2010

PERFORMANCE ANALYSIS OF SRCP IMAGE BASED SOUND SOURCE DETECTION ALGORITHMS

Praveen Reddy Nalavolu
University of Kentucky, praveennalavolu@gmail.com

ABSTRACT OF THESIS

PERFORMANCE ANALYSIS OF SRCP IMAGE BASED SOUND SOURCE DETECTION ALGORITHMS

Steered Response Power based algorithms are widely used for finding sound source location using microphone array systems. SRCP-PHAT is one such algorithm that has a robust performance under noisy and reverberant conditions. The algorithm creates a likelihood function over the field of view. This thesis employs image processing methods on SRCP-PHAT images, to exploit the difference in power levels and pixel patterns to discriminate between sound source and background pixels. Hough Transform based ellipse detection is used to identify the sound source locations by finding the centers of elliptical edge pixel regions typical of source patterns. Monte Carlo simulations of an eight microphone perimeter array with single and multiple sound sources are used to simulate the test environment and area under receiver operating characteristic (ROCA) curve is used to analyze the algorithm performance. Performance was compared to a simpler algorithm involving Canny edge detection and image averaging and an algorithms based simply on the magnitude of local maxima in the SRCP image. Analysis shows that Canny edge detection based method performed better in the presence of coherent noise sources.

KEYWORDS: Steered Response Power, Sound Source Localization, Hough Transform based Ellipse Detection, Canny Edge Detection, Area under Receiver Operating Characteristic Curve.

Praveen Reddy Nalavolu

12/10/2010

PERFORMANCE ANALYSIS OF SRCP IMAGE BASED SOUND SOURCE DETECTION ALGORITHMS

By

Praveen Reddy Nalavolu

Kevin D. Donohue

Director of Thesis

Stephen D. Gedney

Director of Graduate Studies

12/10/2010

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Unpublished theses submitted for the Master's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with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

Praveen Reddy Nalavolu

The Graduate School

University of Kentucky

2010

PERFORMANCE ANALYSIS OF SRCP IMAGE BASED SOUND SOURCE DETECTION ALGORITHMS

THESIS

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

By

Praveen Reddy Nalavolu

Lexington, Kentucky

Director: Dr. Kevin D. Donohue, Professor of Electrical Engineering

Lexington, Kentucky

2010

MASTERS THESIS RELEASE

I authorize the University of Kentucky

Libraries to reproduce this thesis in

whole or in part for purposes of research.

Signed: _____

Date: _____

Dedicated to my family.

Acknowledgments

I owe my deepest gratitude to my advisor Dr. Kevin Donohue, for his valuable guidance throughout this work. He has made his support available in a number of ways and encouraged me come up with the thesis idea. I would like to thank Dr.-ing Jens Hannemann and Dr. Robert Heath for being a part of my thesis committee and providing their views on my work.

I would like to thank Dr. Naoki Saito for his assistance during the initial stage of this work. I would like to thank Harikrishnan Unnikrishnan for his enthusiasm and willingness to discuss ideas.

I would like to thank my parents whose foresight and values paved the way for a privileged education. Thanks are due to all my friends, whose support has enabled me to complete this thesis and have a wonderful time along the way.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Automatic sound source localization has a wide array of applications including talker tracking, human computer interaction (HCI) and robotics[1]. Sound source localization using microphone arrays have been popular since long. Different methods based on steered beamformers, high resolution spectral estimation and time difference of arrival (TDOA) are used for sound source localization[2]. Localization strategies based on one of these methods have limited applications as they are either computationally expensive or are less robust to reverberant and noisy conditions.

Steered response power (SRP) algorithm is a localization algorithm based on steered beamformers and TDOA methods. The algorithm uses filter and sum beamforming operation. The microphone signals received are time aligned by applying suitable time shifts and their correlation terms are summed together to obtain the steered response power. Auto correlation terms are independent of the sound source position and are subtracted from the SRP values to obtain coherent power values termed as steered response coherent power(SRCP). The SRP beamformer creates a likelihood function over the field of view (FOV), that can be represented as an intensity image of the acoustic environment. Sound source positions in the intensity image are associated with higher SRP values and the presence of coherent noise and reverberations induce false peaks in the intensity image.

Performance of SRP algorithm under coherent noise conditions can be improved by using phase transform (PHAT)[2]. Applying phase transform effectively whitens the signal spectrum and PHAT processing results in better acoustic images with sharper targets and attenuated noise fields[3]. However PHAT tends to over amplify the noise spectral regions especially in case of narrow band signals or when there is significant independent noise present over the whole frequency band[4][5]. A variation of phase transform called modified phase transform or PHAT-β is introduced to control the magnitude of spectral whitening. The value of parameter β to be employed depends on the nature of the sound sources present in the system[6].

[7] Presents an image processing method based on Canny edge detection for detecting mines in Acoustic and Radar images. The edges found by Canny edge detector are usually strong and form a boundary of mine when exists. Acoustic images created using SRP-PHAT β are similar to the images considered in [7] and have a few distinct regions of high response power pixels corresponding to sound sources. Detecting these pixels is equivalent to finding sound source locations inside the FOV. Edge detection techniques can be employed to separate regions of high contrast, typical of sound source locations. This thesis presents an image processing method based on Canny edge detection and Hough transform based ellipse detection, to automatically detect sound sources present inside microphone array FOV. Sound source detection performance is analyzed using area under receiver operating characteristic curve[8].

Chapter 2. provides an introduction to sound source localization strategies and steered response power computation. The chapter also explains phase transform and modified phase transform used to improve the localization performance.

Chapter 3. introduces the concept and mathematical background of Hough transform based ellipse detection.

Chapter 4. focuses on the implementation of Hough transform based ellipse detection. A detailed explanation of the simulation used in this thesis work is provided and the parameters considered for practical implementation are explained. Results obtained using Hough transform based ellipse detection are presented and discussed.

Chapter 5. describes a simplified method based on Canny edge detection. Results are compared with the results obtained using Hough transform based ellipse detection and direct peak detection method.

Chapter 2. Concept of Steered Response Coherent Power with PHAT-β

2.1. Introduction

Distributed microphone arrays are used for a variety of applications including beamforming[9][10], human-computer interaction and talker tracking[11][12]. Sound source localization is an important part of many of these applications. Steered response power algorithm with phase transform (SRP-PHAT) is one robust algorithm used for sound source localization in reverberant and multiple speaker environments[2].

This Chapter explains the concept and mathematical background behind the SRP-PHAT algorithm. Section 2.2. introduces the basic classification of existing microphone array based sound source localization procedures. Section 2.3. explains the concept of the robust localization algorithm based on SRP-PHAT model and the modified SRP-PHAT model used in this thesis work known as SRP-PHAT- β .

2.2. Sound Source Localization Strategies

Sound source localization Strategies using microphone arrays can be classified under three categories[2].

1. Steered beamformer based locators.
2. High resolution spectral estimation based locators
3. TDOA based locators.

2.2.1. Steered beamformer based locators

These locators use a focused beamformer, to steer the microphone array to various locations in the FOV and searches for a peak in the resultant output power in order to estimate the maximum likelihood sound source location[2]. Delay and sum beamformers, the simplest of these locators time align each of the microphone channel responses and adds them up to get the resultant power. These locators are computationally expensive

and the steered response of a conventional beamformer depends heavily on the spectral content of the sound source signal.

2.2.2. High resolution spectral estimation based locators

These are based on beamforming techniques adapted from the field of high-resolution spectral analysis methods such as autoregressive modeling, minimum variance spectral estimation and Eigen analysis-based techniques[2]. They are used in a variety of array processing applications but they have the following limitations. These algorithms are less robust to source and sensor modeling errors and assume ideal source radiators, uniform sensor channel characteristics, exact knowledge of the sensor positions[2].

2.2.3. TDOA based locators

The third category is TDOA based locators. These locators use the time delay data for each pair of microphones along with known microphone locations, to generate hyperbolic curves which are intersected in an optimal fashion to find the sound source location. The time delay estimation in these locators is complicated by the presence of background noise and room reverberations. In the noise only case with known noise statistics, the maximum likelihood time-delay estimate is obtained from a SNR-weighted version of the generalized cross correlation (GCC) function[2]. A more robust version of GCC locators known as GCC-PHAT uses phase transform (PHAT) to obtain a peak in the GCC-PHAT function corresponding to the dominant delay in the reverberated signal.

The TDOA based methods are computationally less expensive, but they have limitations as they assume a single source model. multiple simultaneous sound sources, which is often a case in sound source localization applications, excessive ambient noise or moderate to high reverberation levels in the acoustic field typically results unreliable sound source locations.

The limitations listed above restrict the usage of these locators in realistic acoustic environments. Brandstein et al. have introduced a localization algorithm known as SRP-PHAT based on the concept of Steered beamformer based locators and TDOA based

methods[2]. The localization scheme is shown to perform better in moderate ambient noise and reverberation levels compared to the previous locators.

Donohue et al. have introduced a modification of SRP-PHAT called as SRP-PHAT- β to further improve the sound source localization performance in reverberant and noisy environments[4]. The parameter β is used to control the extent of spectral whitening of the magnitude spectrum. This thesis work uses SRP-PHAT- β for sound source localization and a detailed explanation of the technique is given in section 2.3.

2.3. **SRP-PHAT-β**

This section explains the concept behind the SRP localization algorithm and the application of modified phase transform (PHAT- β) for enhanced robustness in low and moderate reverberant conditions.

The sound wave field in a room is considered to be linearly related to the sound source signal. This concept is based on the assumption that sound waves propagate as predicted by the linear wave equation as mentioned in[13]. Consider a setup of microphones and sound sources distributed inside a 3-D field of view (FOV) as shown in Figure2.1. Let $u_i(t;\vec{r}_i)$ be the pressure wave resulting from the i^{th} sound source at location r_i, where r_i is a position vector denoting the *x*, *y* and *z* axis coordinates. The waveform received at the p^{th} microphone $v_{p,i}(t;\vec{r}_p,\vec{r}_i)$ is given by[4]:

$$v_{p,i}(t;\vec{r}_p,\vec{r}_i)=\int_{-\infty}^{\infty} h_{p,i}(\lambda;\vec{r}_p,\vec{r}_i)\,u_i(t-\lambda;\vec{r}_i)\,d\lambda + \sum_{k=1}^{K}\int_{-\infty}^{\infty} h_{p,k}(\lambda;\vec{r}_p,\vec{r}_i)\,n_k(t-\lambda;\vec{r}_k)\,d\lambda + n_p(t) \tag{2.1}$$

where $h_{p,i}(.)$ represents the overall impulse response of the propagation path from $\vec{r}_i$ to $\vec{r}_p$. $h_{p,i}(.)$ is a combination of the microphone channel response and the room impulse response. The microphone channel response takes in to account of the different electrical, mechanical and acoustical properties of the microphone system. The room impulse

response depends up on room temperature, humidity and the position and motion of different physical objects inside the room. $n_k(t)$, $n_p(t)$ are the correlated and uncorrelated noise sources present inside the room.

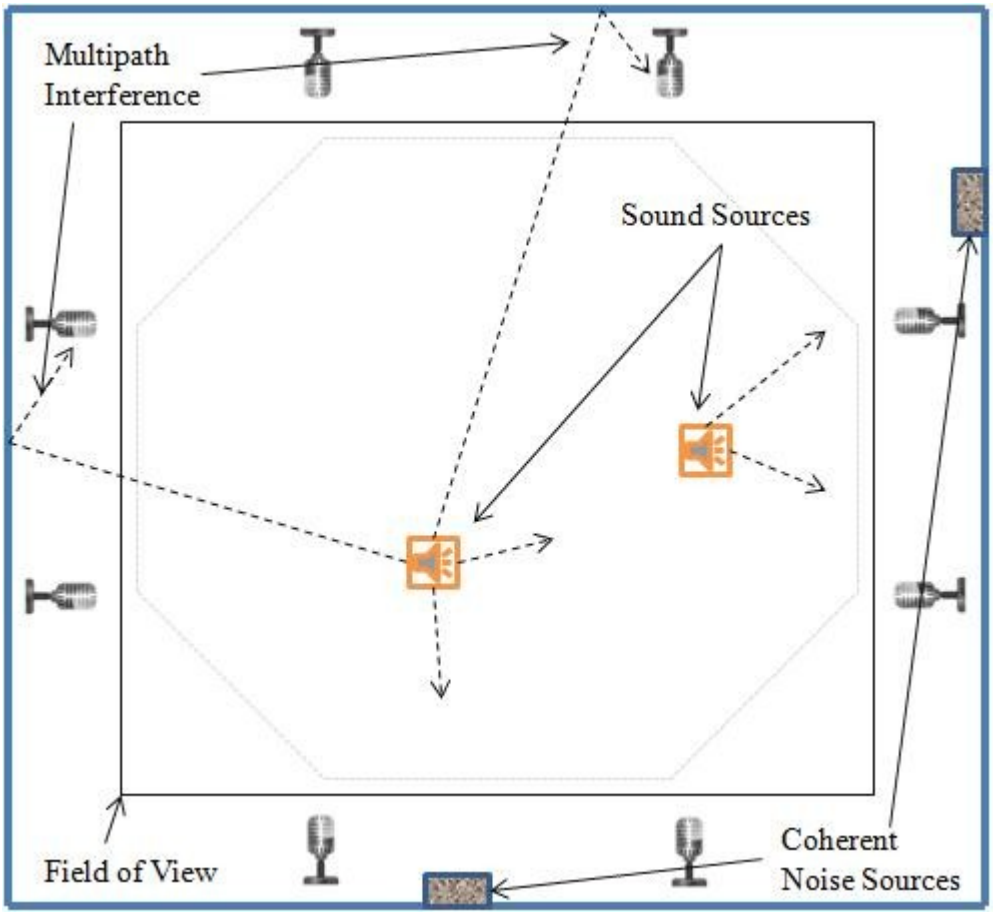


Figure 2.1: Schematic of sound sources and interfering sources in a perimeter microphone array system.

