1-1-2011

Locating and extracting acoustic and neural signals

Na Zhu
Wayne State University,

LOCATING AND EXTRACTING ACOUSTIC AND NEURAL SIGNALS

by

NA ZHU

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2012

MAJOR: MECHANICAL ENGINEERING

Approved by:

_____

Advisor Date

_____

_____

_____

DEDICATION

To my parents

ACKNOWLEDGMENTS

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. Thank you:

My Advisor, Dr. Sean F. Wu for providing the encouragement to set high goals and the resources and advice to achieve them;

Dr. Jinsheng Zhang for providing interesting direction and giving me the opportunity to get involved in the fields of bio-mechanical and medicine research activities;

Dr. Emmanuel Ayorinde and Dr. Hao Ying for their encouraging words, thoughtful criticism, and time and attention during busy semesters;

Logesh Kumar Natarajan and Elias Taxakis for their help during the experimental work;

The machine shop and engineering computer center personnel in Wayne State University: Eugene Snowden and Timothy Jones for their help in building the experimental setup and allowing me access to their facilities;

And a very special thanks to my husband, Chenglin Wu, for his continues support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Sound sources localization, extraction, and separation have always been topics of interest in the engineering research for decades, yet they still face significant challenges. There are many cases in practice where the locations of sound sources are highly desired. For example, soldiers in a battle field want to know the directions and distances of explosions and gun shots; police officers monitoring the traffic conditions need to locate accidents as soon as they happen; intelligence agents try to track and trace a suspect or moving vehicle in a crowded area; engineers hope to find the precise location of the noise sources so as to eliminate or reduce noise emission, etc. At the same time, extraction of target source and separation of sources are also important such as in extracting a target voice from overall noisy signals for homeland security, separating specific signals in order to detect abnormalities in monitoring machinery health , analyzing biomedical signal [1] in EEG [2-4] and fMRI [5-7], and removing noise involved in the signals measured on a factory floor for in-line and end-of-line product quality control. In all these applications, we use multiple sensors, usually microphones or electrodes, to measure the overall signals at various positions, process the data, and perform source localization, extraction, and separation in presence of various unknown interfering signals and random background noise.

For sound source localization, three algorithms are presently being used, which include triangulation [8-34], beamforming [35-69], and time reversal (TR) [70-97]. Each method has its pros and cons. Triangulation has the longest history among these three methods, and requires a small number of microphones. The main limitation of triangulation alone is that the accuracy of

source localization is highly dependent on the signal to noise ratio (SNR). Therefore, it is more applicable for the cases [10, 14, 24, 26, 33, 34] where SNR is high. Besides, traditional triangulation usually covers an area of radius up to 1.5 microphone spacing [31], which is not enough for many applications.

Beamforming is another popular method to locate sound sources that has been productized and used in the manufacturing industry. Beamforming is suitable for impulsive, broadband and high frequency sound waves, and can give a general idea of the distribution of sound sources in the target area [51-53, 67, 69]. However, the number of microphones required in beamforming algorithm is relatively high, which makes the device costly. Moreover, the spatial resolution of beamforming is no better than one wavelength of the acoustic signal [62], and its lower frequency limit is determined by the overall diameter of the microphone array. A beamforming system can obtain the directions of the target sound sources but not their ranges. It can find the ranges of the sound sources only when the matched-field method [98] is used simultaneously. TR can accurately locate the sound sources even in the presence of background noises, especially when the microphones surround the target sources [73-75]. However, the TR algorithm relies on spatial scanning, which is time consuming. Therefore, it is impossible for TR to produce real time source localization.

The present dissertation aims at developing an innovative model based methodology for sound sources localization in three-dimensional (3D) space in real time. In particular, the hardware based on the proposed method should be portable, affordable, and easy to use, and the results be displayed in real time. To this end, the number of microphones required in the new method must be minimal. Moreover, it must be able to handle a wide variety of sound sources,

including broad- and narrow-band, random, continuous, and impulsive signals, and cover a large frequency range.

The second aim of this dissertation is to develop new ways to separate sound sources. The existing sound sources separation technologies include Computational Auditory Scene Analysis (CASA) [99-104] and Blind Source Separation (BSS) [2, 3, 5, 105-113]. CASA is mainly used for speech recognition and music segregation [104], while BSS can be used for a much wider ranges of applications than CASA does. A good example of BSS applications is the cocktail party problem [114, 115], where one desires to separate a target signal, for example, the voice of a particular person from the overall signals. Several algorithms have been developed [7, 108, 109, 112, 113, 116-129] for the BSS algorithms, depending on specific types of sound sources, yet none of them are applicable to all types of signals.

