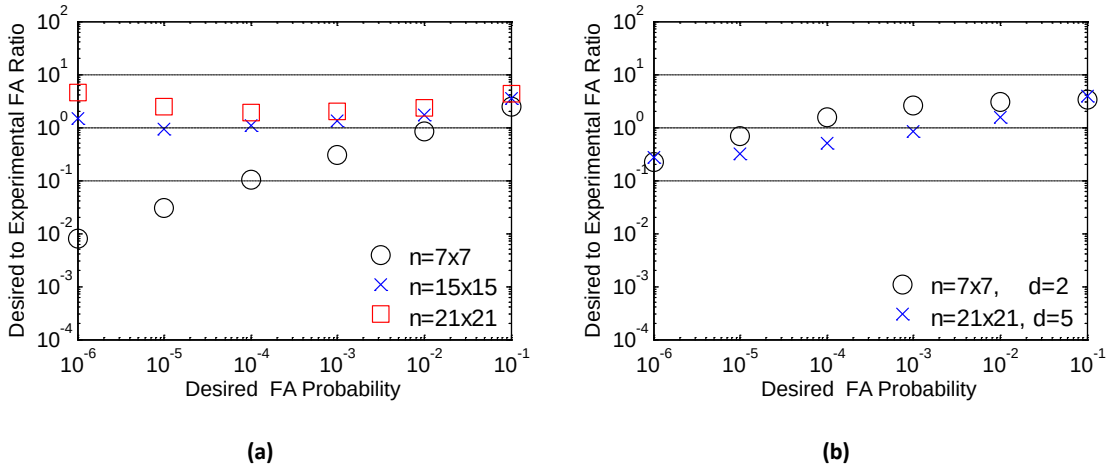

Figure 4.13 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut-off frequency of 300 Hz. The beta is equal to 0.85. The *mean-cx2* approach is applied (a) variations of the neighborhood size using Chi-square distribution degree of freedom of 4 (b) for neighborhood size of 7x7 pixels the degree of freedom should be set to 2 while for neighborhood size of 21x21 pixel it can be either 4 or 5 to achieve a reasonable CFAR performance.

By increasing the degree of freedom, the ratio of experimental to desired FA probability decreases for the planar geometry like the linear and perimeter arrays. Furthermore, similar to the linear geometry, a reasonable CFAR performance cannot be achieved using the neighborhood size of 7x7 pixels for the planar array. Also, if the neighborhood size is equal to 21x21 pixels, the degree of freedom needs to be increased to achieve a reasonable CFAR performance. (Although the degree of freedom equal to three leads to a reasonable CFAR performance using neighborhood size of 21x21 pixels, the best performance for the neighborhood size of 21x21 is achieved when the degree of freedom is set to 4 using the *mean-cx2* method). Fig 4.14 shows the CFAR performance for different neighborhood sizes for the planar array. The signals are high-pass filtered with the cut-off frequency of 300 Hz and also they have been whitened with a PHAT-β parameter of 0.85.

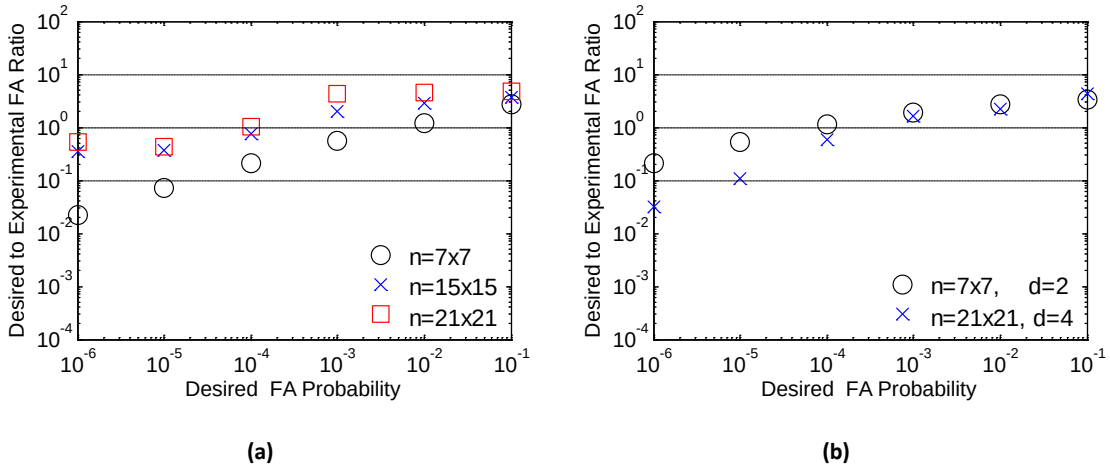

Figure 4.14 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut-off frequency of 300 Hz. The beta is equal to 0.85. The *mean-cx2* approach is applied (a) variations of the neighborhood size using Chi-square distribution degree of freedom of 3 (b) for neighborhood size of 7x7 pixels a reasonable CFAR performance is not achievable while for neighborhood size of 21x21 pixel the degree of freedom can be either 3 or 4 to achieve a reasonable CFAR performance.

In summary, the neighborhood size of 15x15 pixels has the best CFAR performance for all microphone geometries. In addition, the linear array is the most sensitive geometry in respect to the neighborhood size, whereas the planar array is the least sensitive geometry. Furthermore, the *mean-cx2* approach outperforms among other scaling approaches. In the situation where the *mean-cx2* method is used to model the coherent power and the local neighborhood size is set to 15x15, the optimal Chi-square distribution degree of freedom is equal to 4 for the linear and perimeter geometries, while to achieve a good CFAR performance for the planar array, the degree of freedom should be equal to 3.

4.3.3 Weibull distribution

This section presents CFAR threshold performance where the coherent power is modeled with Weibull distribution. The same distribution with the same parameters is considered for both negative and positive coherent power values. Variations in the shape parameter, the influence of beta value (PHAT-β parameter), and the high pass filter cut-off frequency on the performance for three microphone distributions will be investigated. For the high-pass filter designing purpose, only the *mic-distribution* factor will be considered.

As it was mentioned in chapter 3, the Weibull distribution is used because of its ability to model the skewness in the tail of the distribution by adjusting the shape parameter. In addition, some powerful distributions, such as *Rayleigh* and *Exponential,* are special cases of Weibull distribution. In this approach, the coherent power (test pixel) is modeled by Weibull distribution with the parameters of *a* as the *scale parameter* and *b* as the *shape parameter*:

$$S_0 \sim Wei(a, b) \tag{4.20}$$

If the original noise distribution is assumed to be Gaussian distribution and be locally stationary, then the coherent power can be modeled by Chi-square distribution. Experiments show that Weibull distribution with shape parameter of about 1.26 corresponds to Chi-square distribution. Therefore, the shape parameter of Weibull distribution is initially set to 1.26.

Using the maximum likelihood estimation, the scale parameter is computed for a known shape parameter, *b*, from only negative pixels in the local neighborhood.

$$\hat{a} = \left(\frac{1}{N} \sum_{S_i \in N_0^-} |S_i|^b \right)^{\frac{1}{b}} \tag{4.21}$$

where *N* is the total number of the negative pixels in the neighborhood and N_0^- denotes the set of negative pixels in the local neighborhood. Note that since the Weibull distribution is not defined for the negative values, the absolute values of the negative pixels are taken before being raised by the shape parameter.

The scale parameter can be estimated by exploiting the expected value of Weibull distribution alternatively. Using expected value of Weibull distribution, the scale parameter is estimated for a known shape parameter, *b*, from only the negative pixels in the local neighborhood.

$$\hat{a} = \frac{E\{S_0\}}{\Gamma\left(1+\frac{1}{b}\right)} \tag{4.22}$$

where $\Gamma(.)$ denotes gamma function. The expected value of the Weibull distribution is estimated by finding the absolute value of the average of the negative pixels in the local neighborhoods.

By finding the inverse value of Weibull distribution cdf at the compliment of the desired FA probability, the adaptive threshold is computed:

$$T = \hat{a}\ln\left(\frac{1}{P_{FA}}\right)^{\frac{1}{b}} \tag{4.23}$$

where P_{FA} represents the desired FA probability.

Experimental results show no differences between the CFAR performances when the scale parameter is computed by the maximum likelihood approach and exploiting the expected value of Weibull distributions for all possible scenarios. However, note that the computational cost is less if the expected value method is applied to estimate the scale parameter for real time applications, rather than exploiting the maximum likelihood approach as the gamma function needs to be computed once in the expected value method. On the other hand, in the maximum likelihood approach, the negative values need to be raised by a fractional power.

Figure 4.15 illustrates the ratio of experimental to desired FA probability versus desired probability of FA for the linear array with high-passed filtered signals using 300 Hz as the cut-off frequency when the coherent power is modeled by Weibull distribution. The CFAR threshold performance for the linear array with low cut-off frequencies such as 300 Hz is not accurate. For low desired probability of FAs, the experimental FA probabilities tend to be overestimated relevant to the desired one. This over estimating FA probability means the experimental threshold is less than the desired one.

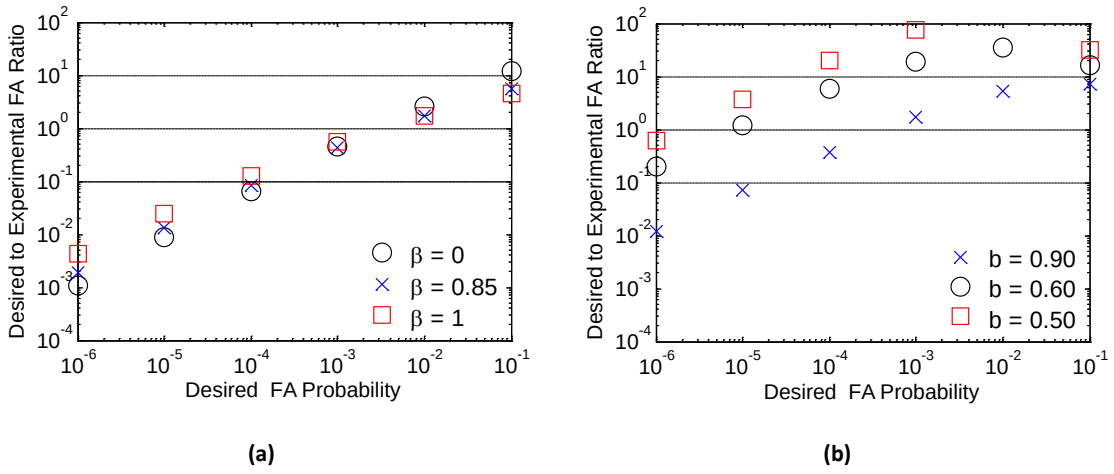

Figure 4.15 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. (a) Variations of PHAT-β parameters using shape parameter of 1.26 (b) variations of shape parameters using beta equal to 0.85.

On the other hand, for high desired probability of FA, the experimental FA probabilities tend to be underestimated which means the experimental threshold is higher than the desired one. As can be seen from Fig 4.15.a, whitening improves the performance. However, the previous works showed that partial whitening resulted in significantly better detection rates, rather than total whitening. Therefore, beta-- partial spectral whitening is set to 0.85 and the shape parameter is adjusted to make the experimental FA probability as close to the desired one as possible (Fig 4.15.b). However, it is too

complicated to find a shape parameter that makes the ratio of the desired to the experimental FA probability close to one. In other words, fitting Weibull distribution to the data is too complicated. In the linear array, lower shape parameter performs better at low FA probabilities, while higher shape parameter performs better at the high probability of false alarms. The best performance that can be achieved for the linear array with a cut-off frequency of 300 Hz is for 0.6 as the shape parameter. Even a low shape parameter such as 0.6 which cause a high skewness, could not keep the ratio at least in order of 1 magnitude at the high FA probabilities. In fact, Weibull distribution is not a good distribution to model both the positive and negative values of the coherent power (as seen from the figures), and the reason is the inability of the specific microphone distribution to decorrelate the lower frequency components. Among all microphone distributions, the linear array has the least inter-path distance variance relative to the source wavelength. Therefore, greater deviations from symmetry are expected for the linear array. As a consequence, near-symmetry for the linear array is only possible at higher source frequencies.