The correlated noise term $n_k(t)$ is a result of other sound sources present inside the FOV and ambient noise sources outside FOV. The uncorrelated noise term $n_p(t)$ is a result of channel noise in microphone system. Correlated noise is hard to suppress and in general is the significant noise present in the system[2][4].

The propagation path impulse response $h_{p,i}(.)$ is a combination of direct path component and reflected path component. Consequently the impulse response can be expressed as[4]:

$$h_{p,i}(t;\vec{r}_p,\vec{r}_i)=h_{p,i}(t)=a_{p,i,0}(t-\tau_{p,i,0})+\sum_{n=1}^{\infty}a_{p,i,n}(t-\tau_{p,i,n}) \tag{2.2}$$

Where $a_{p,i,0}(t)$ represents the impulse response of direct path component and $a_{p,i,n}(t)$ represents the impulse response of n^{th} reflected path component between source at $\vec{r}_i$ and microphone at $\vec{r}_p$, $\tau_{p,i,n}$ is the corresponding path delay. The SRP pixel estimate is based on the sound events limited to those received over a finite time frame denoted by Δ_l. The value of Δ_l depends upon the steering delay for focusing the array at the appropriate source spatial location and compensation for the direct path propagation delay associated with the desired signal at each microphone[2]. The waveform received by p^{th} microphone, resulting from signal segments during the interval Δ_l can be represented in frequency domain as[4]:

$$\begin{aligned}\hat{V}_{p,l}(\omega)=&\sum_{i=1}^{N_T}\hat{U}_{i,l}(\omega)\sum_{m/\tau p,i,m\epsilon\Delta_l}\hat{A}_{p,i,m}(\omega)\exp(j\omega\tau_{p,i,m})\\&+\sum_{k=1}^{K}\hat{N}_k(\omega)\sum_{m/\tau p,i,m\epsilon\Delta_l}\hat{A}_{p,k,m}\exp(j\omega\tau_{p,k,m})+\hat{N}_p(\omega)\end{aligned} \tag{2.3}$$

Where $\hat{U}_{i,l}(\omega)$ is the Fourier transform of the i^{th} sound source field $u_i(t)$ over the interval Δ_l. N_T is the number of target sound sources inside the FOV and K is the number of noise sources. The summation index in the above equation indicates the summing of signal components whose path delay is falling with in the interval Δ_l.

During propagation, the attenuation of a sound signal depends upon its frequency and in general higher frequencies are more attenuated compared to the lower frequencies. This condition makes the estimate values obtained in Eq.2.3 dependent up on the magnitude spectrum of the sound source. Phase transform (PHAT) is introduced to make SRP values independent of the magnitude spectrum. The application of PHAT whitens the whole spectrum to equally emphasize on all frequencies[2][14][15].

The PHAT is a robust weighting scheme and does not require signal and noise characteristic information[5]. However PHAT tends to over amplify the noise spectral

regions especially in case of narrow band signals or when there is significant independent noise present over the whole frequency band[4][5].

To overcome these defects a variation of Phase transform known as PHAT-β [5] is introduced to control the extent of whitening the spectrum and to limit the amount of degradation due to independent noise. PHAT-β is shown to improve sound source location performance for both narrow band and wide band signals[3][4][5][6][16].

PHAT-β is defined as[4]:

$$\hat{\theta}_{p,l}(\omega,\beta)=\frac{|\hat{V}_{p,l}(\omega)|}{|\hat{V}_{p,l}(\omega)|^{\beta}}\measuredangle\hat{V}_{p,l}(\omega) \tag{2.4}$$

The parameter β takes values between 0 and 1. When β is 1 the magnitude of the Fourier transform is one for all the frequencies. It is worth mentioning that for conventional PHAT β value is always equal to 1. When β is 0 PHAT-β has no effect on the signal.

Based on experimental results from[6], PHAT-β improves sound source localization for β values ranging from 0.65 to 0.9 for broadband signals and 0.4 to 0.75 for narrowband signals under low reverberation conditions. When the reverberation levels are high suggested β values are 0.6 to 1 for broadband signals and 0.2 to 0.7 for narrowband signals.

The Steered response coherent power (SRCP) value is obtained by subtracting self power terms of microphone channel responses from the sum of cross power terms. This can be expressed in the discrete frequency domain as[4]:

$$S_{i,l}(\beta)=(\Delta_{\omega}/T)\sum_{k=K1}^{K2}(|\sum_{p=1}^{P}\hat{B}_{p,i}\hat{\theta}_{p,i,l}(\omega_k,\beta)|^2-\sum_{p=1}^{P}|\hat{B}_{p,i}\hat{\theta}_{p,i,l}(\omega_k,\beta)|^2) \tag{2.5}$$

Where T is the length of the interval Δ_l, $K1$ and $K2$ are upper and lower frequency limits of the signal bandwidth. Δ_{ω} is the frequency domain sampling interval and $\hat{B}_{p,i}$ is the

complex weight representing the delay and filtering associated with the image location and array geometry. The simulation used in this thesis work considered $\hat{B}_{p,i}$ values equal to the reciprocal distance between the p^{th} microphone and i^{th} SRP pixel location. The values are normalized by the sum of reciprocal distances over all array elements[4]. Thus pixels which are closer to the microphone are weighted more compared to the pixels which are farther from the microphone.

The coherent power values $S_{i,l}(\beta)$ calculated at every r_i inside FOV using Eq.2.5 will become the pixel values of the SRCP Image.

Chapter 3. Concept of Hough Transform based Ellipse Detection

3.1. Introduction

Sound source locations in an SRCP image are usually associated with high response power and in general can be thought of as discontinuities from their background power levels. Image processing techniques can be employed to detect these discontinuities in SRCP images and Hough transform based ellipse detection is one such robust method, that can be used for sound source detection.

This chapter explains the concept of Hough transform based ellipse detection (HTED). The chapter is divided in to 3 sections. Section 3.2. explains the concept of Canny edge detection used for pre-processing of the input image data. Section 3.3. introduces the concept and mathematical background behind ellipse fitting. Section 3.4 explains the process of detecting centers of elliptical shapes in the processed image data using Hough transform.

3.2. Canny Edge Detection

In image processing an edge is basically a local discontinuity in pixel values that exceeds a particular threshold[17]. SRCP image explained in previous section is a representation of the sound field inside a FOV and in general contains a lot of data. However a few high power pixels in the SRCP image contain the most important information about sound source locations. Edge detection techniques can be used to extract necessary data about these pixels. An effective edge detection technique should extract necessary information about sound sources, at the same time should reject unwanted data such as data related to the background. This will reduce the amount of data fed to the subsequent steps, and Canny edge detection (CED) is used to achieve this purpose.

Like most edge detection schemes CED consists of 3 stages: Filtering, Differentiation and Detection[17]. An input image is convolved with a filter during the filtering stage. CED employs a Gaussian filter for smoothing the input SRCP image

during the filtering stage. The Gaussian filter uses weighted averaging of the SRCP pixel values and the pixel weight is inversely proportional to the distance from the center pixel of the filtering window. The level of smoothing depends up on the σ value used.

The differentiation stage gives the preliminary edge data information. Edges in an image can be detected by performing a first order derivative, as the derivative is associated with a high value at the edge. Consider δ_x and δ_y be the gradients of the input SRCP image in x and y directions. The gradient magnitude G and direction η can be expressed as

$$\eta = \tan^{-1}\left(\frac{\delta_y}{\delta_x}\right) \text{ and } G = \sqrt{(\delta_x{}^2 + \delta_y{}^2)}$$

The edge orientation (slope) data θ can be obtained from the gradient data by adding 90 degrees to the edge gradient direction.

The detection stage of CED has two parts. The first part involves performing a non-maxima suppression on the gradient magnitude using the gradient direction information. In general if a pixel gradient magnitude is not varying significantly in the direction of gradient then that pixel is probably not an edge point. Consider $G(x, y)$ as the gradient magnitude and $G(x_1, y_1)$, $G(x_2, y_2)$ as the gradient magnitude values on either side of the edge pixel in the direction of edge gradient. Mathematically non maxima suppression can be expressed as:

$$G(x, y) = \begin{cases} G(x, y) & \text{if } G(x, y) > G(x1, y1), G(x2, y2) \\ 0 & \text{otherwise} \end{cases}$$

This step eliminates all the points which are not potential edge points.

Hysteresis thresholding is applied to the non-maxima suppressed magnitude during the second part of the detection stage. CED uses two thresholds a lower threshold (t_{low}) and a higher threshold (t_{high}) to deal with the problem of streaking. If the gradient magnitude at a pixel is greater than t_{high}, it is considered as an edge pixel. Pixels which are 8 connected to the edge pixels determined above and whose gradient magnitudes exceed the t_{low} are also considered as edge pixels.

The success of Ellipse detection method depends on the effectiveness of CED. CED is very efficient at finding the edge pixels around sound sources in an SRCP image. Figure3.1 shows an SRCP image produced using the simulation with two sound sources at the center of circles marked in red. It can be observed that the sound source pixels have higher power compared to the background levels, which is evident from the color of the pixels. The result of applying Canny edge detection on Figure3.1 is shown in Figure3.2.

The canny edge detector used is a MATLAB version of an implementation of the Robot Vision Group in the Department of Artificial Intelligence at the University of Edinburgh[18].

The program output is an array of edge magnitudes and edge orientations. Except edge pixels, most of the pixels in the edge magnitude array are set to zero. This significantly reduces the number of computations required to implement ellipse detection.

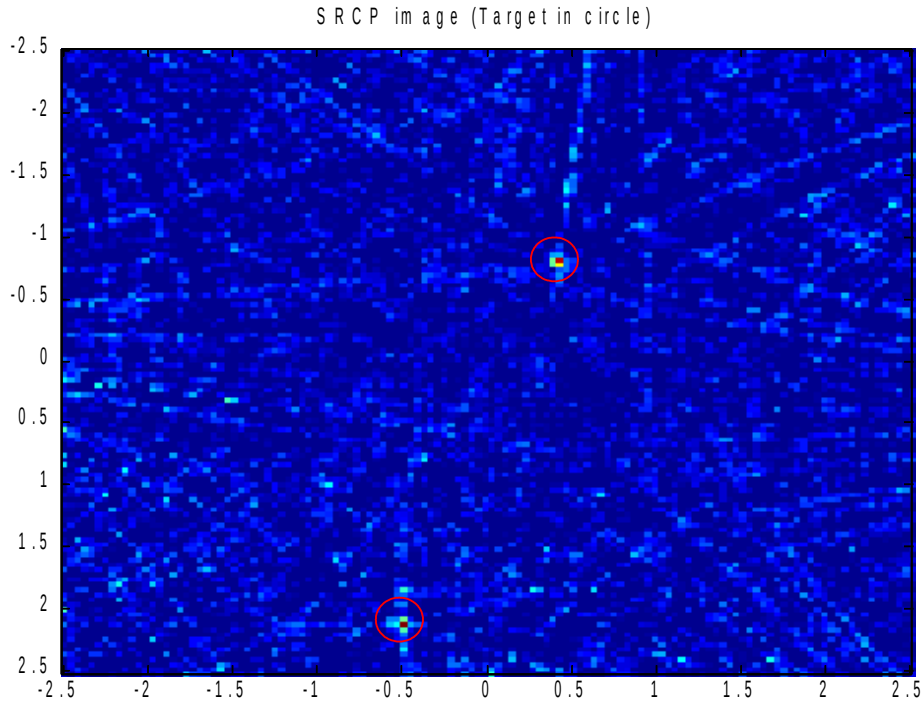


Figure 3.1: SRCP image with sound sources positioned at the center of circles.

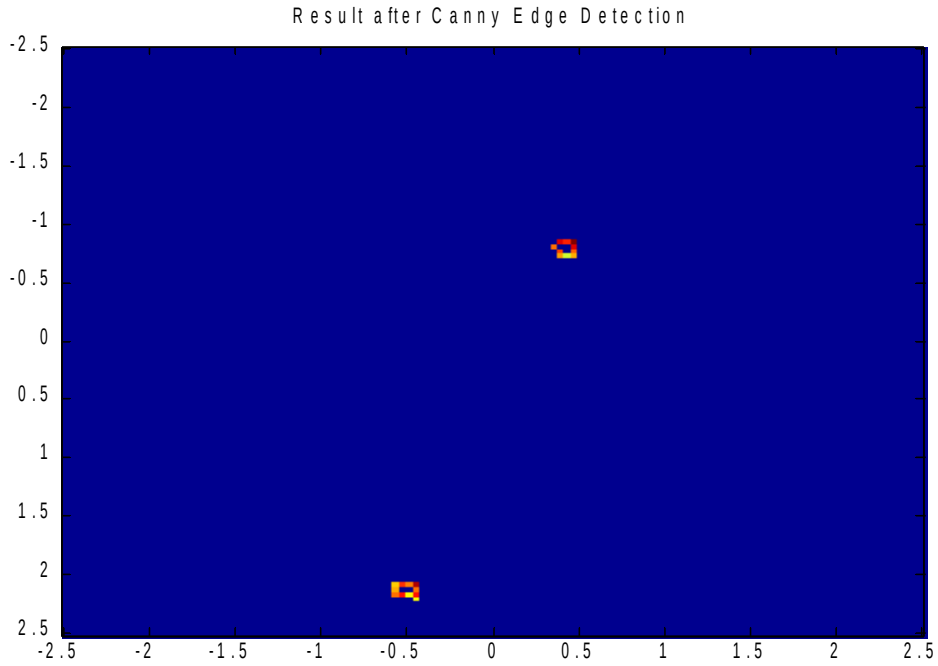


Figure 3.2: Result of applying Canny edge detection on Figure3.1.