In this dissertation, a model based source separation algorithm called point source separation (PSS) is developed, which enables one to separate mixtures of any type of time domain signals that cannot be accomplished by any previous separation methods. Moreover, PSS can work together with the proposed source localization method and become a truly blind sources localization and separation (BSLS), which can separate sound signals as well as locate their precise positions in space. Experimental validations of the proposed source localization method, PSS, and BSLS are conducted in the Acoustics, Vibration, and Noise Control (AVNC) laboratory Machine Shop, auditorium room, and hall way inside the Engineering Development Center.

The final aim of this dissertation is to apply TR algorithms to locate hyper-active neurons inside the brain auditory structure that are directly related to tinnitus perception. Currently, tinnitus is analyzed based on the neural activities measured by using electrodes arrays implanted

inside the brain. The spatial resolution of the measured data is very low and therefore, diagnosis of tinnitus is not very reliable. This dissertation shows that by processing the data measured by electrode arrays using TR algorithm, it is possible to significantly enhance the accuracy and spatial resolution of neural network activities in a very cost-effective manner. The spatial resolution in locating neuron activities can be down to the micrometer level. Validations of using TR algorithm to locate hyper-active neurons are conducted jointly with the Auditory Prosthesis Research Laboratory (APRL), led by Dr. Zhang in the School of Medicine at Wayne State University.

1.2 Literature review

In this section the existing research works reported in the areas of sound source localization, sources separation, and de-noising methodologies as well as the background information on neuron source localization are presented.

1.2.1 Sound sources localization

As mentioned in the Introduction, there are currently three methodologies developed for the sound source localization problem, namely, triangulation, beamforming, and time reversal algorithms. These methodologies are reviewed below.

1.2.1 (a) Triangulation

Triangulation is based on the assumption that the source radiates signals to all directions and sound waves travel along straight lines with a constant speed in a free field [8-10]. Depending on the relative positions of a source and sensor, there is a time delay as the sound signal travels from the source to sensor, which is called time of arrival (TOA). Since the speed of

sound is constant, TOA is equal to the distance between the source and sensor divided by the speed of sound. Consequently, by measuring TOA the distances from the source to the sensor can be calculated. When there are enough sensors, namely, at least N+1 sensors for an N-dimensional space, source localization using triangulation becomes a geometry problem.

Figure 1.1 shows an example of the geometry distribution of the sensors and sources discussed in Tobias' paper [10]. Here a point source is indicated by a red dot in a two dimensional space. Three sensors, namely, Ch.1, 2, and 3 as indicated by blue, green, and black squares in the figure, respectively, are located on the same plane as that of the source. Suppose that TOAs from the source to three sensors are specified. The distances from the source to the sensors can be calculated by multiplying them to the speed of sound. However, as there is no knowledge of the position of the source a priori, the number of possible locations of the source is infinite. For example, for the sensor marked as Ch.1, the source distance r_1 calculated based on TOA can be anywhere along the blue circle of radius r_1 (see Figure 1.1). Similarly, for sensors marked as Ch.2 and 3, the source distances r_2 and r_3 calculated based on TOA can be anywhere along the green and black circles of radii r_2 and r_3, respectively. These three circles only share one particular point on the plane, which is the location of the sound source. In this way, the precise location of the source can be determined.

In many cases, however, TOA cannot be obtained directly. Thus the time difference of arrival (TDOA) is used to solve the localization problem. Instead of measuring the time delay from the source to any sensor, the time delay between one sensor to another is measured. Since the distances among individual sensors are specified a priori, the source position can be determined in terms of the sensor positions and relative TDOAs. For example, in Figure 1.1, if Ch.1 is considered as the reference, then r_2 and r_3 can be rewritten as $r_2=r_1+c\Delta t_{12}$ and

$r_3=r_1+c\Delta t_{13}$, where c is the speed of sound, Δt_{12} and Δt_{13} indicate the TDOAs from Ch.2 to Ch.1 and Ch,3 to Ch.1, respectively. The geometric position of the source can now be determined once Δt_{12} and Δt_{13} are measured and the relative positions of Ch.1, Ch.2, and Ch.3 are specified. There are some improved triangulation algorithms that attempt to define the direction of arrival (DOA) first and find the crossing point of multiple directions [26, 27], while others uses more sensors than the minimum number required together with complicate equations to improve the accuracy of localization results [32, 33].