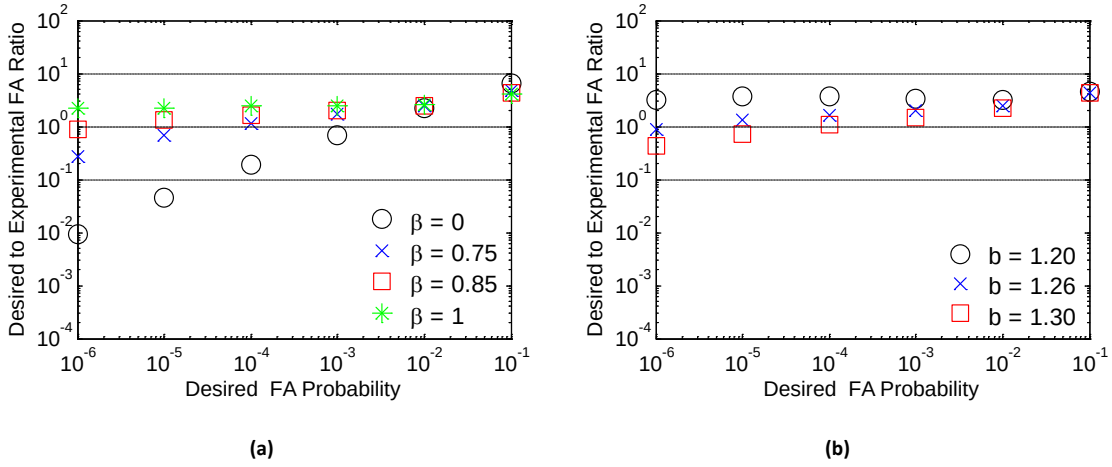

Figure 4.16 Ratios of specified to empirical (experimental) FA probabilities for linear array for high-pass filtered signals with cut off frequency of 1500 Hz when the coherent power is modeled by Weibull distribution. (a) Variations of PHAT-β parameters using shape parameter of 1.26 (b) variations in shape parameters using beta equal to 0.85.

Improved CFAR performance can be achieved in the linear array by removing the lower frequency content of the source signals using an increased cut-off frequency for the high-pass filter (Fig 4.16). The higher the cut-off frequency results in a higher inter-path distance variance relative to the source wavelength and consequently, the noise distribution will be more symmetric. By increasing the cut-off frequency, the improvement of CFAR performance via whitening becomes more observable. It should be noted that increasing the cut off frequency results in a loss of lower source frequencies, which skew the distribution. Figure 4.16.a shows the effect of whitening when 1500 Hz is used as the high-pass filter cut-off frequency. As Fig 4.16.b shows partial whitening performs better. The purpose of Fig 4.16.b is to demonstrate the effect of shape parameter on the CFAR performance for a fixed beta value equal to 0.85 when using 1500 Hz as the cut-off frequency.

Among all microphone distributions, the perimeter array has the highest inter-path distance variance relative to the source wavelength. Therefore, noise distribution has a mean value closer to 0 and is more symmetric (see equation 2.18). Consequently, the perimeter array should have the best performance relative to the two other geometries.

Figure 4.17 illustrates the CFAR performance for the perimeter array using a cut-off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. As Fig 4.17.a shows, whitening improves the performance of CFAR threshold in the perimeter array (the shape parameter is set to 1.26). Also, it can be seen that partial whitening has better performance relative to non-whitening, or total whitening cases.

For a better detection rate, beta is set to 0.85 [20,21]. In fact, Weibull can be fitted to data reasonably by adjusting the shape parameter. And the experimental probability of false alarm is very accurate (Fig 4.17.b).

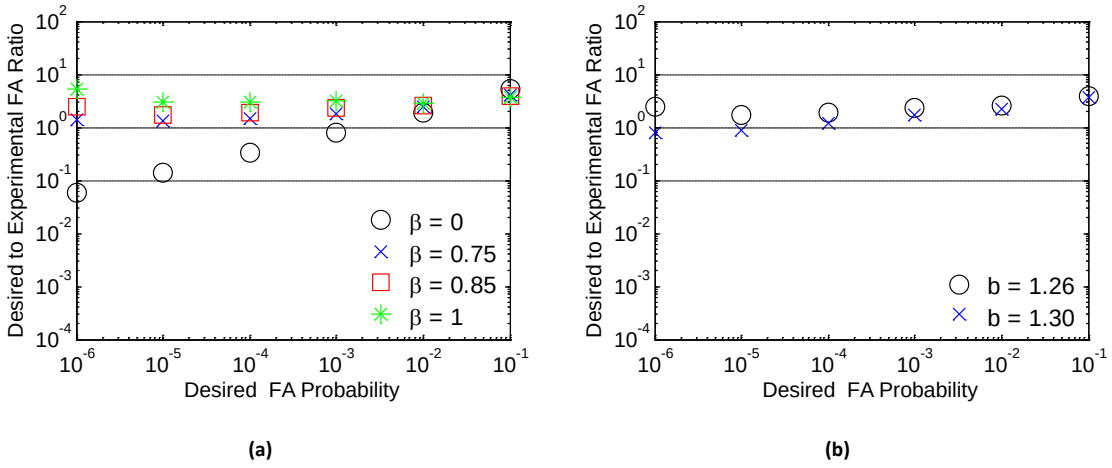

Figure 4.17 Ratios of specified to empirical (experimental) FA probabilities for perimeter array for high-pass filtered signals with cut off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. (a) Variations in PHAT-β parameters using shape parameter of 1.26 (b) variations in shape parameters using beta equal to 0.85.

In the planar array, whitening does not have an observable effect on the CFAR threshold performance. In fact, the shape parameter of Weibull distribution has the main influence on the performance in the planar geometry. Decreasing the shape parameter, or in other words increasing skewness, will cause the experimental FA probability to decrease relevant to using a higher shape parameter; especially, for the low FA probabilities. In consequence, the ratio of desired to experimental FA probability will increase.

Figure 4.18 exhibits the CFAR performance for the planar distribution using a cut-off frequency of 300 Hz and exploiting Weibull distribution to model the coherent power. In Fig 4.18.a the shape parameter is set to 1.26. Since the probability of false alarm is

overestimated at the low FA probabilities, the shape parameter should be decreased (Fig 4.18.b)

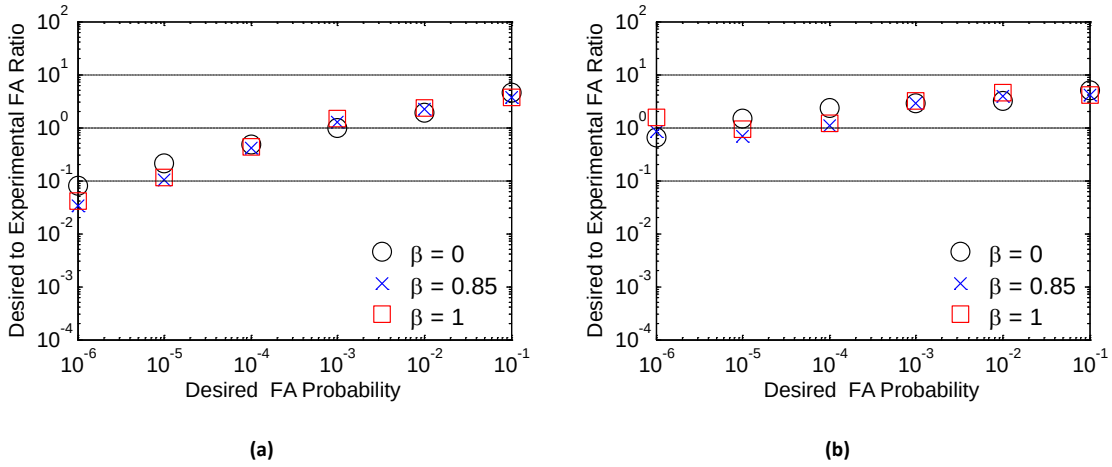

Figure 4.18 Ratios of specified to empirical (experimental) FA probabilities for planar array for high-pass filtered signals with cut off frequency of 300 Hz when the coherent power is modeled by Weibull distribution. (a) Variations in PHAT-β parameters using shape parameter of 1.26 (b) variations in PHAT-β parameters, using shape parameter of 1.12.

There are two issues which need to be discussed in the situation where Weibull distribution is used. One issue is to examine the CFAR performance when the coherent power is squared and the squared coherent power is modeled by Weibull distribution. The other issue is to select an appropriate neighborhood size.

If modeling the squared of the coherent power by Weibull distribution is desired, the scale parameter is estimated from only the negative pixels in the local neighborhood.

$$S_0^2 \sim Wei(\theta, \varphi) \tag{4.24}$$

The negative pixels are first squared and then projected into equation (4.21) to estimate the scale parameter, θ for a known shape parameter, φ. Experimental results show that by modeling the squared of the coherent power via Weibull distribution, the exact same CFAR performances are achieved as the CFAR performances when the

coherent power is modeled by Weibull distribution, with the difference that the shape parameter φ becomes half of the shape parameter b.

$$\varphi = \frac{b}{2} \tag{4.25}$$

Similar to the Chi-square approach, there is a similar discussion about the local neighborhood size for Weibull distribution. The CFAR performances shown above were a result of selecting 15x15 pixels for the local neighborhood size.

The scale parameter of the Weibull distribution is computed from the negative pixel values in the local neighborhood. The size of the local neighborhood should be selected such that the data in the local neighborhood is stationery. On the other hand, reducing the size of the local neighborhood causes inaccuracy in the estimation of the scale parameter due to lack of enough samples from which the scale parameter is estimated. As a result, the CFAR performance will be degraded by reducing the size of the local neighborhood. Reducing the size of the local neighborhood will cause a reduction in the experimental threshold. In consequence, the experimental FA probability will be higher than the desired one. Therefore, the shape parameter needs to be reduced in order to make the ratio of desired to experimental probability of FA closer to one. Figure 4.19 shows the performance of CFAR processor for different neighborhood sizes for the three microphone distributions. The partial whitening value is set to 0.85. The shape parameter of Weibull distribution is selected such that the best performance can be achieved using 15x15 as the size of the local neighborhoods.