3.3. **Ellipse Fitting**

Sound source locations yield clear boundaries when CED is applied on SRCP images as evident from Figure 3.2. The shape of these edge pixel boundaries can be approximated to an ellipse. The basic idea is to detect the center of these elliptical boundaries which in most cases represents a sound source location in an SRCP image.

A detailed description of translating image data in to inferences about possible ellipses is presented in[18]. This Section explains the mathematical characterization of the family of ellipses passing through a pair of points. The ellipse directions are determined by the associated normal directions at the pair of points obtained during CED.

3.3.1. Parameterization of the ellipse

Consider two points P_1, P_2 such that $P_1=(x_1, y_1)$ and $P_2=(x_2, y_2)$. Let the normal directions associated with each point be $N_1=(p_1, q_1)$ and $N_2=(p_2, q_2)$. The normal directions are supposed to be pointing into the angular sector between the tangent lines containing the other point as shown in Figure 3.3. That is,

$$p_1(x_2-x_1)+q_1(y_2-y_1)>0 \text{ and}$$

$$p_2(x_1-x_2)+q_2(y_1-y_2)>0 \tag{3.1}$$

The equation of line *P1P2* is given by

$$L(x,y)=(y_1-y_2)x+(x_2-x_1)y+x_1y_2-x_2y_1=0 \tag{3.2}$$

The tangent lines at P_1 and at P_2 are given by

$$l_1(x,y)=p_1(x-x_1)+q_1(y-y_1)=0 \text{ and}$$

$$l_2(x,y)=p_2(x-x_2)+q_2(y-y_2)=0$$

The functions L, l_1 and l_2 are each linear in x and y. The values p_1, q_1, p_2, q_2 are obtained from the orientation data generated during the Canny edge detection. For a given constant λ,

$$C(x,y)=L^2(x,y)-\lambda l_1(x,y)l_2(x,y)=0 \tag{3.3}$$

is quadratic in x and y and represents a conic section C. For $\lambda=0$ the conic C represents the line $L=0$ and for $\lambda=\infty$ the conic represents the pair of lines $l_1=0$ and $l_2=0$. For intermediate values of λ, C passes through the intersection of L, l_1 and l_2 touching l_1 at point P_1 and l_2 at point P_2 [18], as shown in Figure 3.4. The conic C can be rewritten as:

$$C(x,y)=ax^2+2hxy+by^2+2gx+2fy+c \tag{3.4}$$

where a, b, c, f, g, h are linear functions of λ and also depend on x_1, y_1, p_1, q_1, x_2, y_2, p_2, q_2.

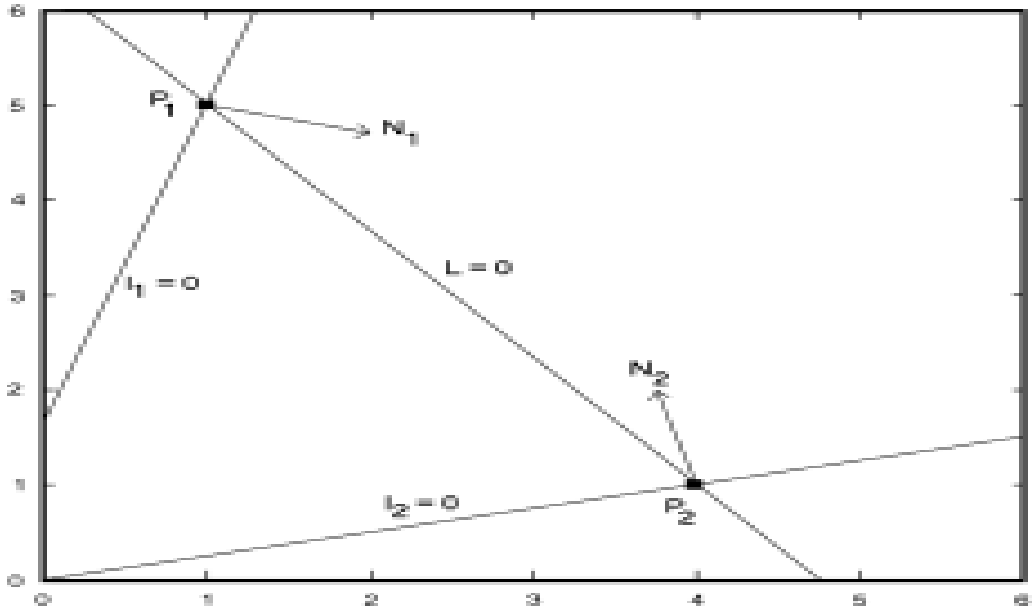


Figure 3.3: Graphical representation of L, l_1 and l_2 (Figure adapted from [18])

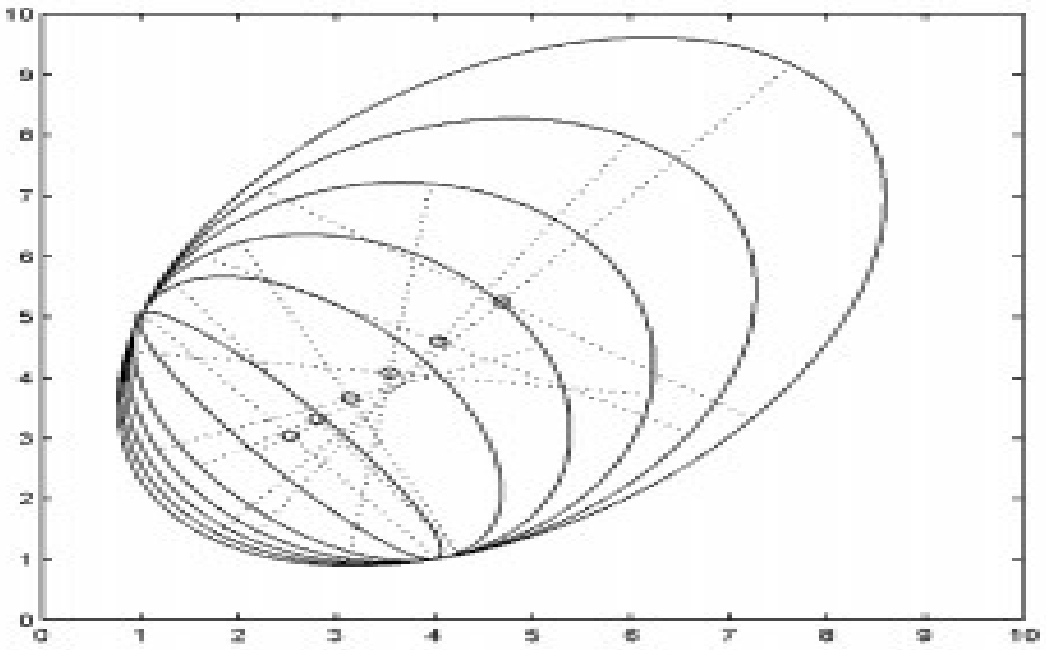


Figure 3.4: Graphics of Eq. 3.3 for various values of λ ranging from 1 to 39 (Figure adapted from [18])

Eq. 3.4 can be written in matrix form as

$$X^T AX + 2\mathrm{F}^T X + c = 0 \tag{3.5}$$

Where $X = (x, y)^T$ and $F = (g, f)^T$ and A is a matrix

$$A = \begin{pmatrix} a & h \\ h & b \end{pmatrix} \tag{3.6}$$

The center of the conic $X_0 = (x_{0,} y_0)^T$ is written as[18]:

$$X_0 = -A^{-1} F \tag{3.7}$$

3.3.2. Range of λ for which the conic is an ellipse

The properties of the conic section C as in Eq.3.3 varies significantly depending upon the value of λ . Substituting L , l_1 and l_2 in Eq.3.3 and comparing with Eq.3.4, a , b , c , f , g , h can be expressed as[18]:

$$a(\lambda) = (y_1 - y_2)^2 - \lambda(p_1 p_2)$$

$$b(\lambda) = (x_1 - x_2)^2 - \lambda(q_1 q_2)$$

$$c(\lambda) = (x_1 y_2 - x_2 y_1)^2 - \lambda(p_1 x_1 + q_1 y_1)(p_2 x_2 + q_2 y_2)$$

$$f(\lambda) = (x_1 y_2 - x_2 y_1)(x_2 - x_1) + 1/2\lambda[q_2(p_1 x_1 + q_1 y_1) + q_1(p_2 x_2 + q_2 y_2)]$$

$$g(\lambda) = (x_1 y_2 - x_2 y_1)(y_1 - y_2) + 1/2\lambda[p_2(p_1 x_1 + q_1 y_1) + p_1(p_2 x_2 + q_2 y_2)]$$

$$h(\lambda) = (y_2 - y_1)(x_2 - x_1) + 1/2\lambda(p_1 q_2 + p_2 q_1) \tag{3.8}$$

When $\lambda<0$ the conic C represents a hyperbola outside the sector formed by l_1 and l_2. When $\lambda=0$ the conic represents the line segment P_1P_2, which can be verified by substituting the value of λ in Eq.3.3. When λ is small and positive, the conic C is close to the line segment P_1P_2. This variation of behavior of conic C can also be expressed in terms of matrix A defined in Eq.3.6. When $\lambda=0$, The matrix A is singular. When λ increases, A at first becomes positive definite and then becomes singular (indefinite) again for a positive value value of $\lambda=\lambda_0$. For λ greater than λ_0, the matrix A becomes indefinite. This behavior corresponds to the center of the conic $X_0=(x_0, y_0)^T$ at first receding to infinity, so that the conic tends to a parabola for $\lambda=\lambda_0$, and then, for $\lambda>\lambda_0$, the conic becomes a hyperbola, but this time within the sectors for which l_1l_2 is positive. Thus, the range of λ for which the conic C of Eq.3.3 is an ellipse is the interval $0<\lambda<\lambda_0$. Hence, to find the range of λ values for which the conic is an ellipse, we must solve

$$det(A)=\begin{vmatrix} a & h \\ h & b \end{vmatrix}=0 \tag{3.9}$$

For λ, one root is zero; the other is the required value λ_0 given by [18]:

$$\lambda_0=4[p_1(x_2-x_1)+q_1(y_2-y_1)][p_2(x_1-x_2)+q_2(y_1-y_2)]/(p_1q_2-p_2q_1)^2 \tag{3.10}$$

Thus, the conic C represents an ellipse for $0<\lambda<\lambda_0$. For λ in this range, the coordinates of the center of the ellipse, $(x_0(\lambda), y_0(\lambda))$ can be calculated using Eq.3.7.

3.4. **Ellipse Center Detection**

The purpose of finding the centers of elliptic edge boundaries can be achieved by using Hough transform. Given the edge magnitude and orientation data, the parameterization explained in the above section can be used to fit ellipses for each pair of edge points for λ varying in the range $0<\lambda<\lambda_0$. Vote for the center position of these ellipses passing through the edge points and satisfying predefined conditions in an array. The array cells

receiving maximum number of votes are the probable centers of the elliptical edge pixel groups[18][19][20].

The Hough Transform uses an accumulator array for collecting the votes. The accumulator array has the same size as that of an input SRCP image and can be thought of as a discretized position space for the center of an ellipse. The array contents are initially set to zero.

Consider a pair of edge pixels P_1 and P_2 such that the distance between the two points d_{p1p2} is less than distance threshold d_{thresh} . The distance threshold is necessary so as to consider pixels which correspond to the same edge boundary. Given the edge data the range of λ values for which the conic passing through the edge points P_1 and P_2 will be an ellipse can be calculated using the position and orientation information. The upper bound of λ value, λ_0 is calculated using Eq.3.10. For $0 < \lambda < \lambda_0$ ellipses passing through the edge points are fitted, whose centers are given by Eq.3.7. Every time an eligible ellipse is fitted for the pair of pixels, the accumulator array cell corresponding to the center of the ellipse is incremented.

The process is repeated for all the pairs of edge pixels. If there exists an ellipse in the edge data so that many pairs of edge points correspond to it, the center of that ellipse accumulates many votes[18]. The accumulator array cell receiving maximum number of votes in a region is the probable center of the edge pixel regions and consequently is the probable sound source location.

Chapter 4. Implementation of Hough Transform based Ellipse Detection

4.1. Introduction

Chapter 3 outlined the procedure and necessary background to detect the centers of elliptical groups of edge pixels using Hough Transform based ellipse detection. This chapter explains the procedure and implementation of the HTED algorithm, to detect sound sources inside a microphone array environment. Section 4.2. explains the simulated environment used for creating the SRCP image. The parameters considered during practical implementation are explained in section 4.3. Results obtained using this algorithm are discussed in section 4.5.