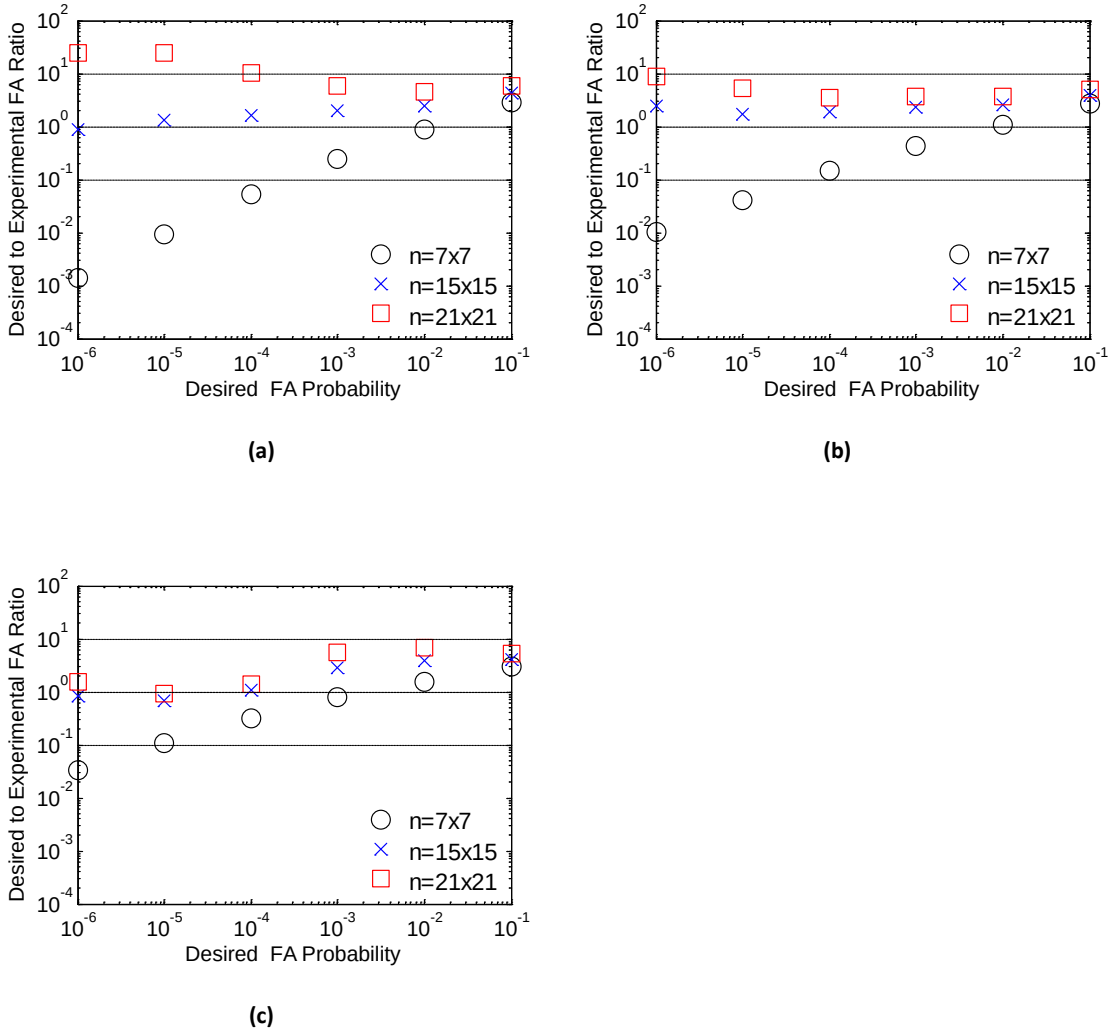

Figure 4.19 CFAR Performance for 3 different neighborhood sizes using Weibull distribution, using partial whitening value of 0.85 for (a) linear array with 1500 Hz cut of frequency and using shape parameter of 1.26 (b) perimeter array with 300 Hz cut of frequency and using shape parameter of 1.26 (c) planar array with 300 Hz cut-off frequency and using shape parameter of 1.12.

Fig 4.20 illustrates the performance of the CFAR processor, using a local neighborhood size of 7x7 for three microphone geometries. As can be seen from Fig 4.20, the linear array is most sensitive whereas the planar array is less sensitive relative to the size of the local neighborhood. In the linear geometry, reducing the size of the local neighborhood to 7x7 degrades severely the performance of the CFAR processor. Even

using a high value cut-off frequency such as 1500 Hz cannot cause the ratio of desired to experimental FA probability to be within at least one order of magnitude. Although reducing size of the local neighborhood to 7x7 degrades the CFAR performance in the perimeter and planar arrays, the performances of the CFAR processor in the perimeter and planar geometries are still reasonable and the ratios of desired to experimental FA probability are still within one order of magnitude by exploiting a lower shape parameter value.

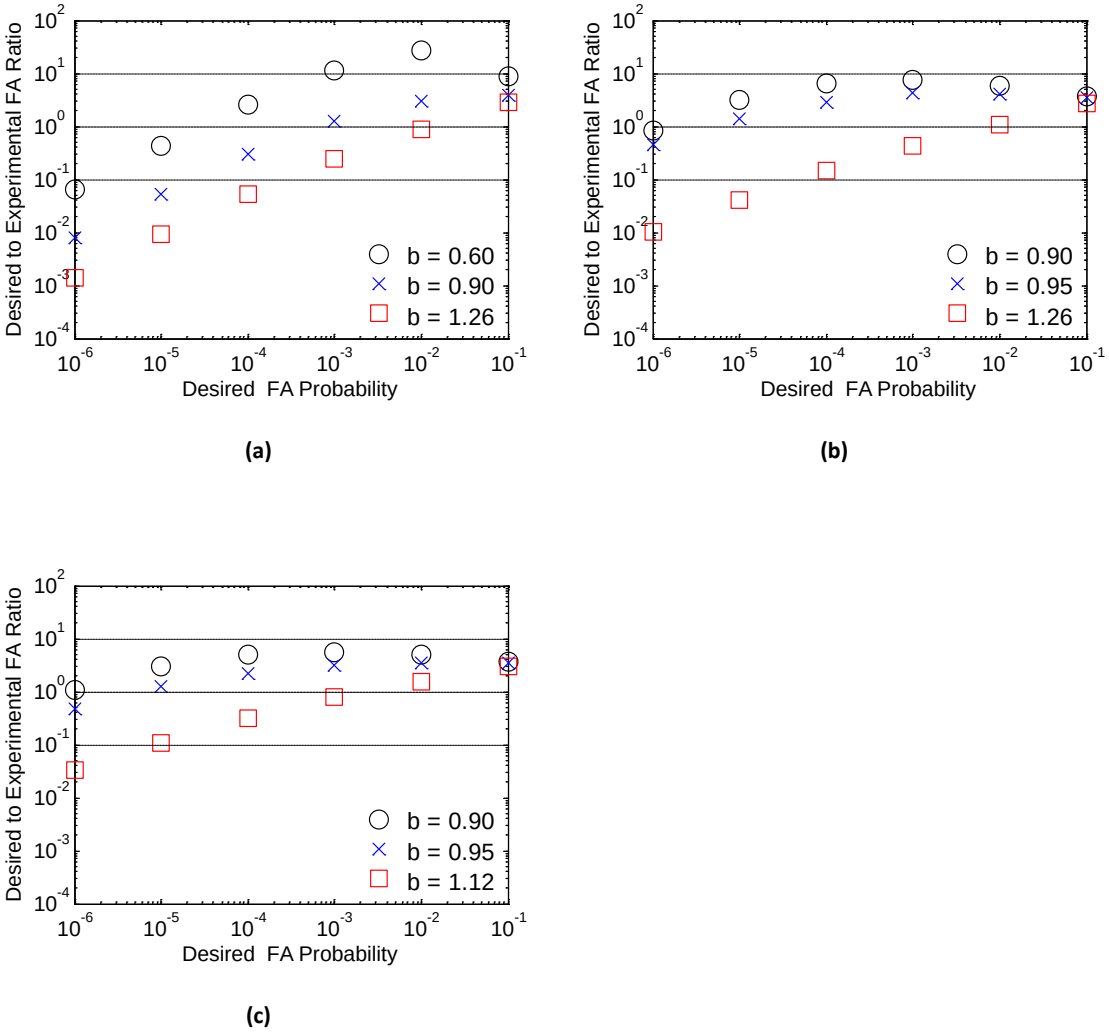

Figure 4.20 CFAR Performance using Weibull distribution for neighborhood size of 7x7, using partial whitening value of 0.85 for 3 different shape parameters. (a) linear array with 1500 Hz cut-off frequency (b) perimeter array with 300 Hz cut-off frequency (c) planar array with 300 Hz cut-off frequency.

4.4 Low frequency limit

According to the equation (2.18), it was demonstrated that in order to provide a symmetric condition for the noise-only distribution, it is required to increase sufficiently the standard deviation of the inter-path distances relative to the source wavelength. Earlier in this chapter it was shown that the minimum inter-path distances standard deviation is equal to 0.21, 0.38 and 0.67 for the linear, perimeter and planar arrays respectively. Applying a high-pass filter to the microphone signals is an effective approach to eliminate the low frequency components which the specific microphone geometry cannot decorrelate. The above experimental results show that the high-pass filter cut-off frequency for the linear, perimeter and planar arrays should be equal to 1500 Hz, 300 Hz and 300 Hz respectively to achieve a good CFAR performance. Based on equation (2.18) and the experimental CFAR performances, the third null of the *sinc* function in equation (2.18) can be used to determine the low frequency limit for all microphone geometries, if only the *mic-distribution* factor is considered:

$$f_L = \frac{3c}{\sigma_\Delta \sqrt{12}} \tag{4.26}$$

where σ_Δ is the smallest standard deviation of inter-path distance over the FOV and c is the speed of sound.

4.5 Adaptive high-pass filters

In our conventional CFAR processor, based on the minimum inter-path distance standard deviation, the microphone signals are first high-pass filtered using equation (4.26). The high-passed filtered signals are then whitened. After whitening signals, the whitened signals are time aligned. The time alignment of signals is based on the propagation delays which depend on the inter-path distance of each point in the FOV to microphone pairs. Instead of applying a fixed high-pass filter to the microphone signals based on the minimum inter-path distance standard deviation, a more sophisticated approach is to first whiten the microphone signals and then time align them. Right after time aligning the signals and before creating the coherent powers, adaptive high-pass filters are applied to the time aligned signals based on the inter-path distance of each point in the FOV to the microphone pairs. In fact, for each point in the FOV an adaptive high-pass filter is designed based on the inter-path distance of that point to the microphone pairs.

In the situation where for each point in the FOV an adaptive high-pass filter is exploited, the equation (4.26) is modified in order to demonstrate that the cut-off frequencies of adaptive high-pass filters are a function of the position of the focal point.

$$f_L(x,y) = \frac{3c}{\sqrt{12}\sigma_\Delta(x,y)} \tag{4.27}$$

where $\sigma_\Delta(x,y)$ denotes the standard deviation of inter-path distances of the point inside the FOV plane, located at position (*x,y*) to the microphone pairs.

Fig 4.21 shows the CFAR performance using adaptive high-pass filters with cut-off frequencies computed by equation (4.27) for the linear geometry in the situation where the *mean-cx2* approach and Weibull distribution are exploited to model the coherent power. The PHAT-β parameter is set to 0.85.

As can be seen by Fig 4.21, a reasonable CFAR performance cannot be achieved for the linear array using either the *mean-cx2* approach or Weibull distribution if an adaptive high-pass filter is applied to each point of FOV.

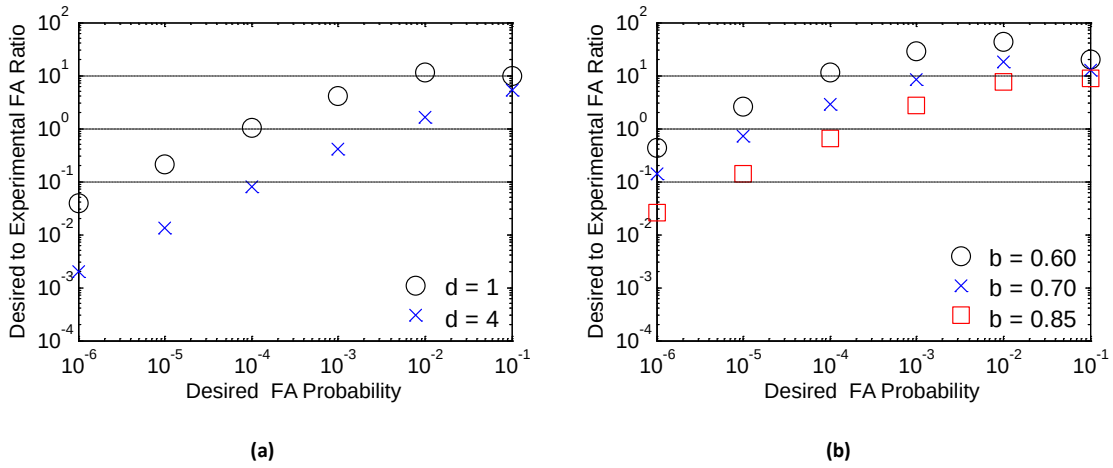

Figure 4.21 CFAR performances using adaptive high-pass filters for each FOV point for linear array with beta value equal to 0.85 (a) using *mean-cx2* method and variations in degree of freedom (b) using Weibull distribution and variations in shape parameter.