4.2. Simulation Design

This thesis work employed a simulation for the purpose of generating SRCP images under various operating conditions. The simulation is similar to the one used in [4] except for a change in few parameters and dimensions. The simulation is a part of the Array Toolbox developed by Audio Systems Laboratory at the University of Kentucky.

The simulation is inspired from an actual audio cage array in Audio Sensing and Rendering Lab at University of Kentucky. The dimensions of the rectangular room used in simulation are outlined in Table 4.1.

Parameters	Value
Length and Width(room)	7m * 8m
Height	3.5m
Length and Width (FOV)	5m * 5m
Source Height	1.5m
Speed of Sound	348 m/s
Number of Microphones	8(Perimeter Array)
Microphone Spacing	2.8995
SRCP Computation Grid	4 cm

Table 4.1: Array Simulation Parameters

The simulation produces impulse like sound signals inside the FOV by placing sound sources at random locations inside the FOV. The sound signal is the impulse response of a Butterworth filter with 3 dB cutoff frequencies of 300 Hz and 3000 Hz. Two coherent noise sources representing room noise under practical conditions are simulated outside the field of view on the actual room wall. The strongest signal on the microphone array is used for adjusting the noise power. In addition a -30 dB white noise signal representing microphone channel noise is added to every microphone with respect to the strongest signal. The room reflection coefficient values are set at 0.8 for walls and 0.7 for floor and ceiling.

An 8 microphone perimeter array is used to record the sound produced. The microphones are omnidirectional and are offset by 0.25m toward the center of the room from the room walls. The microphone array is steered to each point in the FOV to generate an acoustic image according to the SRCP values. The microphone signals received over each channel from a point in FOV are time aligned and are weighted according to the distance of the microphone from the point under consideration. The weight of a microphone signal is inversely proportional to the distance between the microphone and the point under consideration, thus closest microphones have more weight compared to others.

Coherent power is computed from these time aligned, weighted microphone signals. Coherent power can be negative and is obtained by subtracting the self power terms from the signal correlation products. Since coherent power only has cross correlation terms its value is large and positive if the microphone signals are correlated and is negative if there is a strong out of phase coincidence between the signals. Calculating coherent power at each pixels results in an SRCP image with its magnitude representing the likelihood of sound source at a particular position.

Sound source detection performance is improved by using a partial whitening transform known as PHAT- β as explained in[6]. The parameter β varies the magnitude of spectral whitening, with β =0 representing no whitening (original signal) and β =1 representing total whitening (Phase Transform). For most audio systems operating under different frequency conditions, an intermediate value of β was shown to improve the performance of the system. A more detailed explanation about PHAT- β can be found in [4][15][16]. According to the results presented in [6], β values between 0.5-0.8 will lead to significant performance improvements in sound source detection and a β value of 0.75 is considered for this experiment.

4.3. **Practical Implementation of SSD using Hough Transform based Ellipse Detection**

This section explains procedure and various parameters considered for the practical implementation of SSD using HTED method.

Given a SRCP image with known number of sound sources, CED is used to obtain groups of edge pixel boundaries around probable sound source locations. Gradient values are computed for the Gaussian filtered SRCP image. Pixels whose gradient values exceed t_{high} after non maxima suppression are considered as edge pixels. Pixels which are greater than t_{low} and are 8-connected to above pixels are also considered as edge pixels. The program outputs arrays of edge pixel magnitudes and normal directions at these pixels, computed using the equations mentioned in Section 3.2.

Figure 4.1 shows an SRCP image with known sound source position marked by the red circle. The application of canny edge detection results in edge pixels around sound source location and the result is as shown in Figure 4.2. 160 and 110 are used as the gradient magnitude thresholds t_{high} and t_{low} during CED. Detecting the centers of these elliptical edge pixel groups should give the sound source locations. It is also worth mentioning that except the edge pixels most other pixels in Figure 4.2 are set to zeros which significantly reduces the amount of data to be considered during the HTED stage.

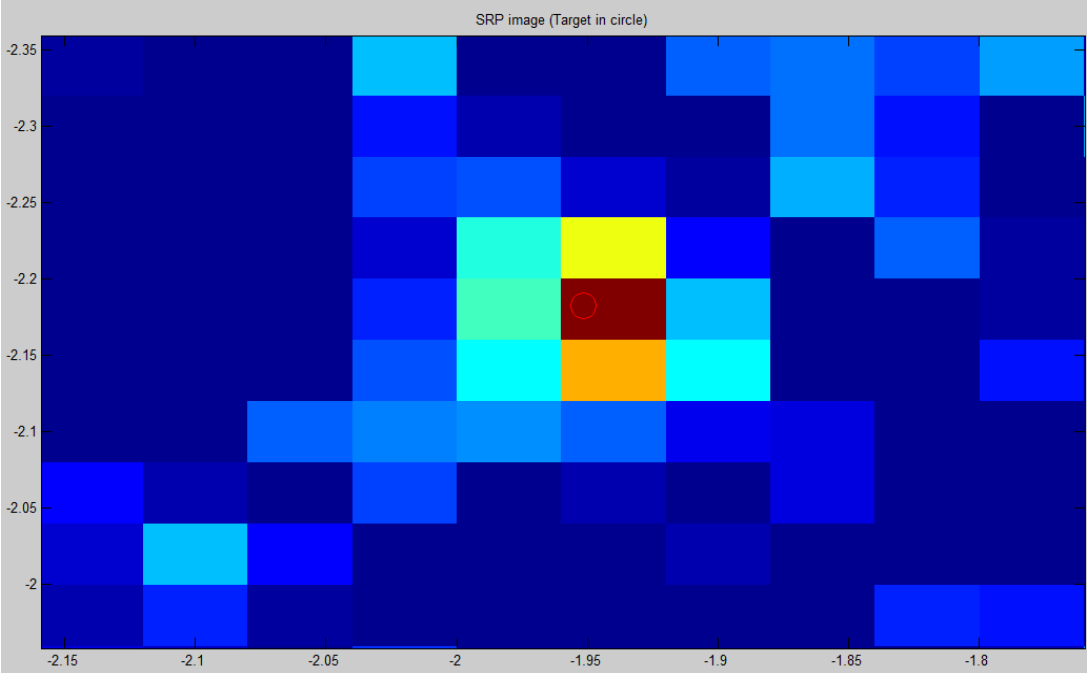


Figure 4.1: SRCP image with known sound source location

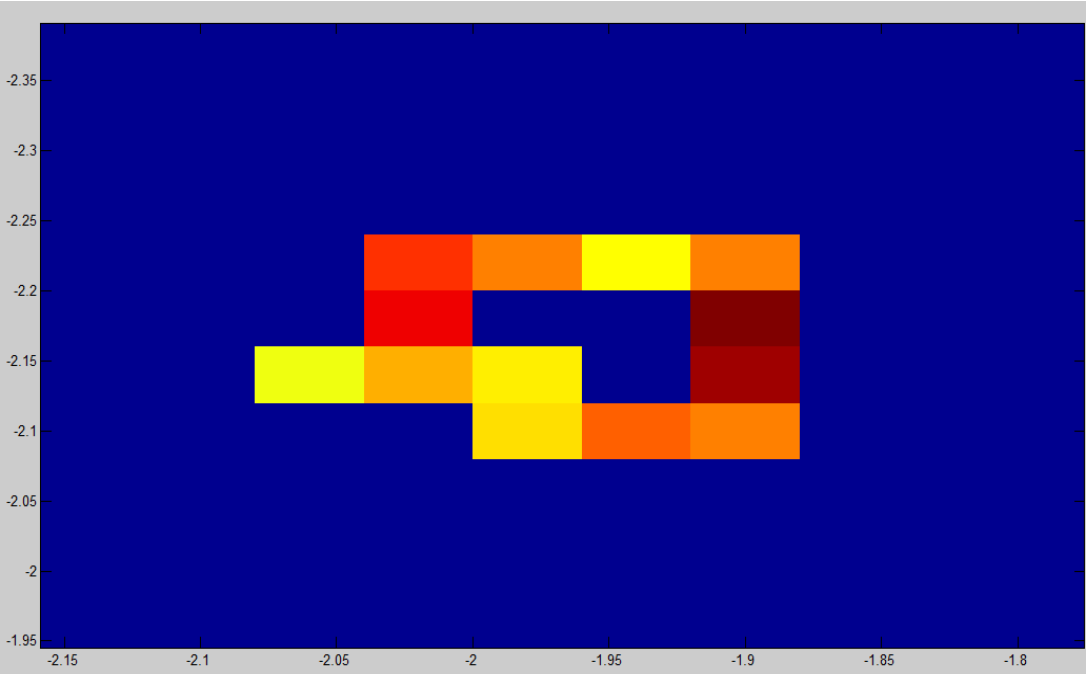


Figure 4.2: Edge pixels detected around the sound source location by Canny edge detector

Given the edge magnitude and orientation data, Hough transform is used, for detecting elliptical shapes in the edge image. Consider a pair of data points from the edge data information $\vec{e}_i$, $\vec{e}_j$, and their corresponding orientations. The values of p_i, q_i are set such that $q_i = 1$ and $p_i = \tan \eta_i$ as suggested in [18]. For each pair check if the edge points satisfy the quadrant conditions specified in Eq. 3.1. The sign of the normal vector is reversed if one of the conditions is not met i.e normal vector $-N_i$ is considered instead of N_i .

The value of λ_0 which gives the range of parameterization $0 < \lambda < \lambda_0$ over which the conic described in Eq. 3.3 is an ellipse is then computed using Eq. 3.10. once the λ_0 value is calculated, values of a , b , c , f , g , h , which characterize the ellipse are calculated using Eq. 3.8. The interval $0 < \lambda < \lambda_0$ is divided in to n increments of $d\lambda$ where $d\lambda = \lambda_0 / n$. Starting from $\lambda = d\lambda$ and for each $\lambda = \lambda + d\lambda$ an ellipse is fitted for the pair of edge points, the center of which is obtained using Eq. 3.7.

An Accumulator array of size 126*126 (same as the size of the SRCP image) is constructed to poll for the center position of ellipse fitted in the above step. This procedure is repeated for each pair of edges in the canny output image and corresponding pixel values of accumulator array are incremented. A distance threshold ($threshold_{ed}$) of $\sqrt{18}$ is used while selecting pairs of edge pixels. This threshold helps in selecting pixels which are sufficiently close together, which are more likely to be the edge pixels around a sound source. The distance threshold also minimizes the number of noise pixels in edge data considered for ellipse fitting.

Figure 4.3 shows the accumulator array for the region considered in Figure 4.2. It can be observed that the ellipse center values are distributed over a region but there are a few distinct maximum values whose corresponding position can be considered as the probable center of the elliptic region shown in Figure 4.2. MATLAB command "imregionalmax" is used to identify the local maxima in the accumulator array and the result is shown in Figure 4.3 . The local maxima are represented by a 1 in the result where as the remaining pixels are all set to 0.

score

File Edit View Graphics Debug Desktop Window Help

Stack: Base

	7	8	9	10	11	12	13	14
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	0	3	3	0	0	0
12	0	0	5	4	14	0	3	0
13	0	0	3	19	11	41	30	0
14	0	1	12	74	40	67	3	0
15	0	0	0	115	61	53	22	0
16	0	0	0	85	62	96	2	0
17	0	0	2	14	18	8	0	0
18	0	0	0	0	0	0	0	0

Figure 4.3: The Accumulator array after voting for the ellipse centers

	6	7	8	9	10	11	
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	0	0	0	1	0
15	0	0	0	1	0	0	0
16	0	0	0	0	0	1	0
17	0	0	0	0	0	0	0

Figure 4.4: Local maxima in the accumulator array(represented by 1)

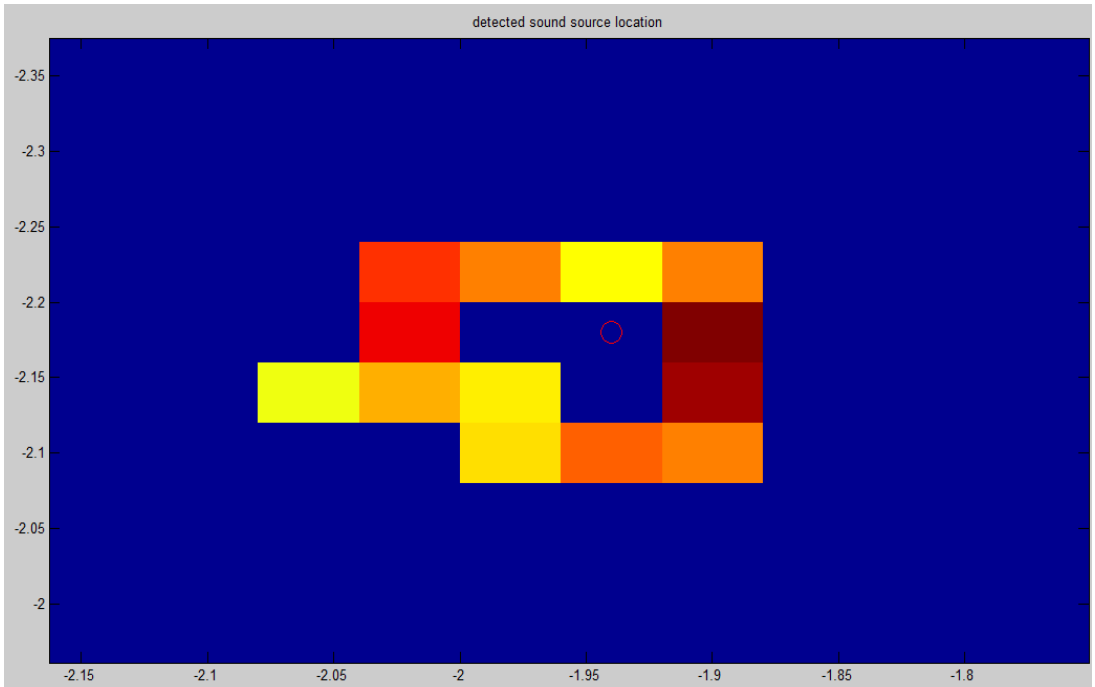


Figure 4.5: Sound source location detected (indicated by a red circle)

Accumulator array values which are local maxima are then sorted according to the number of votes received for the corresponding positions. The position of the Accumulator array with maximum number of votes is considered as the center of an elliptic region and hence a sound source location. The position corresponding to the second highest number of votes in the sorted array (decreasing order) is considered as a center of an elliptic region if and only if it is at a distance greater than target distance threshold ($threshold_{td}$) from the position corresponding to highest number of votes. $threshold_{td}$ Value is a user control and a value of $\sqrt{18}$ is used as the target distance threshold in the experiment. The positions of subsequent values in the sorted array are considered as centers of elliptical regions if they are at a distance greater than $threshold_{td}$ from each of the previously identified center locations.

The Centers of the elliptical regions in the edge data image are the probable sound source locations. Figure 4.5 shows the result, where sound source location is marked in a red circle. This location corresponds to the position (15, 9) in Figures 4.3, 4.4 receiving 115 votes as the probable center location. The local maxima corresponding to positions (14, 11), (16, 11) with 67 and 98 votes respectively are not considered as centers because they are at a distance less than $threshold_{td}$ from the known center position at (15, 9).

The MATLAB program used for identifying the center of elliptical regions is inspired from the 'C' program used for the “center finding experiment” in [18].

4.4. **Analysis Method**

In any sound source detection algorithm false alarms can not be completely eliminated and performance of the algorithm is measured in terms of its ability to discriminate between a true detection and a false alarm. Area under Receiver Operating Characteristic curve (ROCA) is used to analyze the performance of the algorithm. The ROCA is a variable between 0 and 1 and represents the priority given to a true detection over a false alarm.

True detections and false alarms in a SRCP image are determined using the known sound source location information. The ROCA program is supplied with true detection and false alarm intensities. The program sweeps a threshold over the range of values present in the two sets to compute 'pd' (probability of detection) and 'pfa' (probability of false alarm). ROCA value is obtained by computing the area under (1 - pfa) vs pd curve.

This method of ROCA computation may not be applied when there are no true detections, i.e when no sound sources are detected and all the detections are false alarms. This situation is more pronounced when canny threshold value is too high and under high coherent noise conditions, when the algorithm fails to detect any sound sources inside FOV. To overcome this constraint ROCA program is fed with true detection and false alarm intensities detected over 25 SRCP images created under identical conditions. This in a way can be explained as considering true detections and false alarms over a 25m * 25m FOV. This resulted in sufficient number of true detections under normal conditions. The experiment is repeated 4 times in each case and the results are averaged for consistency.

4.5. **Results and Discussion**

The parameters used in the simulation are explained in section 4.2. To emulate real world conditions two coherent noise sources are placed on the room walls. Performance of HTED method is explained using the ROCA values achieved during the simulation experiments.

Figure 4.6 shows the results obtained using this method. Two coherent noise sources of -25 dB are placed on room walls for the experiment. The number of sound sources is varied between 1-4, so as to test the performance of the algorithm in presence of multiple sound sources inside the FOV.

The Figure shows that ROCA value decreases as the number of sound sources inside the FOV increases. This fall in ROCA value can be attributed to the fact that noise in the system increases as the number of sound sources increases. While computing

SRCP power of a sound source at a particular location, all other sound sources are treated as noise sources. Also, increase in the number of sound sources results in increased reflections and hence increased noise in the system.

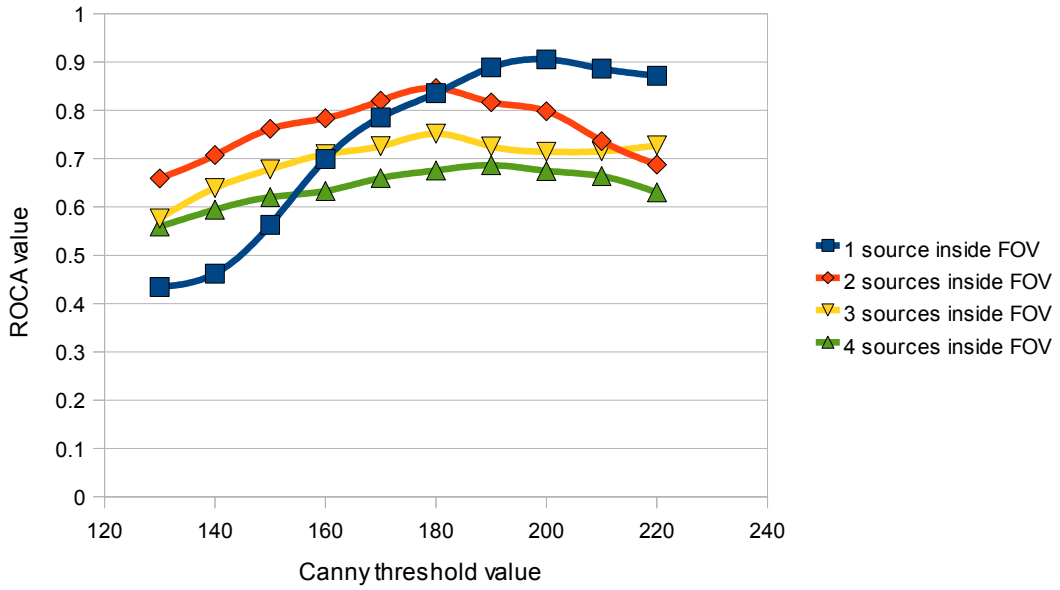


Figure 4.6: Performance of HTED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -25dB are used for the experiment.

The horizontal axis in Figure 4.6 represents Canny threshold value, which is varied from 220 to 130. This value represents the higher threshold (t_{high}) used for deciding whether a pixel is an edge pixel or not during Canny edge detection step. A threshold of t_{high} - 20 is used as lower threshold during the same step.

Figure 4.6 shows that ROCA values for a fixed number of sound sources inside FOV varies with a change in Canny threshold value. This scenario can be explained as follows. When Canny threshold is decreased, the number of pixels classified as edge pixels increases. This might result in two cases.

(1) Increased number of edge pixels around actual sound source location, which increases the number of edge pixel pairs for the ellipse detector stage. This results

in a greater ellipse count for the sound source location and hence a stronger detection statistic.

(2) Increased number of edge pixels around pixels other than actual sound source locations, resulting in a higher number of noise pixels. This increases the ellipse count around noise peaks and hence results in a stronger false alarm statistic. This effect decreases ROCA values.

Thus ROCA value reaches a maximum for a particular canny threshold value and increases or decreases for other threshold values as shown in Figure 4.6. In general ROCA value is higher for Canny threshold values around 180 – 200 for 2, 3 and 4 sound sources inside FOV.

To find the performance of these algorithms under increased noise conditions, two coherent noise sources of -10dB are placed on the room walls. Figure 4.9 shows the results obtained using HTED method under these circumstances. The number of sound sources present inside FOV is varied between 1-4.

Comparing Figure 4.6 and Figure 4.9; the detection performance of the algorithm decreases with an increase in the coherent noise present in the system. There is a steep decrease in the ROCA values in the second case, when coherent noise sources of SNR -10dB are placed on the room walls. The presence of stronger coherent noise sources results in stronger noise peaks in the SRCP image. Under these conditions noise pixel magnitudes are comparable to sound source pixel magnitudes and the algorithm can no longer distinguish between a sound source peak and a noise peak.

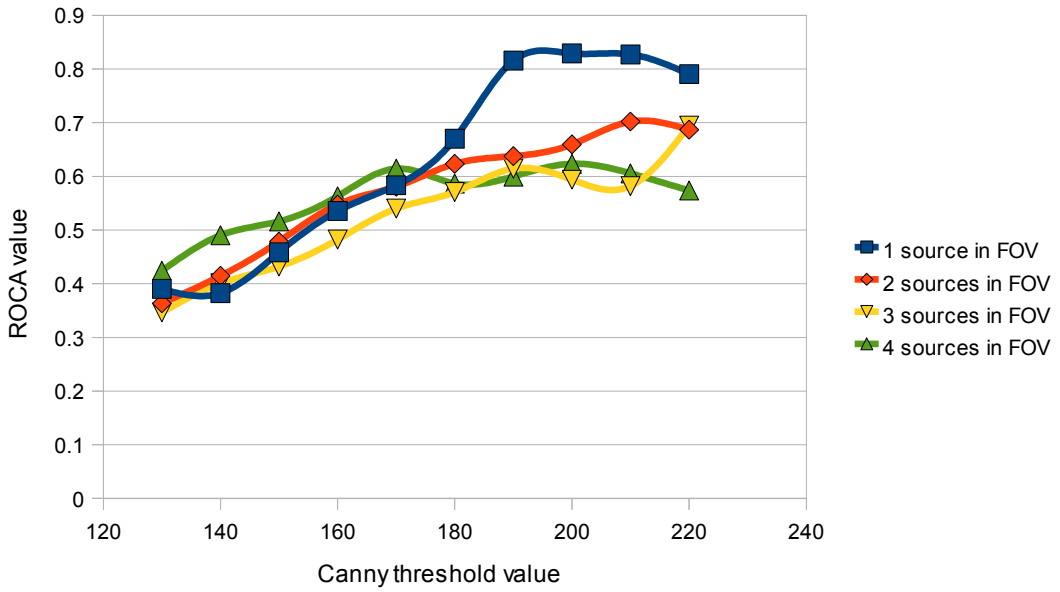


Figure 4.7: Performance of HTED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -20dB are used for the experiment.

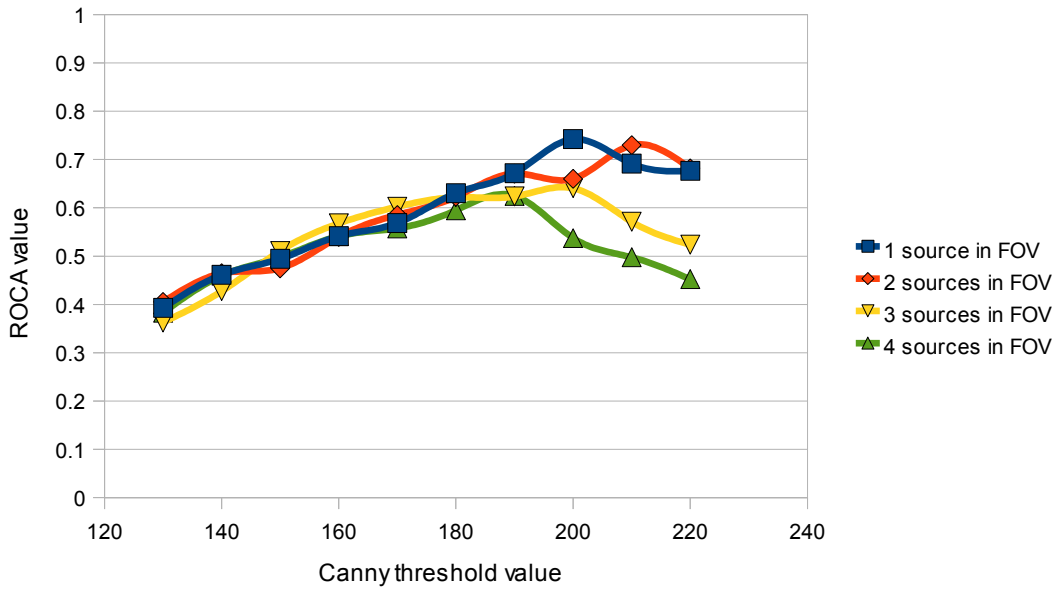


Figure 4.8: Performance of HTED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -15dB are used for the experiment.

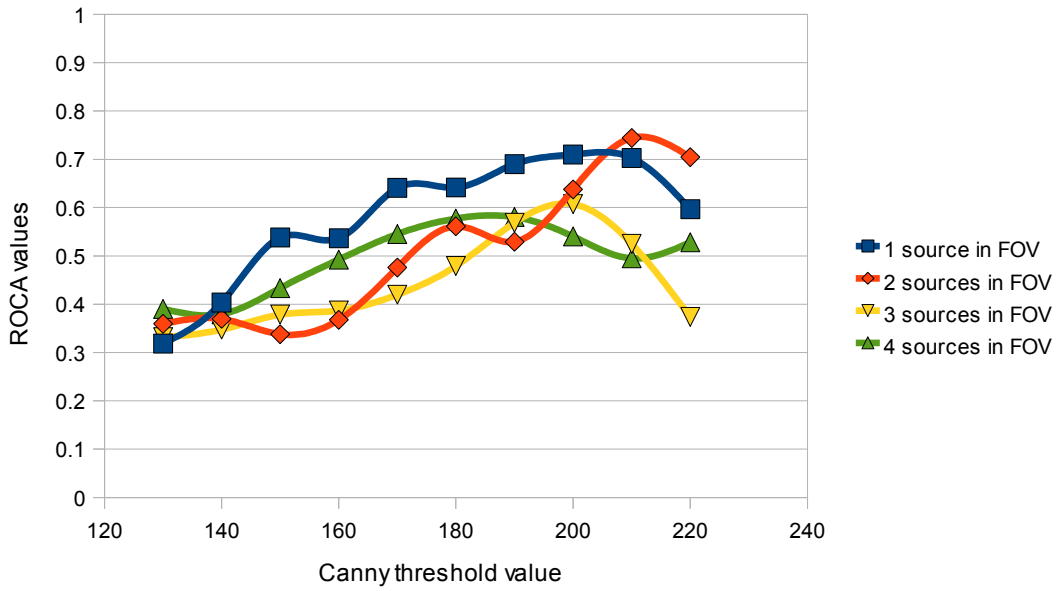


Figure 4.9: Performance of HTED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -10dB are used for the experiment.