On the other hand, a reasonable CFAR performance can be achieved in the perimeter and planar geometries using adaptive high-pass filters. In the case of using the *mean-cx2* approach, the CFAR performance is degraded in both the perimeter and planar arrays using adaptive high-pass filters relative to using only one high-pass filter for all FOV points. Furthermore, if Weibull distribution is used to model the coherent power, the shape parameter needs to be decreased to achieve a good CFAR performance for both the perimeter and planar geometries in the situation where adaptive high-pass filters are exploited relative to the situation where microphone signals were high-passed filtered by the same filter for all FOV points.

Fig 4.22 demonstrates the CFAR performance for the perimeter and planar geometries using adaptive high-pass filters in the cases that the *mean-cx2* approach and Weibull distribution are exploited to model the coherent power. The PHAT-β parameter is set to 0.85.

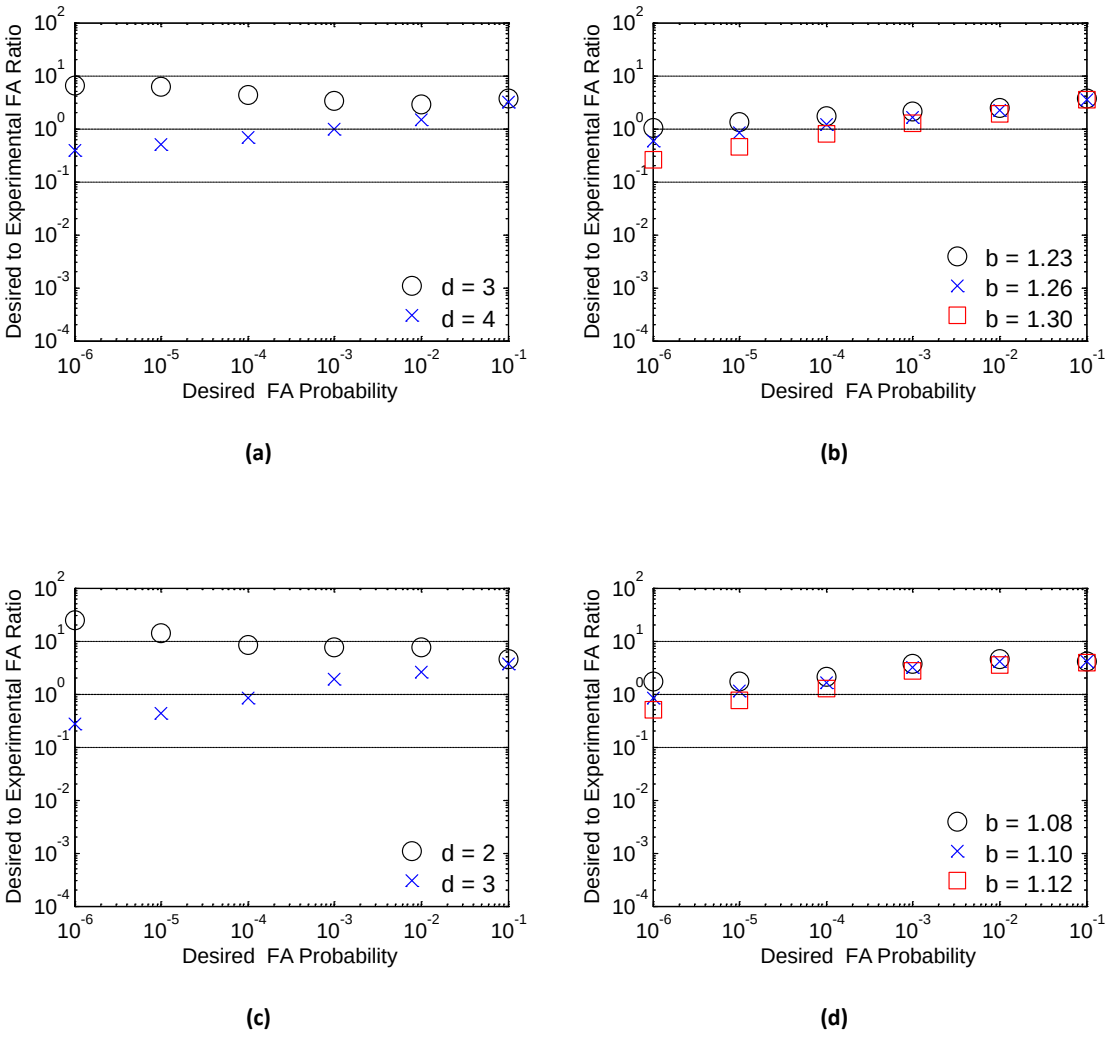

Figure 4.22 CFAR performance using adaptive high-pass filters for each FOV point for beta value equal to 0.85. (a) and (b) for perimeter array. (c) and (d) for planar array. (a) and (c) using *mean-cx2* method and variations in degree of freedom (b) and (d)using Weibull distribution and variations in shape parameter.

4.6 Combining mic-distribution and noise-path factors

In chapter 2, based on equation (2.10), it was declared that *mic-distribution* and *noise-path* factors are two sources of incoherency or decorrelation. The mic-distribution factor depends on the microphone geometry and since it scales all noise components, it provides a convenient point to design the high-pass filters. Consequently, based on only the mic-distribution factor, high-pass filters were designed, including a simple one high-pass filter for all FOV points as well as adaptive high-pass filters for each FOV point, and the resultant CFAR performances were investigated in the previous sections.

In this section, in addition to the mic-distribution factor, *noise-path* factor will be exploited to determine the low frequency limit. The *noise-path* factor depends on the inter-path distances between the position of noise sources and microphone pairs. Therefore, the positions of noise sources need to be located. As it was mentioned earlier, there were two noise sources outside the FOV plane. One noise source is located in the left side (endfire) of the linear array (left relative to a person facing the linear array from the front), and the other is located directly in front of the linear array (broad side).

The perimeter array is exploited to locate noise source positions. The position of a noise source is located by two approaches. First of all, the steered response coherent power images are created for the pixels of the FOV plane which most likely contains the noise source position. A local neighborhood may contain only positive SRCP value or may consist of both positive and negative values. If a local neighborhood contains both positive and negative coherent power values, then the position of the test pixel is recorded in the situation where it is the local maximum value and where it exceeds the threshold computed by the inverse value of the Weibull distribution cdf. In addition, the positions of the test pixels which have a maximum value in the neighborhoods that contain all positive coherent power values are recorded. In fact, the location of the noise source is among these recorded positions. These recorded positions have the potential to represent the position of a noise source. In the first approach, the noise

source position is computed by finding the mean value of the recorded positions during all times and frames. In the second approach, the related position to the maximum value of the coherent power values corresponding to the recorded positions is reported as the position of the noise source.

The position of the broad side noise source is [-2.94, 1.3, 1.07]' meters using the first approach and by using the second approach, [-2.98, 1.34, 1.07]' meters is the broad side noise source position. Also, the position of the endfire noise source is [0.54 -2.72 1.45]' meters using the first approach to determine the position of the noise source (the average approach) and if the second approach (maximum approach) is exploited, the position of the endfire noise source will be [0.46 -2.64 1.45]' meters. Fig 4.23 shows the SRCP images at the time and frame number corresponding to the above approaches to locate the position of noise sources.

After determining the noise source positions, the inter-path distances between noise source position and microphone pairs are computed. The standard deviation of the noise source inter-path distances is being used in order to determine the low frequency limit resulting from the noise-path factor. Table 4.1 lists the low frequency limit resulting from the broad side and endfire noise sources for the three microphone geometries in the situation where the n^{th} null of the *sinc* function of equation (2.18) is considered as the low frequency limit due to the noise-path factor for the two approaches to locate the position of noise sources. The maximum of low frequency limit resulting from the two approaches as well as the two noise sources, broad side and endfire, should be selected as the low frequency limit due to the noise-path factor.

Clearly, the standard deviation of inter-path distances between the broad side noise source and microphone pairs is much smaller than inter-path distance standard deviation between the endfire noise source and microphone pairs for the linear geometry. Therefore, the low frequency limit resulting from the broad side noise source is greater than the low frequency limit resulting from the endfire noise source in the linear geometry. In consequence, the low frequency limit resulting from the broad side noise source is considered as the noise-path factor for the linear array.

On the other hand, for the perimeter and planar geometries the low frequency limit resulting from the broad side and endfire noise sources are close together. Therefore, it is required to compare the low frequency limits resulting from the broad side and endfire noise sources and then select the largest one as the noise-path factor.

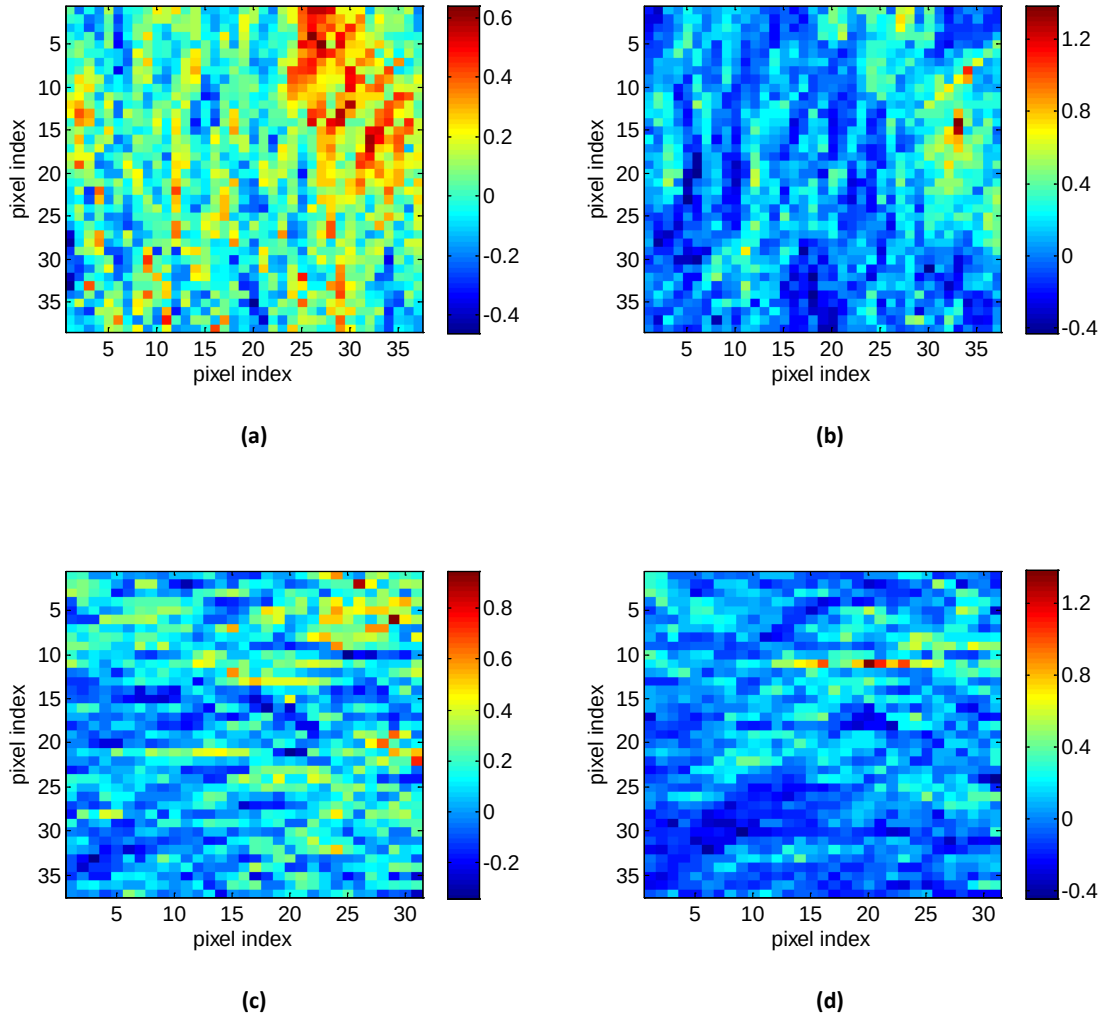

Figure 4.23 SRCP images. (a) and (c) locating noise source by finding mean value of position of all potential noise source positions. (b) and (d)locating noise source by finding the position which have maximum value of SRCP among all potential pixels. (a) and (b) for broad side noise. (c) and (d) for endfire noise.