To test the algorithm at different noise levels, the experiments are repeated using two coherent noise sources of -20dB and -15dB respectively. The results are shown in Figure 4.7 and 4.8. Table 4.2 summarizes the best case ROCA values obtained using HTED method.

Table 4.2: Result summary for SSD using HTED

Strength of coherent noise sources	Number of sound sources inside FOV	ROCA values for SSD using HTED method
-25dB	1	0.91
	2	0.85
	3	0.75
	4	0.69
-20dB	1	0.83
	2	0.7
	3	0.69
	4	0.62
-15dB	1	0.74
	2	0.73
	3	0.64
	4	0.62
-10dB	1	0.71
	2	0.74
	3	0.61
	4	0.58

4.6. **Conclusion**

The experimental results prove that ROCA values decrease with an increase in coherent noise level in the system. Also, ROCA value decreases with an increase in the number of sound sources. The HTED method has done well in distinguishing between a sound source and a false alarm. The primary drawback of this method is, the algorithm is very complex and computationally intensive.

A simplified algorithm based on Canny edge detection is discussed in next chapter.

Chapter 5. SSD using SRCP- Canny edge detection based method

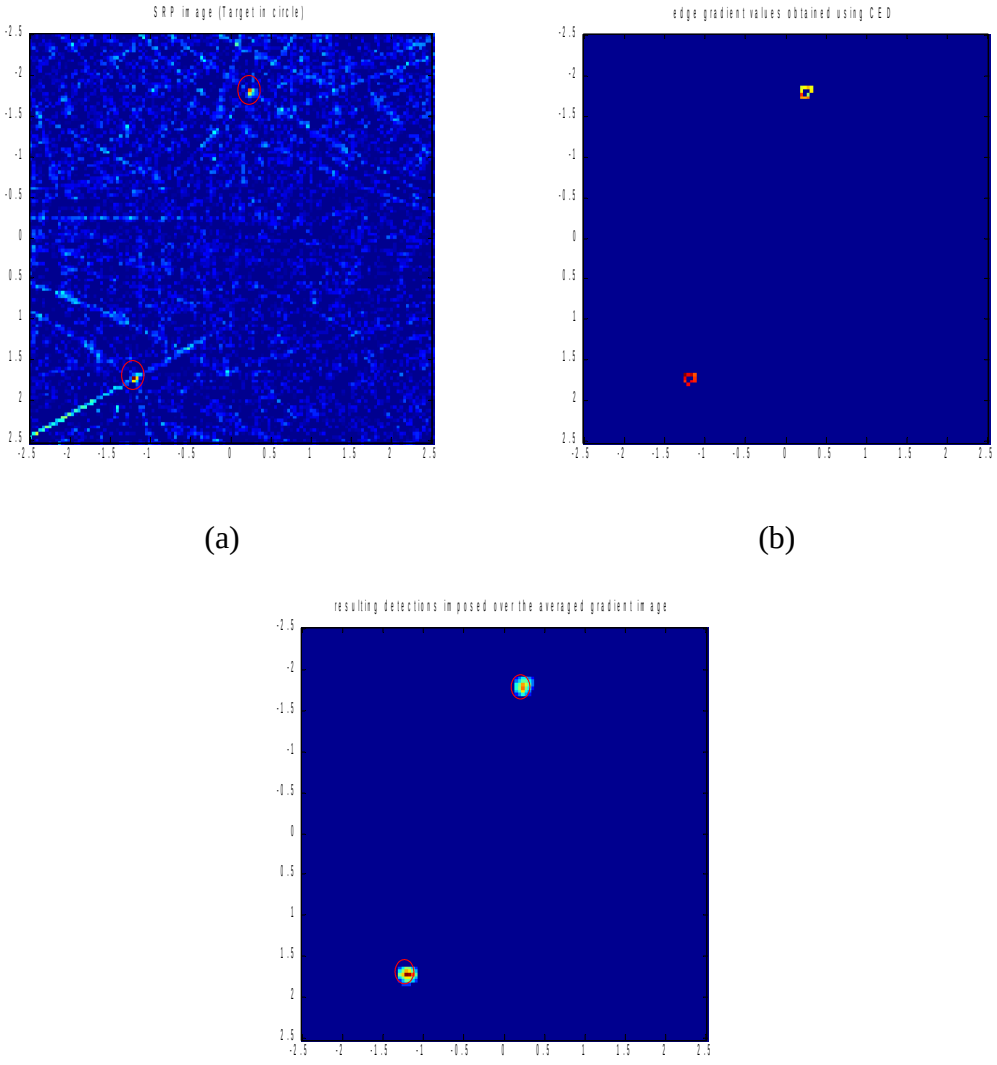
5.1. Introduction

A detailed description of sound source detection algorithm based on Hough Transform based ellipse detection is presented in Chapter 4. This chapter presents a simplified algorithm based on Canny edge detection (SRCP-CED). Section 5.2. introduces the concept of SSD using SRCP-CED method. Results obtained using the method are presented in Section 5.3. A SSD algorithm based on detecting peaks in a SRCP image is described in section 5.4.

5.2. SSD using SRCP-Canny edge detection based technique (SRCP-CED)

Consider a single pixel of considerably higher power compared to the immediate neighborhood pixels. This region when applied with an edge detection algorithm should yield a region of edge pixels around the pixel with high power. Identifying the center of this region will thus give the location of the actual peak pixel. This forms the basic concept behind SRCP-CED method. Canny edge detection when applied to an SRCP image yields edge pixels around strongest peaks in the SRCP image as explained in section 3.2. Finding the center of these edge pixel groups will result in detection of sound source locations.

Figure 5.1.a shows a SRCP image with two sound sources inside FOV at locations marked by red circles. Applying Canny edge detection on Figure 5.1.a yields a gradient magnitude image, shown in Figure 5.1.b. Average filter with a 3*3 window is applied on the gradient magnitude image. This dilates the gradient pixel values and pixels around the center of the edge pixel groups will have maximum magnitudes. Finding local maxima in these dilated groups of edge pixels is equivalent to finding the center of edge pixel groups, which are the probable sound source locations. Figure 5.1.c shows the resulting average gradient magnitude image after applying averaging filter with a 3*3 window. Finding local maxima in the image resulted in two true detections which are imposed in Figure 5.1.c.



(a) (b)

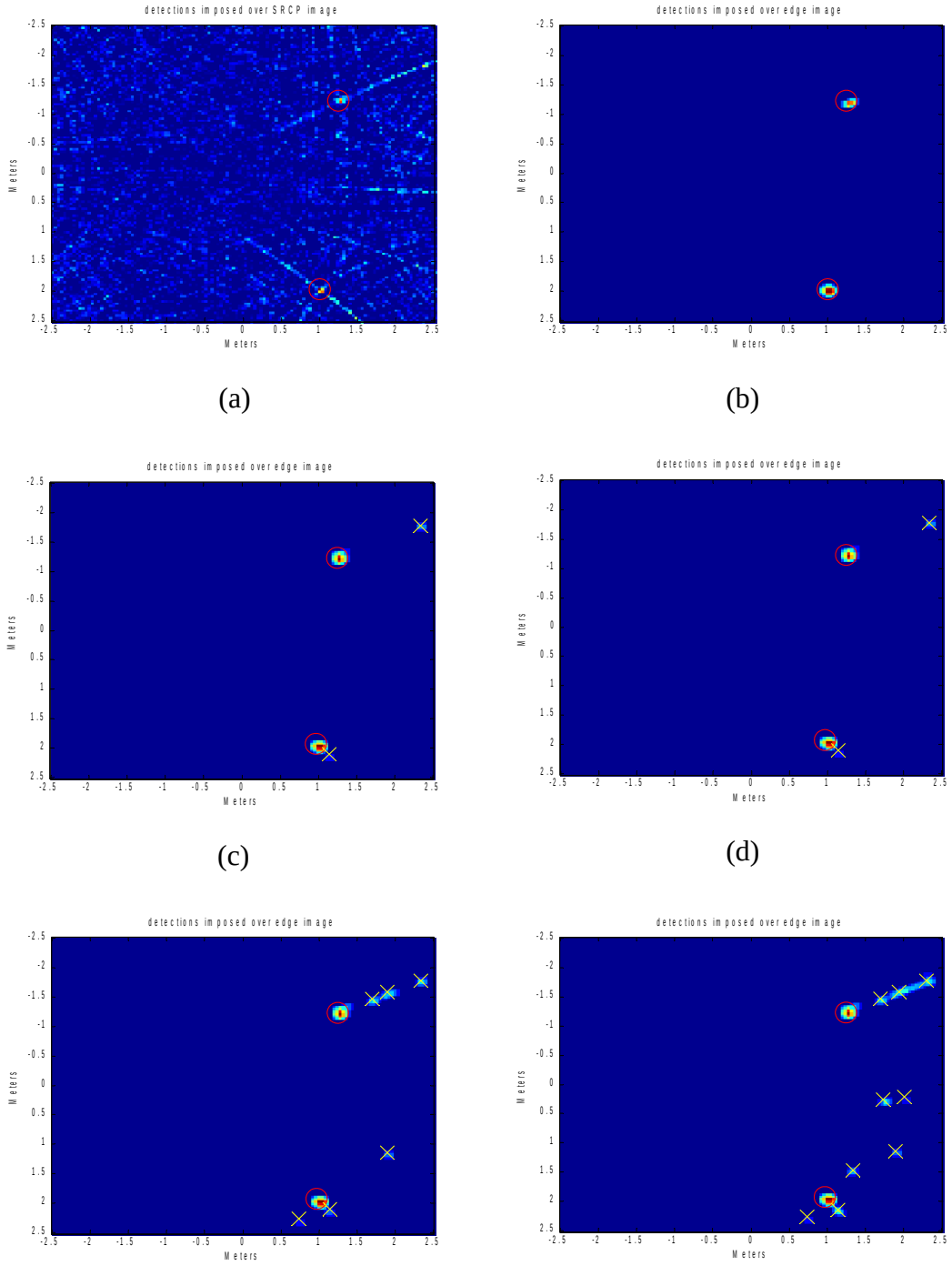
(c)

Figure 5.1: Sound source detection using SRCP-CED method. (a) is the input SRCP image with two sound sources inside the FOV. (b) is the result after applying Canny edge detection on (a). (c) shows detected sound sources using the algorithm.

Figure 5.1.c is a perfect case, where the algorithm is able to detect all the sound sources present inside the FOV and there are no false alarms. However results are not always perfect. Presence of coherent noise in the system significantly degrades the

performance of the algorithm. Coherent noise and reflections induce false peaks in the SRCP image and will result in false alarms.

Canny edge detection uses two thresholds, a lower threshold (t_{low}) and a higher threshold (t_{high}) to deal with the problem of streaking, as explained in section 3.2. Selecting very high values for t_{high} and t_{low} will result in loss of true detections and selecting very low values for t_{high} and t_{low} will result in false alarms. Figure 5.2 shows a situation where decrease in canny threshold values resulted in false alarms.



(a) (b)

(c) (d)

(e) (f)

Figure 5.2: True detections and false alarms using SRCP-CED method. False alarms are added as canny threshold value is lowered. (a) is input SRCP image. (b) Canny threshold of 220 resulted in only true detections. (c),(d) shows added false alarms for a canny threshold of 170 and 160. (e),(f) shows increased false alarms as threshold is lowered to 150 and 140.

5.3. **Results and Discussion**

ROCA analysis explained in section 4.4. is used to analyze the detection performance of the algorithm. Figure 5.3 shows the results obtained using SRCP-CED method. Two coherent noise sources of -25dB are placed on room walls for the experiment. The number of sound sources is varied between 1- 4.

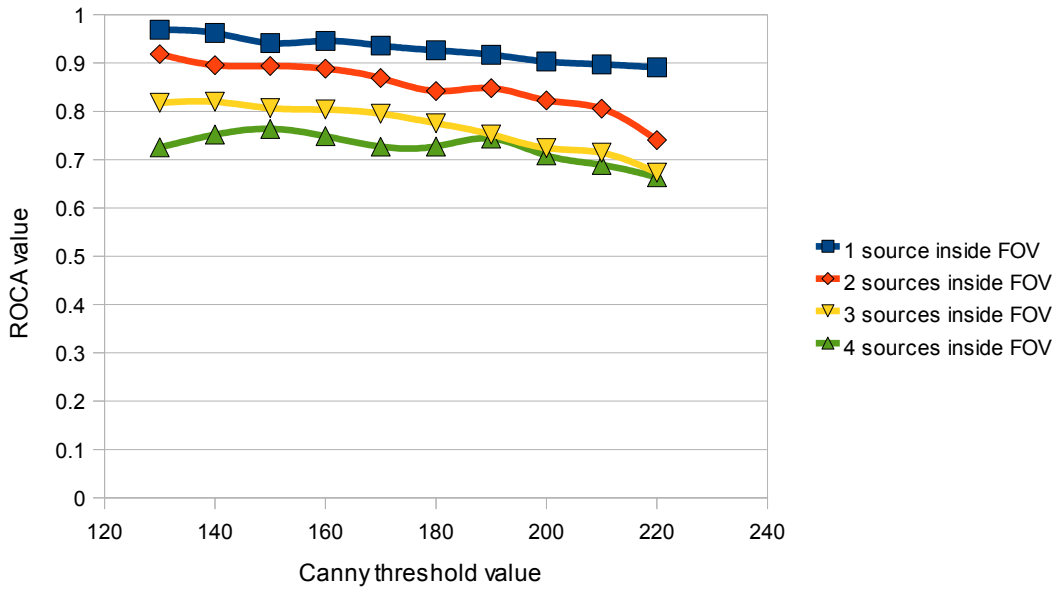


Figure 5.3: Performance of SRCP-CED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -25dB are used for the experiment.