Table 4.1: Low frequency limit resulted by noise-path factor

noise source position	Geometry	n^{th} null	Frequency (Hz) Avg. approach	Max approach
broad side	Linear	1	241	236
		2	481	473
		3	722	709
	Perimeter	1	52	51
		2	102	102
		3	154	153
	Planar	1	69	68
		2	136	136
		3	204	204
endfire	Linear	1	73	74
		2	146	149
		3	219	223
	Perimeter	1	53	53
		2	106	107
		3	159	160
	Planar	1	69	69
		2	138	139
		3	206	208

Both mic-distribution and noise-path factors can be analyzed to determine the best cut-off frequencies for the adaptive high-pass filters. For each point in the FOV, an adaptive high-pass filter is applied with the cut-off frequency equal to the maximum of lower frequency limits resulting from the mic-distribution and noise-path factors.

The goal is to achieve a good CFAR performance with the lowest possible frequency limit. Therefore, in this section, different combinations of the n^{th} nulls of the sinc functions related to the mic-distribution and noise-path factors are examined to achieve reasonable CFAR threshold performances with the lowest possible frequency limit and the effect of different null combinations are investigated.

In the previous sections, it was illustrated that applying adaptive high-pass filters to each FOV point resulted in a poor CFAR performance for the linear array using either the

mean-cx2 or Weibull distribution approaches in the situation where only the mic-distribution factor was considered and the 3rd nulls of the sinc function related to this factor was selected as the high-pass filter cut-off frequencies. Therefore, the combination of the 3rd nulls for the mic-distribution factor and zero as the low frequency limit resulting from the noise-path factor will lead to a poor CFAR performance for the linear array. On the other hand, it was also shown that applying a fixed high-pass filter with a cut-off frequency of 800 Hz (approximately equal to the 3rd null of the sinc function related to the noise-path factor) for all FOV points did not result in a good CFAR performance for the linear array; especially if Weibull distribution was used to model the coherent power. Therefore, the combination of zero as the low frequency limit resulting from the mic-distribution factor and the 3rd null for the noise-path factor results in a poor CFAR performance for the linear geometry.

Experimental results show that for the partial whitening case, a reasonable CFAR performance is achieved for the linear array if the 3rd nulls of the related sinc functions are selected as low frequency limit resulting from the mic-distribution and noise-path factors. In this situation, the signals are high-pass filtered with a cut-off frequency of at

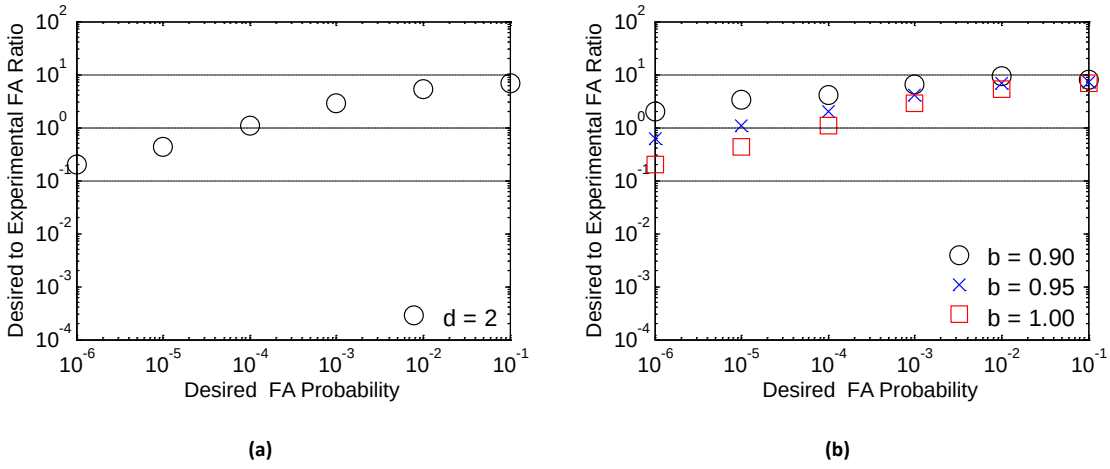

Figure 4.24 CFAR performance using 3rd null for both mic-distribution and noise-path factors to design adaptive high-pass filters for each FOV point for linear array with beta value equal to 0.85 (a) using *mean-cx2* method with degree of freedom equal to 2 (b) using Weibull distribution and variations in shape parameter.

least 722 Hz. Figure 4.24 illustrates the CFAR performance for the combination of the 3rd null for both mic-distribution and noise-path factors.

It is worth noting that a better result is achieved by modeling the coherent power with Weibull distribution because of its powerful capability in adjusting the tail of the distribution by exploiting different shape parameters. Furthermore, by exploiting adaptive high-pass filters such that the cut-off frequencies are determined by finding the maximum value of the 3rd null of sinc functions related to the mic-distribution and noise-path factors, the required frequency limit to achieve a good CFAR performance is decreased from 1500 Hz to at least 722 Hz in comparison with the situation where only a fixed high-pass filter with a cut-off frequency of 1500 Hz is applied to all FOV points for the linear array.

By comparing CFAR performances resulting from selecting different null combinations for the mic-distribution and noise-path factors frequency limits, it can be concluded that in the situation where partial whitening is exploited, the noise-path factor is dominant over the mic-distribution factor in terms of the impact amount of these factors on the performance of the CFAR processor, whereas the mic-distribution factor is dominant over the noise-path factor in the case that no whitening is applied. Figures 4.25 and 4.26 compare the CFAR performances for the linear array for different null combinations for the low frequency limits resulting from the mic-distribution and noise-path factors using the *mean-cx2* and Weibull approaches to model the coherent power.

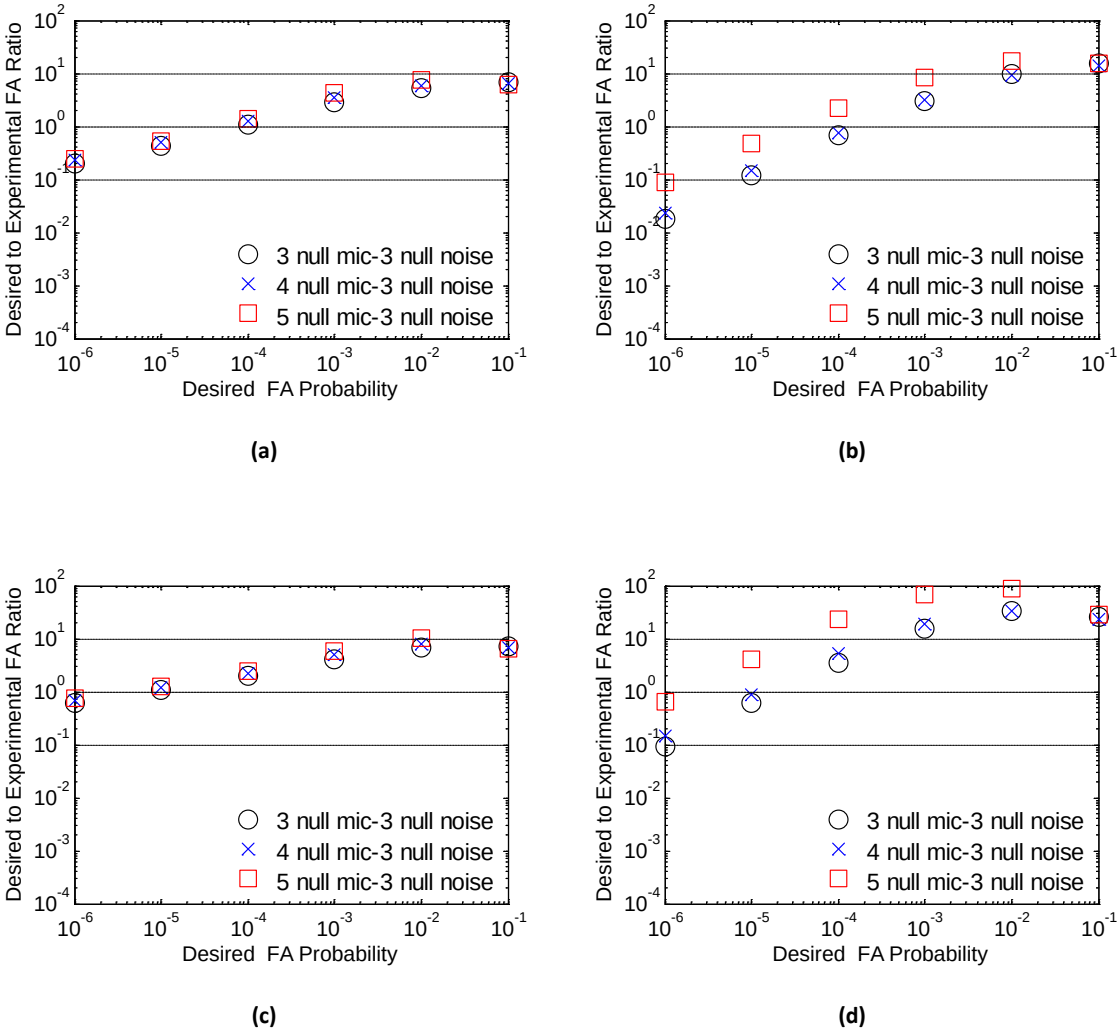

Figure 4.25 Mic-distribution factor is dominant for no whitening case. CFAR performance for linear array: (a) and (b) using mean-cx2 approach with degree of freedom equal to 2. (c) and (d) exploiting Weibull distribution with shape parameter of 0.95. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation.