The ROCA value is decreased as the number of sound sources inside the FOV increase because of a overall increase in system noise as explained in previous chapter. However unlike HTED case under similar conditions, the ROCA value for a fixed number of sound sources inside FOV increases as the Canny threshold value decreases. This can be explained as follows. Under low noise conditions sound source peaks are very strong compared to noise peaks. When Canny threshold is decreased, the number of pixels classified as edge pixels increases. False alarms are added as the threshold is decreased and average gradient value around these peaks is low compared to gradient

values around sound source locations. Addition of these false alarms with low average gradient values increases the ROCA value.

Figure 5.4 shows the results obtained when the number of sound sources is varied between 1- 4 and two coherent noise sources of -20dB are placed on the room walls.

Comparing Figure 5.3 and Figure 5.4; with a decrease in Canny threshold value ROCA values remain almost constant or decrease unlike the previous case because of an increase in coherent noise. The difference in strengths of sound source peaks and noise peaks decreases and this effect offsets some of the increase in ROCA value explained in the previous paragraph.

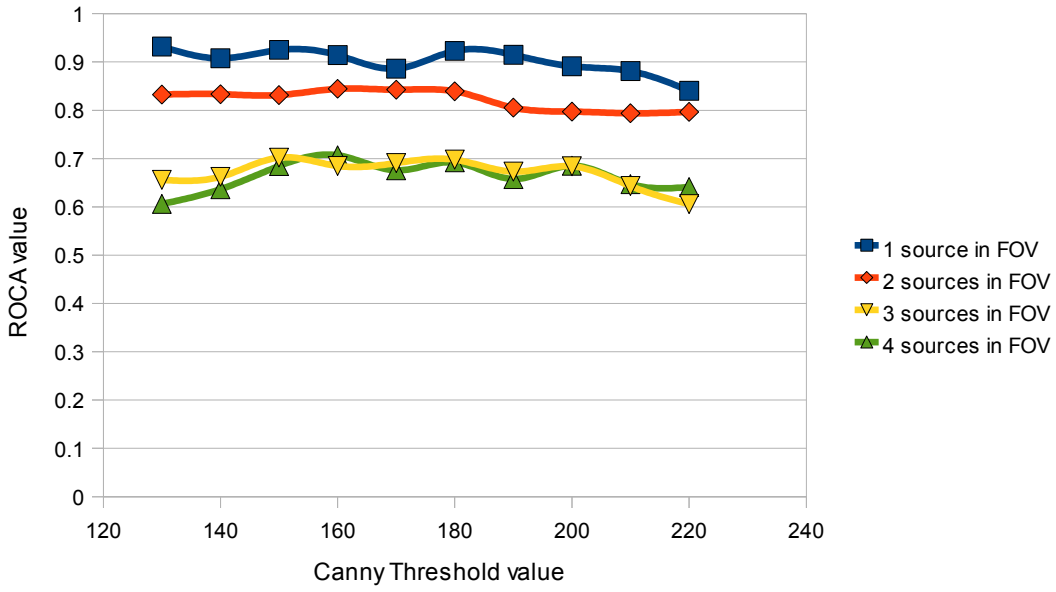


Figure 5.4: Performance of SRCP-CED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -20dB are used for the experiment.

The experiment is repeated for different coherent noise levels, for the purpose of generalizing the results obtained using SRCP-CED method. Figure 5.4 and 5.5 shows the performance of the algorithm using two coherent noise sources of -15dB and -10dB. Number of sound sources is varied between 1-4.

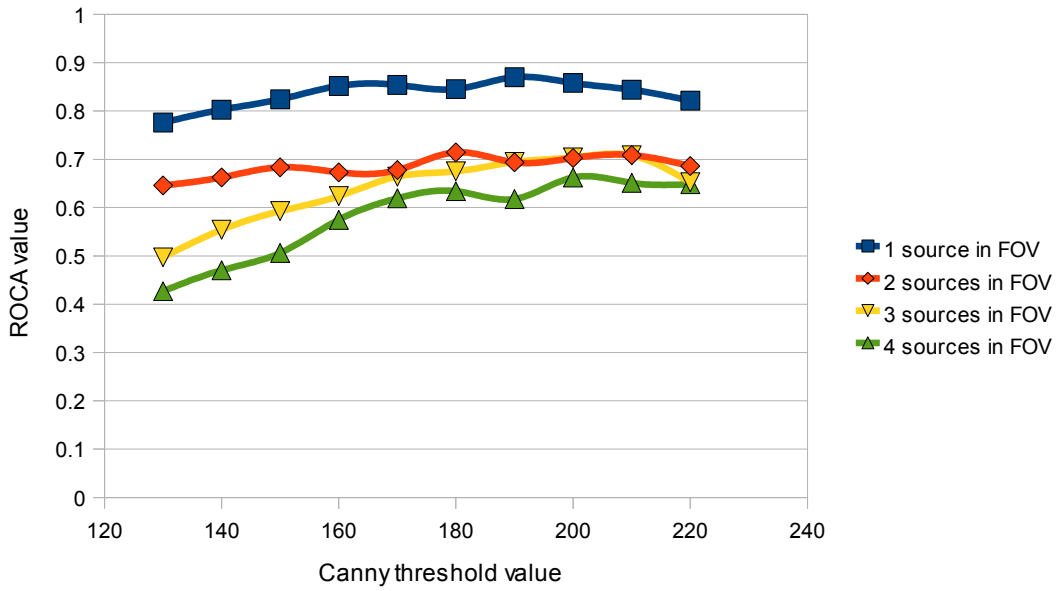


Figure 5.5: Performance of SRCP-CED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -15dB are used for the experiment.

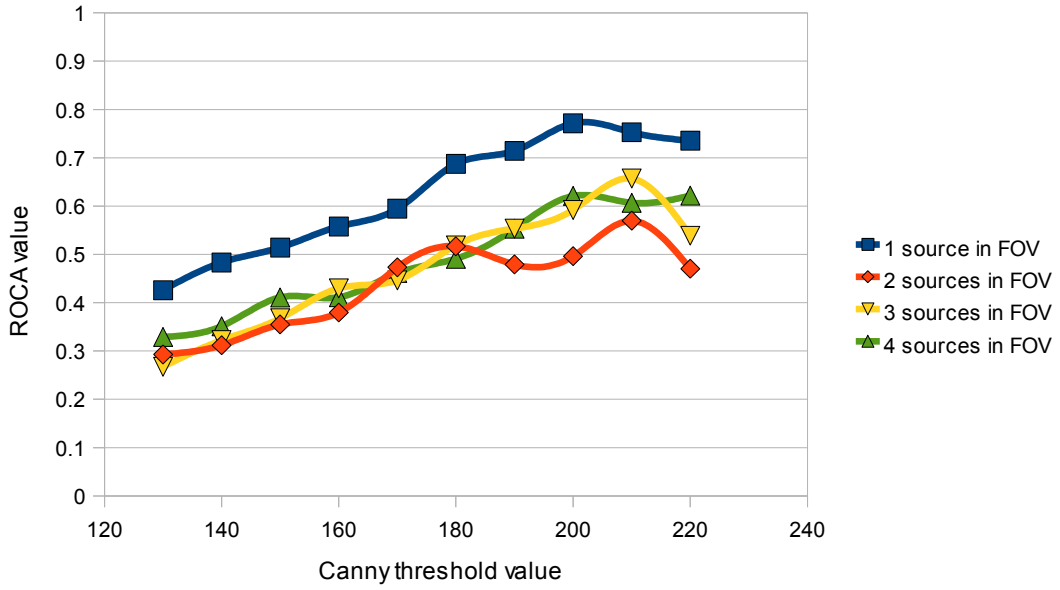


Figure 5.6: Performance of SRCP-CED method while varying the Canny threshold value and number of sound sources present inside the FOV. Two coherent noise sources of -10dB are used for the experiment.

Figures 5.3-5.6 establish a pattern that ROCA values decrease with an increase in coherent noise in the system. Also under similar experimental conditions, ROCA value decreases with an increase in the number of sound sources due to an increase in the overall system noise.

5.4. **SSD using Direct Peak Detection**

This section discusses a third method for sound source detection. The simplest method to find probable sound source locations in an SRCP image is to find the peaks in the SRCP image. Sound source locations are usually associated with large SRCP values, because of the coherent addition of the microphone powers. Detecting peaks in the SRCP image, should thus give an array of probable sound sources. Selecting highest magnitude peaks from the array should give the sound source locations.

Figure 5.7 shows a SRCP image with two sound sources inside the FOV at known locations marked by red circles. Two coherent noise sources are placed on the room walls, indicated by cross marks. Figure 5.8 shows a surface plot of the FOV. Peaks corresponding to sound source locations and probable false alarms are pointed out. Figure 5.9 shows the results of direct peak detection superimposed over the input SRCP image of Figure 5.7. In Figure 5.9 true detections are indicated by red circles and false alarms by cross marks.

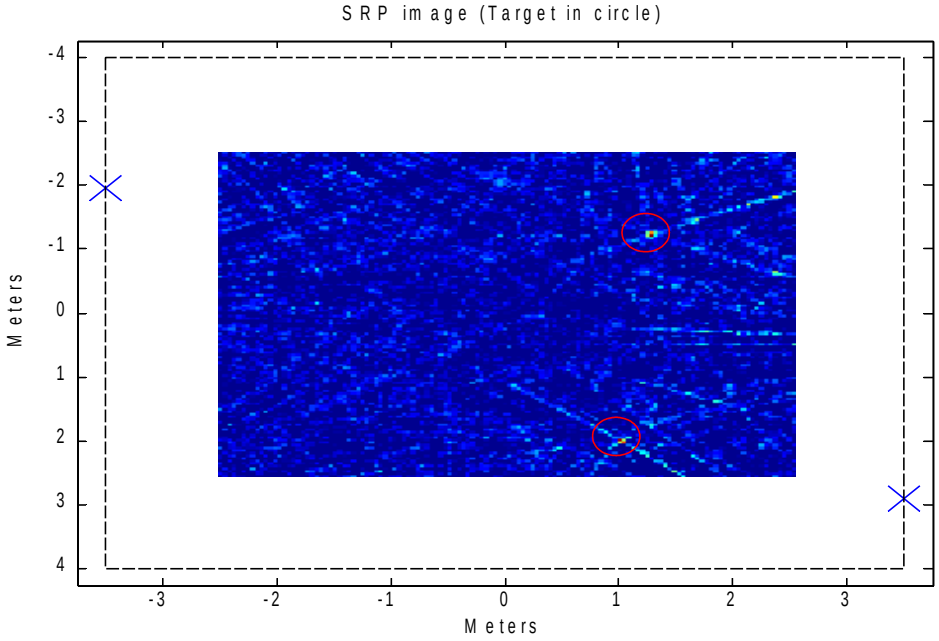


Figure 5.7: SRCP image with sound and coherent noise source locations marked.

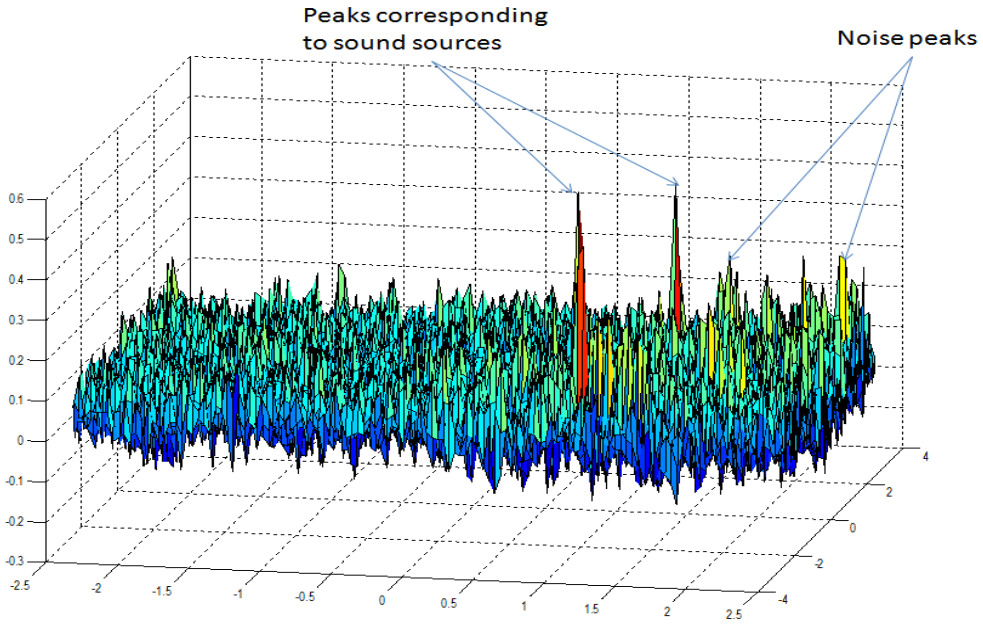


Figure 5.8. Surface plot of SRCP image in Figure 5.7

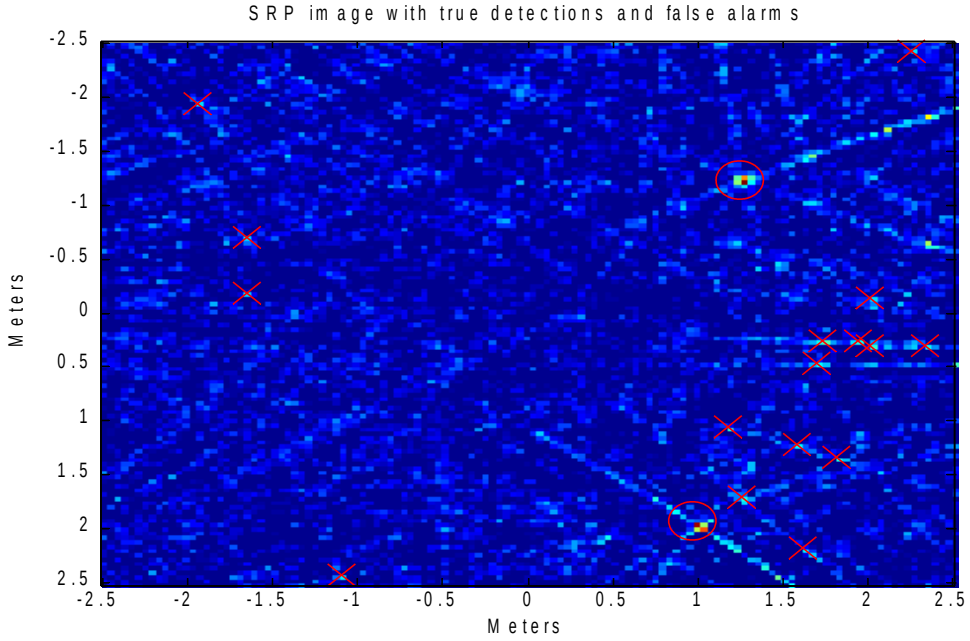


Figure 5.9: Result of applying direct peak detection method. Detected sound sources are marked in red circles and false alarms are marked with crosses.

To analyze the performance of the algorithm known sound source location information is used to separate true detections and false alarms from the array of peak values. The primary drawback of this technique is that there will be a lot of peaks in a SRCP image, and hence a lot of detections in each SRCP image. Considering all these peaks will result in a higher ROC area. To obtain a more practical result a 1-8 ratio is maintained between the target peaks and noise peaks. Thus only the strongest false alarm peaks are considered in the experimental computation of the ROC area values.

The sequence of steps involved in direct peak detection method can be summarized as

1. Find the local maxima in the SRCP image.
2. Identify the true detections and false alarms using known sound source location information.

3. If the number of false alarms exceed 1-8 ratio, then sort the false alarms according to their SRCP value and select the strongest false alarms magnitudes.
4. Compute the ROCA value using true detections and false alarm information.

Table 5.1: Performance of direct peak detection method

Number of sources inside FOV	ROCA values when coherent noise sources of -25dB are used	ROCA values when coherent noise sources of -20dB are used	ROCA values when coherent noise sources of -15dB are used	ROCA values when coherent noise sources of -10dB are used	ROCA values when coherent noise sources of -5dB are used
1 source inside FOV	0.95	0.9	0.63	0.26	0.11
2 sources inside FOV	0.93	0.83	0.51	0.28	0.2
3 sources inside FOV	0.86	0.71	0.49	0.36	0.26
4 sources inside FOV	0.79	0.69	0.46	0.36	0.31

Table 5.1 summarizes the results obtained using direct peak detection method. The experiment is performed varying the number of sound sources inside the FOV and at different noise levels in the system.

The results prove that direct peak detection has a satisfactory performance in determining sound source peaks under very low coherent noise conditions. The ROCA values are high, which suggests that a random sound source peak selected has a higher magnitude compared to a random noise peak inside the FOV.

When SNR of coherent noise sources is increased, the drop in ROCA values for this method is very steep. i.e a random sound source peak selected no longer has a higher magnitude compared to a random noise peak. For a coherent noise source strength of -15dB the ROCA values are around 0.5, which suggests that magnitude of a sound source

peak may or may not be greater than the magnitude of a noise peak. Thus this method can no longer distinguish sound source peaks and noise peaks.

Table 5.2 summarizes the results obtained using the three SSD techniques. The table entries are best case ROCA values obtained during the experiments. Comparing the ROCA values, it can be observed that among the 3 methods, SRCP-CED method performed better for different coherent noise levels and number of sound sources. ROCA values obtained using direct peak detection method show a drastic drop when the overall noise in the system increases. Comparatively HTED method and SRCP-CED method have higher ROCA values even at high noise conditions. Thus these two methods are more robust to coherent noise present in the system and can better detect a sound source present inside the system compared to direct peak detection method.

When SRCP-CED method and HTED methods are compared, SRCP-CED method apart from having better ROCA values, is simple and computationally less expensive. Ellipse fitting outlined in section 3.3. is very complex and intensive.

Table 5.2: Performance comparison of the 3 SSD methods

Strength of coherent noise sources	Number of sound sources inside FOV	ROCA values for SSD using HTED method	ROCA values for SSD using SRCP-CED method	ROCA values for SSD using Direct peak detection method
-25dB	1	0.91	0.97	0.95
	2	0.85	0.92	0.93
	3	0.75	0.82	0.86
	4	0.69	0.76	0.79
-20dB	1	0.83	0.93	0.9
	2	0.7	0.84	0.83
	3	0.69	0.7	0.71
	4	0.62	0.71	0.69
-15dB	1	0.74	0.87	0.63
	2	0.73	0.71	0.51
	3	0.64	0.71	0.49
	4	0.62	0.66	0.46
-10dB	1	0.71	0.77	0.26
	2	0.74	0.57	0.28
	3	0.61	0.66	0.36
	4	0.58	0.62	0.36

5.5. **Conclusion**

This chapter introduced and analyzed two SSD algorithms SRCP-CED method and direct peak detection methods. Simulation outlined in section 4.2. is used to test the performance of the algorithm. It was observed that SRCP-CED method gave the best performance of all of them. The algorithm is very simple and gave better ROCA values compared to the other two techniques.

Chapter 6. Conclusion and Future Work

6.1. Conclusion

This thesis has introduced a sound source detection algorithm based on Hough transform based ellipse detection for detecting sound sources in microphone arrays. The algorithm uses Canny edge detection to pre-screen SRCP-PHAT images. The algorithm based on pairwise fitting of ellipses is robust against breakage of edge pixels. Monte Carlo simulations are carried out and ROCA values are computed to quantify the priority given to a sound source peak over a noise peak. The algorithm is tested, varying the number of sound sources and noise conditions and results prove that HTED method has done well in detecting sound sources. Best performance is achieved for a single sound source inside FOV and best case ROCA values for this case vary between 0.91-0.71. Experimental results prove that performance of the algorithm deteriorates with an increase in the number of sound sources and coherent noise level in the system. For multiple sound sources best case ROCA values vary between 0.85-0.6.

HTED method while being effective at detecting sound sources, is a very complex method. A simplified algorithm, SRCP-CED is also introduced in this thesis. ROCA computations prove that SRCP-CED method has out-performed HTED method. Best case ROCA values for single sound source are in the range of 0.97-0.77. These values drop with an increase in number of sound sources and coherent noise and ROCA values of 0.92-0.6 are obtained for multiple sound sources.

Performance of these algorithms is compared to a straight forward method of detecting sound sources using SRCP-PHAT called direct peak detection method. This method has performed better than other methods under very low noise conditions. However as coherent noise in the system increases noise peak magnitudes become comparable to the magnitudes of sound source peaks and the algorithm has performed very poorly under high noise conditions. Best case ROCA values fall in the range of 0.6-0.3 under noisy conditions.

6.2. **Future Work**

The primary focus of this thesis work is to investigate the performance of an image processing based method to detect sound sources in SRCP-PHAT systems. The simulation specified in [6] is used as the test environment. A perimeter array with eight microphones is used for the experiments. A comprehensive performance evaluation can be achieved by testing the algorithm on a real-time recording using human speakers and practical noise sources. Different microphone array setups, varying the number of microphones and microphone spatial distributions could also be investigated.

PHAT-β is employed to improve sound source detection and a value of 0.75 is assigned to β during the simulation experiments. Decreasing the β value results in an increase in the mainlobe width, which inturn decreases the resolution of the microphone array. A detailed study varying the β value over a range of values can be carried out to find out the effect of β on the algorithm performance.

References

[1] Jean-Marc Valin, Francois Michaud, Jean Rouat & Dominic Letourneau, *Robust sound source Localization using a Microphone Array on a Mobile Robot*, in Proceedings International Conference on Intelligent Robots and Systems, 2003

[2] Joseph H DiBaise, Harvey F Silverman & Michael S Brandstein, *Microphone Arrays Signal Processing Techniques and applications : Robust Localization in Reverberant Rooms*, Springer Publications, pp. 157-180, 2001

[3] Kevin D Donohue, S M Sayed & Jingjing Yu, *Constant False Alarm Rate Sound Source Detection with Distributed Microphones*, University of Kentucky, Lexington, KY, USA, 2009

[4] Kevin D Donohue, Jens Hannemann & Henry G Dietz, *Performance of Phase Transform for Detecting Sound Sources with Microphone Arrays in Reverberant and Noisy Environments*, Signal Processing, Vol. 87, no.1, pp. 1677-1691, Jan. 2007

[5] Kevin D Donohue, A Agrisoni & J Hanneman, *Audio Signal Delay Estimation using Partial Whitening*, Proc. of the IEEE, Southeastcon, pp. 466-471, March 2007

[6] Anand Ramamurthy, Harikrishnan Unnikrishnan & Kevin D Donohue, *Experimental performance Analysis of Sound Source Detection with SRP PHAT*, 2009

[7] Arnold Williams, Rodney Meyer, Peter Pachowicz & George Maksymonko, *A Robust Mine Detection Algorithm for Acoustic and Radar Images*, Science Applications International Corporation, Arlington, VA, 2000

[8] James A Hanley & Barbara J McNeil, *The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve*, Radiology, Vol. 143, pp. 29-36, 1982

[9] M Brandstein & D Ward, *Microphone Arrays: Robust Adaptive Beamforming - signal Processing Techniques and Applications*, Springer publications, 2001

[10] Michael J Peterson & Chris Kyriakakis, *Hybrid algorithm for Robust Real-time source Localization in Reverberant Environments*, IEEE Transactions on Acoustics, Speach and Signal Processing, pp. 1053-1056, 2005

[11] Parham Arabi, *The Fusion of Distributed Microphone Arrays for Sound localization*, EURASIP Journal on Applied Signal Processing, Vol. 4, pp. 338-347, 2003

[12] M Coen, *Design Principles for Intelligent Environments*, Proceedings of Fifth National Conference on Artificial Intelligence, Madison, WI, USA, 1998

[13] L Kinsler, A Frey, A Coppens & J Sanders, *Fundamentals of Acoustics, Third Edition*, John Wiley & sons, 1982

[14] Harikrishnan Unnikrishnan, *Auditory Scene Segmentation using Microphone Array and Auditory Features*, University of Kentucky, Lexington, Ky, USA, 2009

[15] Anand Ramamurthy, *Experimental evaluation of Modified Phase Transform for Sound Source Detection*, University of Kentucky, Lexington, KY,USA, 2007

[16] Kevin D Donohue, Kevin S McReynolds & Anand Ramamurthy, *Sound Source Detection Threshold Estimation using Negative Coherent Power*, IEEE SoutheastCon, Huntsville, Alabama, USA, 2008

[17] Mubarak Shah, *Fundamentals of Computer Vision : Edge Detection*, 1992

[18] Nick Bennett, Robert Burridge & Naoki Saito, *A Method to Detect and Characterize Ellipses using the Hough Transform*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, no. 7, pp. 652-657, July 1999

[19] J Illingworth & J Kittler, *A Survey of the Hough Transform*, Computer Vision, Graphics and Image Processing, Vol. 44, pp. 87-116, October 1988

[20] H K Yuen, J Illingworth & J kittler, *Ellipse Detection using The Hough Transform* , August 1988

Vita

Praveen Reddy Nalavolu was born on February 2, 1985 in Shad Nagar, Andhra Pradesh, India. The author received his Bachelor of Technology (B. Tech.) degree in Electrical and Electronics Engineering from Jawaharlal Nehru Technological University, Hyderabad, India in the year 2006. The author has enrolled for Masters program in Electrical Engineering at University of Kentucky, Lexington in 2007. He has been working at Center for Visualization and Virtual Environments as a Graduate student under Dr. Kevin D. Donohue since February 2009. He is a member of Eta Kappa Nu and UK Solar Car Team since 2008. He received National Merit Scholarship from Central Board of Secondary Education, India and Kentucky Graduate Scholarship from University of Kentucky.