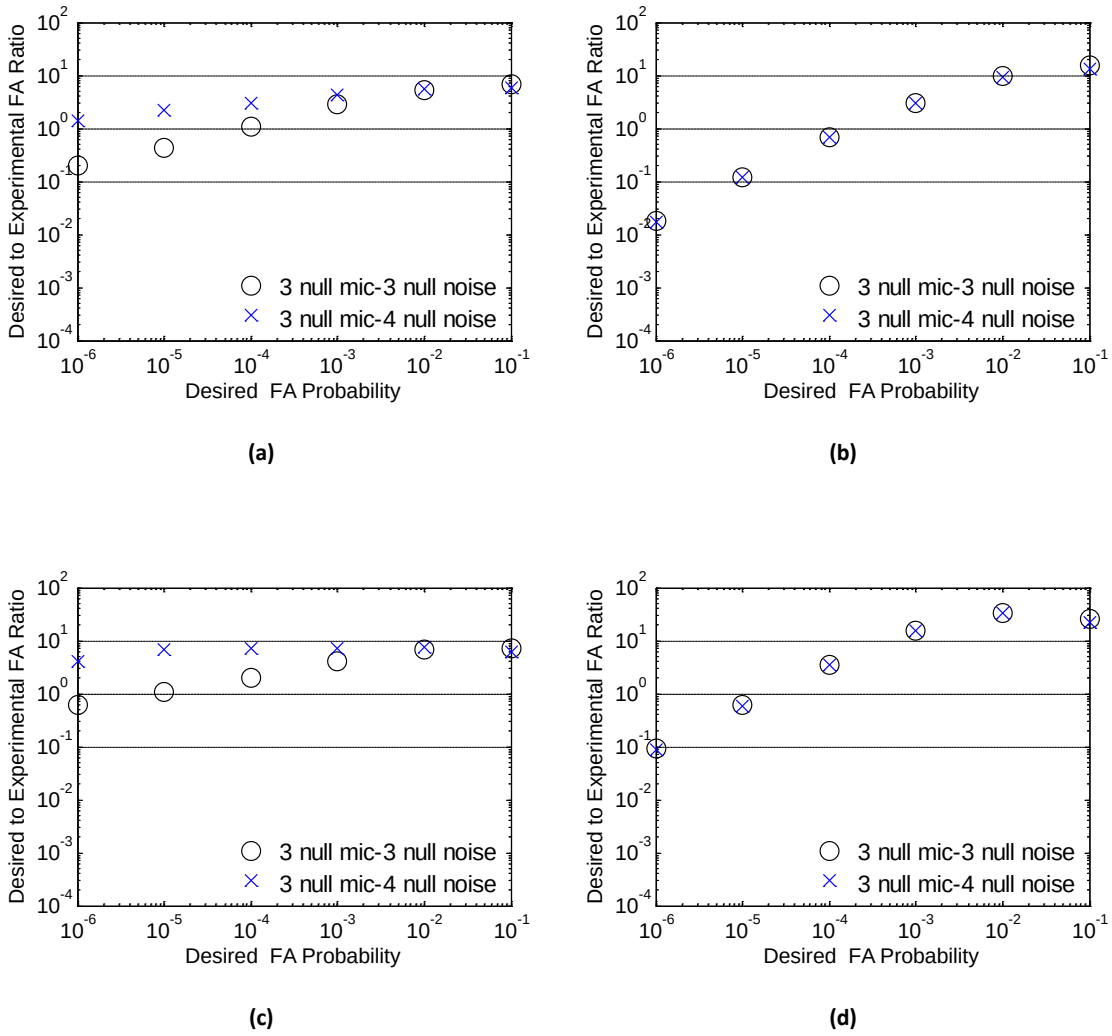

Figure 4.26 Noise-path factor is dominant for partial whitening case. CFAR performance for linear array: (a) and (b) using mean-cx2 approach with degree of freedom equal to 2. (c) and (d) exploiting Weibull distribution with shape parameter of 0.95. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation.

For the perimeter array which has the largest minimum standard deviation of inter-path distances between the focal points and microphone pairs, the smallest possible combinations is selected for the low frequency limits resulting from the mic-distribution and noise-path factors. Therefore, for both mic-distribution and noise-path factors, the 1st nulls of their related sinc function are selected to design the adaptive high-pass filters

cut-off frequencies. From Table 4.1, 53 Hz is the low frequency limit resulting from the noise-path factor for the perimeter geometry if the 1st null of the related sinc function is considered to determine the low frequency limit by the noise-path factor.

Experimental results show that for the partial whitening situation, using either the *mean-cx2* approach or Weibull distribution to model the coherent power results in a good CFAR performance for the perimeter array in the case that even very low frequencies, such as the maximum of the 1st null of the sinc function related to the mic-distribution and 53 Hz, are applied to design the adaptive high-pass filter cut-off frequencies to each FOV points. Consequently, for the partial whitening situation, adaptive high-pass filters for the perimeter array results in a good CFAR performance if the signals for each FOV point are filtered with at least 53 Hz, whereas by applying a fixed high-pass filter for all FOV points the cut-off frequency was required to be set to 300 Hz to achieve a good CFAR performance. Fig 4.27 illustrates the CFAR performance for the perimeter array in the situation where the 1st nulls for both mic-distribution and noise-path factors are selected to design the adaptive high-pass filters and partial whitening is exploited.

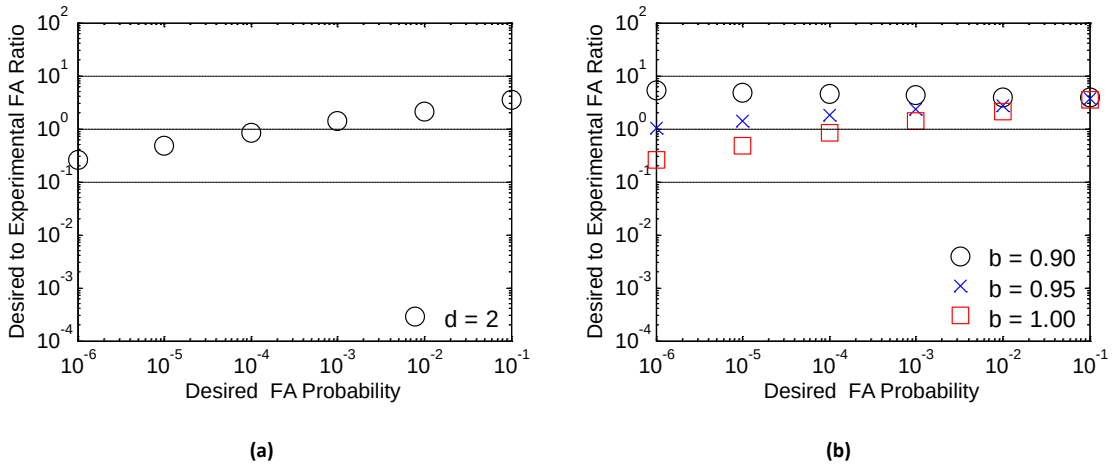

Figure 4.27 CFAR performance using 1st null for both mic-distribution and noise-path factors to design adaptive high-pass filters for each FOV point for perimeter array with beta value equal to 0.85 (a) using *mean-cx2* method with degree of freedom equal to 2 (b) using Weibull distribution and variations in shape parameter.

In the situation where no whitening is performed, likewise in the linear array which is required to apply a high value for the cut-off frequency to achieve a reasonable CFAR performance (even the combination of the 6th null for the mic-distribution factor and the 3rd null for the noise-path factor did not result in a reasonable CFAR performance), a reasonable CFAR performance is achieved for the perimeter array for the no whitening case if only the 1st null of the sinc function related to the mic-distribution factor is combined with the 2nd null of the sinc function related to the noise-path factor (107 Hz) and exploiting Weibull distribution to model the coherent power. Although the CFAR performance is within 1 order of magnitude using the *mean-cx2* method with the degree of freedom equal to 1, the results are not as good as the performances achieved by Weibull distribution. Fig 4.28 shows the results of combining the 1st and the 2nd nulls of the related sinc functions for the mic-distribution and noise-path factors low frequency limits respectively for the perimeter geometry in the situation where no whitening is applied.

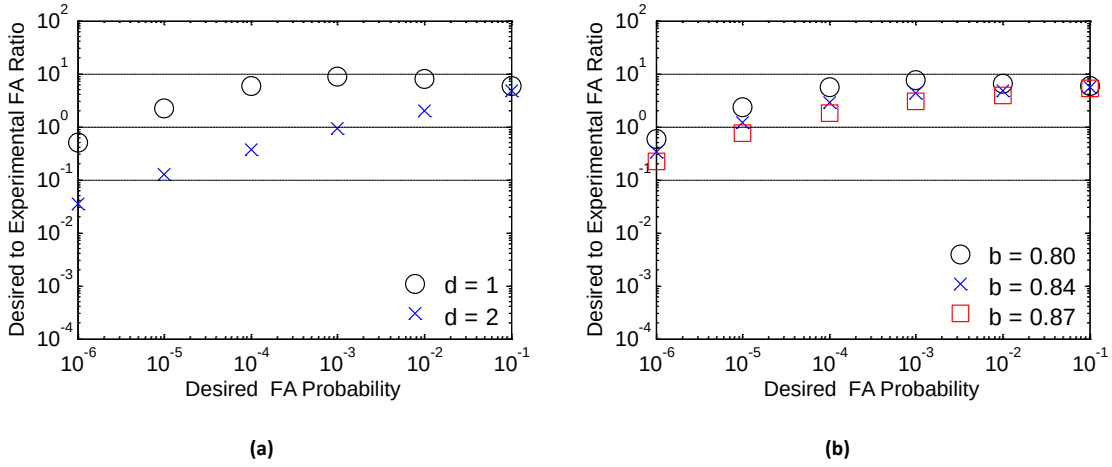

Figure 4.28 CFAR performance using the 1st null for mic-distribution factor and the 2nd null for noise-path factor to design adaptive high-pass filters for each FOV point for perimeter array for no-whitening case (a) using *mean-cx2* method variations in degree of freedom. (b) using Weibull distribution and variations in shape parameter.

If better CFAR performances are required, a higher combination for the null of the sinc functions related to the mic-distribution and noise-path factors can be selected to design the cut-off frequencies of adaptive high pass filters, e.g. the 3rd null of the sinc functions for both mic-distribution and noise-path factors. It should be noted that for the perimeter geometry, the minimum and maximum of the standard deviation of inter-path distance of FOV points to microphone pairs are 0.38 and 1.88 respectively. Therefore, if the 3rd null of the sinc function is chosen for the mic-distribution factor, then the low frequency limit resulting from the mic-distribution factor is between 160 Hz and 789 Hz. On the other hand, if the 3rd null of the sinc function is chosen for the noise-path factor, the low frequency limit resulting from the noise-path factor is 160 Hz (see Table 4.1). Consequently, the combination of the 3rd null for both mic-distribution and noise-path factors is equivalent to ignore the noise-path factor and apply adaptive high-pass filters to each point of the FOV with cut-off frequencies corresponding to only the mic-distribution factor. Fig 4.29 shows the CFAR performance when for both low frequency limits resulting from the mic-distribution and noise-path factors, the 3rd null of the related sinc functions are selected for the perimeter array.

Unlike the linear array, it cannot be concluded whether the mic-distribution is dominant over noise-path factor or vice versa for the perimeter and planar geometries under the condition that either partial whitening or no whitening is exploited.

The same analysis is performed for the planar geometry. In order to determine the smallest possible null combination of the sinc functions related to the mic-distribution and noise-path factors to achieve a reasonable CFAR performance, the 1st null of related sinc functions are selected for both mic-distribution and noise path factors as the low frequency limits caused by these factors. The adaptive high-pass filter cut-off frequencies are the maximum of the low frequency limits. If the 1st null of the sinc function related to the noise-path factor is considered as the low frequency limit resulting from the noise-path factor, then from Table 4.1, 69 Hz is the low frequency limit resulting from the noise-path factor.

Experimental results show that in the case where the 1st null of related sinc functions are considered as the low frequency limits resulting from the mic-distribution and noise-path factors, a reasonable CFAR performance for the planar array can be achieved by applying the adaptive high-pass filters to each FOV point, only if whitening is exploited.

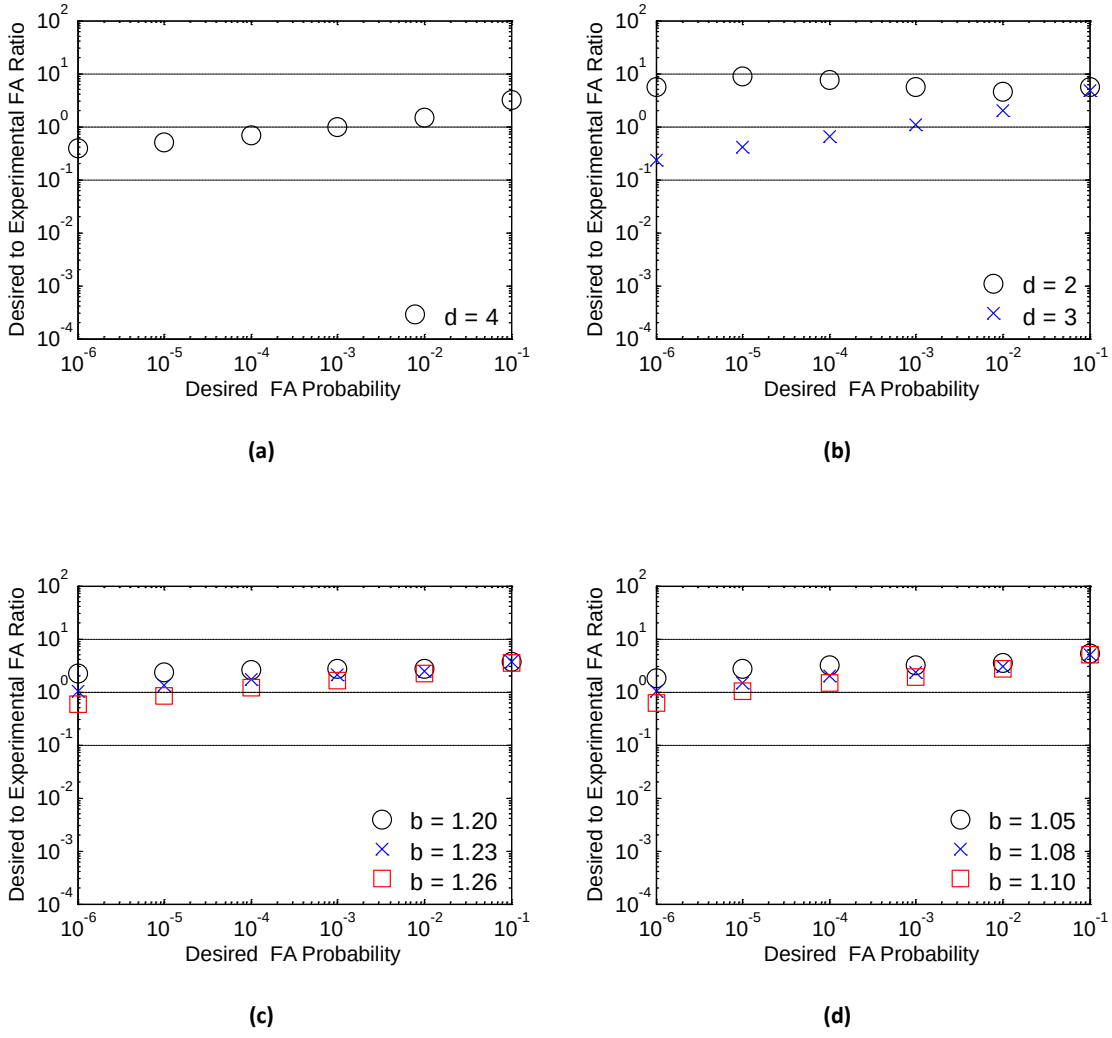

Figure 4.29 CFAR performance for perimeter array when the 3rd null of the related sinc functions are selected for both low frequency limits resulted by mic-distribution and noise-path factors: (a) and (b) using mean-cx2 approach, variation in degree of freedom. (c) and (d) exploiting Weibull distribution, variation in shape parameter. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation.

For the no whitening case, similar to the perimeter array, for the planar geometry, applying adaptive high-pass filters to each FOV point with cut-off frequencies of the maximum of the 1st nulls of the sinc functions related to the mic-distribution and noise-path factors could not compensate for the degradation in the CFAR performance caused by the low frequency components and as a result, the ratio of desired to experimental false alarm probability cannot be limited within 1 order of magnitude. Therefore, larger cut off frequencies should be selected for the no whitening situation. Fig 4.30 represents the CFAR performance for the planar geometry when considering the 1st nulls of related sinc functions as the low frequency limits for both mic-distribution and noise-path factors using the *mean-cx2* approach and Weibull distribution to model the coherent power. The PHAT-β parameter is set to 0.85.

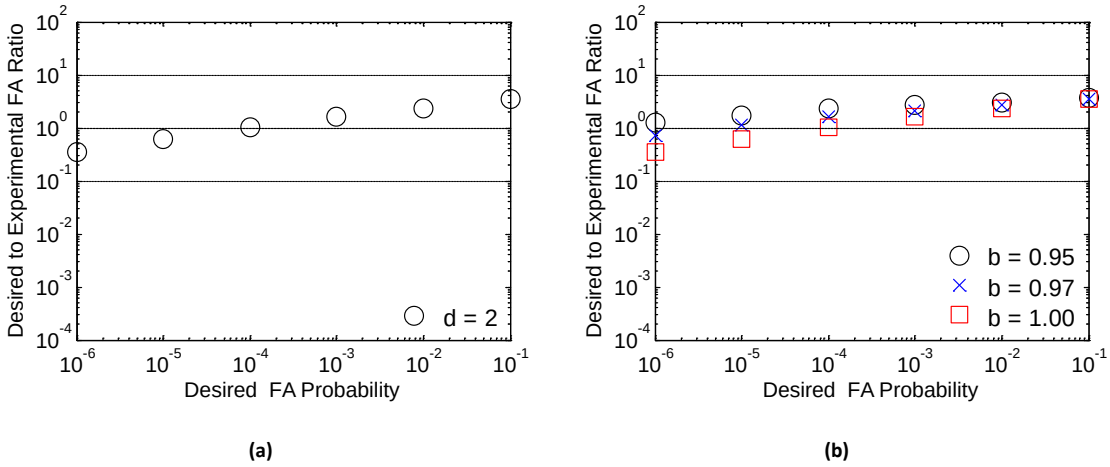

Figure 4.30 CFAR performance by using the 1st null for both mic-distribution and noise-path factors to design adaptive high-pass filters for each FOV point for planar array with beta value equal to 0.85 (a) using *mean-cx2* method using degree of freedom of 2. (b) using Weibull distribution and variations in shape parameter.

In the next step, the low frequency limit resulting from the noise-path factor is increased to the 2nd null of the related sinc function which results in 139 Hz (see Table 4.1) for the noise-path factor low frequency limit, while the low frequency limit resulting

from the mic-distribution factor is kept unchanged (the 1st null of the related sinc function). Since a reasonable CFAR performance is already achieved for the partial whitening case, the main goal of this experience is to focus on the CFAR performance when no whitening is performed on the received signals by the microphones. The CFAR performances of this null combination (the 1st and the 2nd nulls for the mic-distribution and noise-path factors respectively) to design adaptive high-pass filters for the no whitening situation are shown in Fig 4.31.

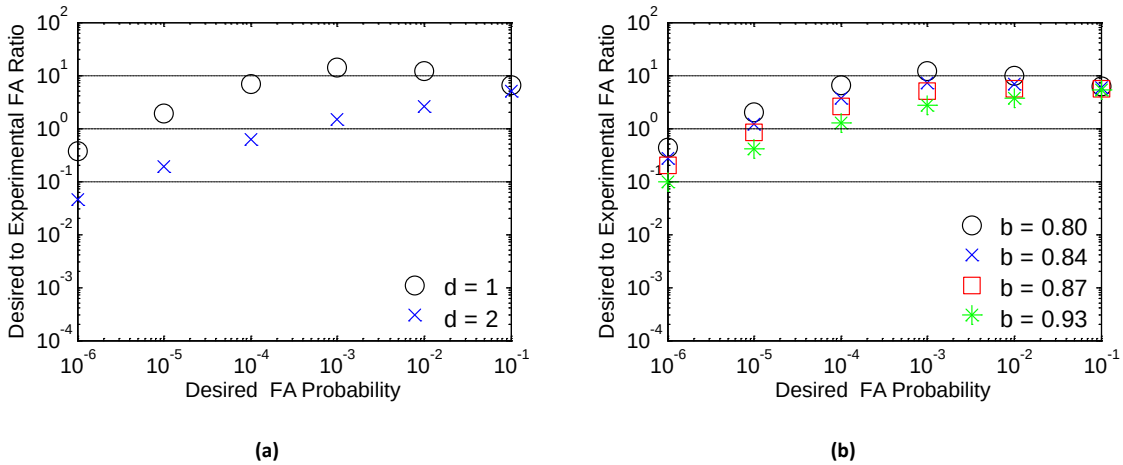

Figure 4.31 CFAR performance using the 1st null for mic-distribution factor and the 2nd null for noise-path factor (139 Hz) to design adaptive high-pass filters for each FOV point for planar array for no-whitening case (a) using *mean-cx2* method variations in degree of freedom. (b) using Weibull distribution and variations in shape parameter.

Increasing high-pass filters cut-off frequency resulted in not only better CFAR performance but also less skewness in the data. As can be seen from Fig 4.31, if the *mean-cx2* method is applied, a reasonable CFAR performance cannot be achieved for the planar array in the situation where for each FOV point, the maximum of the 1st null of the sinc function related to the mic-distribution factor and the 2nd null of the sinc function related to the noise-path factor (139 Hz) is selected as the cut-off frequency of the adaptive high-pass filter. Although for this null combination, modeling the coherent power by Weibull distribution and adjusting the shape parameter between 0.84 and

0.93 results in a reasonable CFAR performance (within 1 order of magnitude) for the no whitening case, the results of this null combination are not sufficiently satisfying for the no whitening situation. Hence, a higher null combination is required to achieve a good CFAR performance for the no whitening case.

In order to achieve a really good CFAR performance for the planar geometry, particularly for the no whitening case as for the partial whitening even a lower null combination results in a good CFAR performance, the 2nd null of the related sinc functions for both mic-distribution and noise-path factors are selected to design the adaptive high-pass filters cut-off frequencies. It should be noted that the minimum and maximum of standard deviation of inter-path distances of FOV points to microphone pairs are respectively 0.38 and 1.88 for the planar geometry. Therefore, if the 2nd null of the related sinc function is desired as the low frequency limit resulting from the mic-distribution factor, then the low frequency limit resulting from the mic-distribution factor is between 135 Hz and 298 Hz. On the other hand, 139 Hz is the low frequency limit resulting from the noise-path factor if the 2nd null of the related sinc function is considered as the low frequency limit resulting from the noise-path factor. In consequence, the CFAR performance when the 2nd null of the related sinc functions for both mic-distribution and noise-path factors are selected is almost identical to the CFAR performance when the noise-path factor is ignored and only the 2nd null of the sinc function related to the mic-distribution is considered in order to design the adaptive high-pass filters cut-off frequencies. Fig 4.32 illustrates the CFAR performance when the 2nd null of the related sinc functions for both mic-distribution and noise-path factors are selected to design the adaptive high-pass filters for the partial and no whitening situations when the *mean-cx2* approach and Weibull distribution are used to model the coherent power.

It is interesting to note that for the planar array a better CFAR performance is achieved for the no whitening case when the *mean-cx2* method is exploited relevant to the CFAR performance for the partial whitening if the 2nd null of the related sinc functions for both mic-distribution and noise-path factors are selected to design the

adaptive high-pass filters. Furthermore, for the planar geometry, if a sufficient high value is chosen for the null combination, the CFAR performances as well as the shape parameter in the no whitening case and partial whitening case are similar. In other words, if a high null combination is selected to design the adaptive high-pass filter cut-off frequencies, whitening does not have a significant effect on the CFAR performance for the planar array. Therefore, if no whitening is desired, one can use the planar geometry to achieve a reasonable CFAR performance.

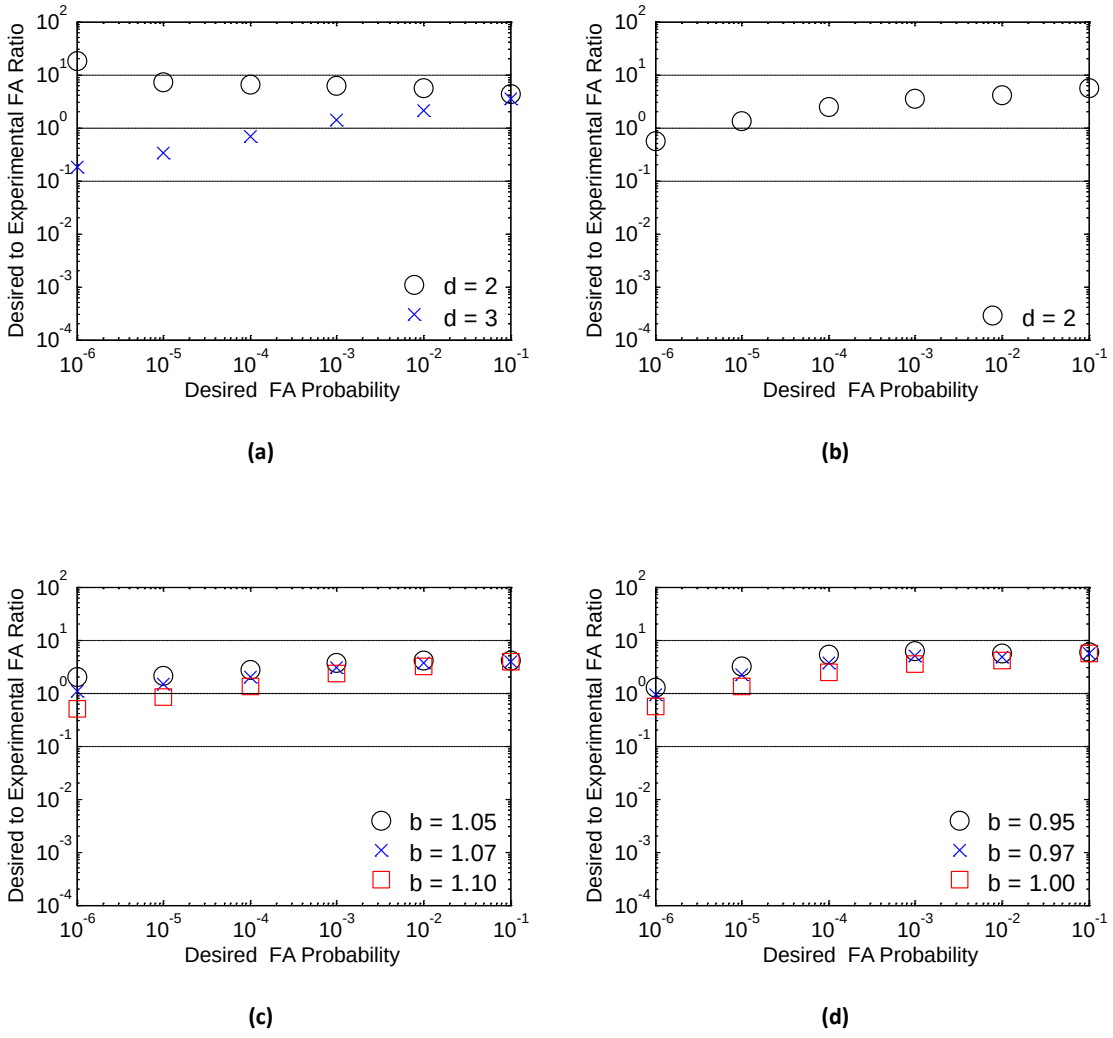

Figure 4.32 CFAR performance for planar array when the 2nd null of the related sinc functions are selected for both low frequency limits resulted by mic-distribution and noise-path factors: (a) and (b) using mean-cx2 approach, variation in degree of freedom. (c) and (d) using Weibull distribution, variation in shape parameter. (a) and (c) applying partial whiting with beta value equal to 0.85. (b) and (d) no whitening situation.

Chapter 5

Conclusion

5.1 Conclusion

This thesis introduced a new CFAR processor which uses only negative coherent power values to estimate the statistics of positive coherent power pixels. In this approach, it is critical that the noise-only distribution is symmetrical in order to achieve a good CFAR performance. It was demonstrated that it is required to remove the low frequency components of the source relative to inter-path distances. Therefore, the perimeter array, which had the largest differential path lengths, outperformed the other geometries. To remove the low frequency components two different high-pass filtering approaches were applied. In the first approach, a fixed high-pass filter was applied to the microphone signals for all FOV points. In this method, only the mic-distribution factor was considered and the smallest standard deviation of inter-path distances between focal points and microphone pairs was exploited to determine the high-pass filter cut-off frequency. In the second approach, adaptive high-pass filters were applied to each FOV point based on the standard deviation of inter-path distances between that point and microphone pairs. Using the second approach to remove the low frequency components, experimental results showed that it is required to combine the mic-distribution and noise-path factors to achieve a good CFAR performance for all microphone geometries. Therefore, the position of noise sources should be located to design the cut-off frequencies of the adaptive high-pass filters. Experimental results showed that the low frequency limit to achieve a reasonable CFAR performance is reduced by combining the mic-distribution and noise-path factors.

To model the coherent power, the Chi-square and Weibull distributions were tested. Experimental results showed that using the *mean-cx2* approach as well as the Weibull distribution results in good CFAR performances. If only a fixed high-pass filter for all FOV points is used to remove the low frequency components there were no significant differences between the *mean-cx2* and Weibull distribution CFAR results despite the fact that the computational cost in the *mean-cx2* approach is less than Weibull distribution. On the other hand, Weibull distribution outperforms because of its powerful ability in adjusting the tail of the distribution if adaptive high-pass filters for each point of FOV were applied.

Bibliography

[1] P. L. Chu. Desktop mic array for teleconferencing. In Proceedings of ICASSP95, IEEE, 1995.

[2] J. L. Flanagan, D. A. Berkley, G.W. Elko, J. E. West, and M. M. Shondhi. Autodirective microphone systems. Acoustica, vol. 73, pp. 58-71, 1991.

[3] W. Kellerman. A self-steering digital microphone array. In Proceedings of ICASSP91, pages 3581-3584, IEEE, May 1991.

[4] F. Khali, J. P. Jullien and A. Gilloire. Microphone array for sound pickup in teleconferencing systems. J. Audio Eng. Soc., vol. 42, no. 9, September 1994.

[5] H. Wang and P. Chu. Voice source localization for automatic camera pointing system in videoconferencing. In Proceedings of ICASSP, volume 1, pages 187-190. IEEE, 1997.

[6] J. E. Adcock, Y. Gotoh, D. J. Mashao and H. F. Silverman. Microphone-array speech recognition via incremental MAP training. In Proceeding of ICASSP96, Atlanta, GA, May 1996.

[7] C. Che, Q. Lin, J. Pearson, B. deVries and J. Flanagan. Microphone arrays and neural networks for robust speech recognition. In Proceedings of the Human Language Technology Workshop, Pages 342-347, Plainsboro, NJ, March 8-11, 1994.

[8] C. Che, M. Rahim and J. Flanagan. Robust speech recognition in a multimedia teleconferencing environment. J. Acoust. Soc. Am., 92(4):2476, 1992.

[9] D. Giuliani, M. Omologo and P. Svaizer. Talker localization and speech recognition using a microphone array and a cross-power spectrum phase analysis. In Proceedings of ICSLP, volume 3, pages 1243-1246, September 1994.

[10] T. B. Hughes, H. Kim, J. H. DiBiase, and H. F. Silverman. Performance of an HMM speech recognizer using a real-time tracking microphone array as input. IEEE Trans. Speech Audio Proc., 7(3):346-349, May 1999.

[11] T. B. Hughes, H. Kim, J. H. DiBiase, and H. F. Silverman. Using a real-time, tracking microphone array as input to an HMM speech recognizer. In Proceedings of ICASSP98, IEEE, 1998.

[12] H. F. Silverman. Some analysis of microphone arrays for speech data acquisition. Trans. on Acoustics, Speech, and Signal Processing, 35(12):1699-1711, IEEE, December 1987.

[13] S. M. Yoon, S. C. Kee, Speaker detection and tracking at mobile robot platform, Proc. 2004 Intl. Symp. on Intelligent Signal Processing and Communication Systems, 2004, pp. 596–600.

[14] Huang, T.S., “Multimedia/multimodal signal processing, analysis, and understanding,” First International Symposium on Control, Communications and Signal Processing. 2004, p. 1.

[15] J. H. DiBiase, H. F. Silverman, M. S. Brandstein, “Robust Localization in Reverberant Rooms, ” Microphone Arrays, Signal Processing Techniques and Applications, Springer Verlag, Berlin, 2001, pp. 157–180.

[16] C. H. Knapp and G.C. Carter. The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech and Signal Processing, 24(4):320-327, August 1976.

[17] P. Svaizer, M. Mattassoni and M. Omologo. Acoustic source localization in a three-dimensional space using crosspower spectrum phase. In Proceedings of ICASSP97, IEEE, 1997.

[18] M. S. Brandstein, J. E. Adcock and H. F. Silverman. A closed-form method for finding source locations from microphone-array time-delay estimates. In Proceedings of ICASSP95, pages 3019-3022, IEEE, May 1995.

[19] K.D. Donohue, K.S. McReynolds, A. Ramamurthy, “Sound Source Detection Threshold Estimation using Negative Coherent Power,” Proceeding of the IEEE, Southeastcon 2008, pp. 575-580, April 2008.

[20] K.D. Donohue, J. Hannemann, and H.G. Dietz, “Performance for Phase Transform for Detecting Sound Sources in Reverberant and Noisy Environments,” *Signal Processing*, Vol. 87, no. 7, pp. 1677-1691, July 2007.

[21] A. Ramamurthy, H. Unnikrishnan, K.D. Donohue, “Experimental Performance Analysis of Sound Source Detection with SRP PHAT-β, *Proceedings of the IEEE Southeastcon* 2009, pp. 422-427, March 2009.

[22] M. Weiss, "Analysis of some modified cell averaging CFAR processors in multiple-target situations," IEEE Trans. on AES, vol. 18, pp. 102114, Jan. 1982.

[23] H. Rohling, “Radar CFAR thresholding in clutter and multiple target situations,” IEEE Trans. on AES, vol. 19, pp. 608-621, July 1983.

[24] N. Levanon and M. Shor, “Order statistic CFAR for Weibull background,” IEE Proc.-F, Commun. Radar and Sig. Proc., vol. 137, pp. 157-162, June 1990.

[25] S. Kuttikkad, R. Chellappa, Non-Gaussian CFAR techniques for target detection in high resolution SAR images, Image Processing, 1994. Proceedings. ICIP-94., Proc. of IEEE International Conference on Image Processing 1994, Vol. 1, pp. 910-914, Nov 1994.

Vita

Sayed Mahdi Saghaian Nejad Esfahani was born on September 22, 1982 in Esfahan, Iran. He received his B.Sc degree in Electrical Engineering-Communication in 2005 from Isfahan University of Technology, Esfahan, Iran. He has been working at Center for Visualization and Virtual Environments as a Research Assistant under Dr. Kevin D. Donohue supervision since 2009. Since Fall 2009, he has been working on high speed digital imaging for vocal cord dysfunction.