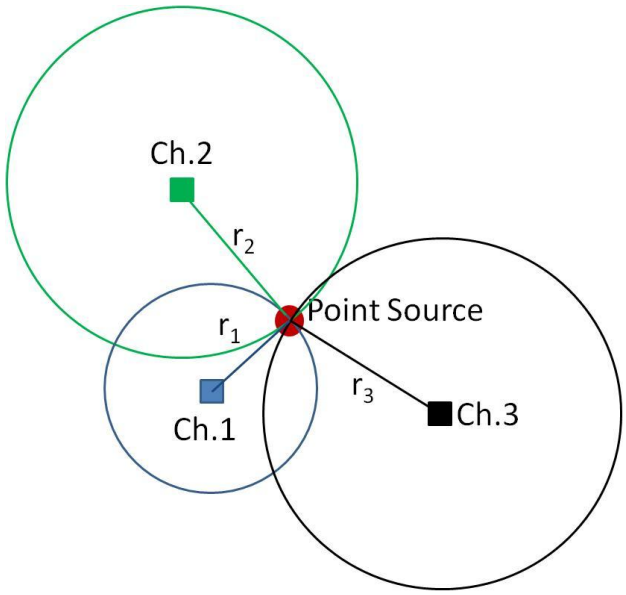


Figure 1.1 Source localization by using the triangulation method. The red dot indicates the source position, and the blue, green, and black squares show three sensors positions, respectively. Blue, green, and black circles indicate the possible source locations with respect to three sensors, and the intersection of these circles is the correct source position.

Generally speaking, the triangulation method is simple and easy to understand. The number of microphones required in this method is relatively small. Mathematically, at least N+1 sensors are required to locate a sound source in an N-dimensional space. For example, if the source is restricted on a two dimensional (2D) plane, then three sensors on the same plane as the

source are needed to locate the source; and if the source is in three dimensional (3D) space, four sensors that are not on the same plane are needed to determine the source location.

The key to a successful source localization using triangulation is to find the correct TOAs or TDOAs. Errors in measuring TOAs and TDOAs can significantly affect the accuracy of source [130]. The presences of sound reflections and reverberation, and interferences by background noises can cause significant errors in the estimation of TOAs and TDOAs. A lot of research has been conducted on developing various algorithms to calculate the time delay [130-154], which aims at reducing the errors in estimating TOA or TDOA in a non-ideal environment. One of the algorithms is to find the time delay by using the criterion of maximum-likelihood (ML) [142, 143, 146, 151], which is effective especially when the sound signals are impulsive. Knapp and Carter suggested the use of generalized cross-correlation (GCC) algorithm [131] to estimate time delays. This concept is widely used today, and various algorithms [135, 139, 141, 145, 148, 153] based on GCC have been developed.

Although the accuracy in estimating TOA and TDOA can be improved by these algorithms, the accuracy in source localization using triangulation alone is still highly dependent on the signal to noise ratio (SNR), which is greatly affected by the test environment. In general, triangulation is effective with a high SNR. Thus in practice, it is usually used in the environments where high SNR can be achieved or guaranteed. For example, triangulation is often used in detecting acoustic emission at high frequencies [10, 34, 97] and ultrasonic sound localization [26, 70-72]. This is because the background noise levels at high frequencies are usually very low, so SNR is very high, making triangulation an effective source localization method. Triangulation is also used to locate the sources that emit impulsive signals because the frequency contents of impulses are very high[31].Error analysis [11, 12, 16] for triangulation

algorithm has been conducted and their results show that the source detection range using traditional triangulation algorithm is quite limited.

1.2.1 (b) Beamforming

Beamforming is based on the delay and sum technique to locate sound sources [62]. The number of microphones required for beamforming is much higher than that in triangulation, and they are usually mounted on a 2D plane. The underlying principle of beamforming is to adjust the time delays in individual microphone channels systematically until they are all in phase. This is equivalent to rotating the microphone array until the incident sound wave arrives at all microphones simultaneously. When this happens, the microphone array is facing the source. So beamforming alone can only determine the bearing of a source, but not its range. Most beamforming employs a planar microphone array for convenience, though Meyer has tested a set of microphones located in a circular shape [47]. Beamforming is commonly used in industry to locate undesirable machine noise, for example, locating noise leakage from a vehicle [51-53]. It has also been used for underwater acoustics [63]. Beamforming has several limitations, as Dougherty described, its spatial resolution is no better than one wavelength of the sound wave emitted by the source [62]. This means that beamforming cannot be used to locate sound sources emitting low frequency sounds. Moreover, since beamforming requires many sensors, it is usually very costly. Note that beamforming can be used to extract source information with only two microphones. Its extraction is based on an assumption that the source is exactly in front of the microphone array, and all signals from other directions are treated as noise. In other word, it has very strict requirements on the microphone configuration and the source direction. This technology has been tested on hearing aid equipment to extract the target source's information and to de-noise the surrounding noise [37, 38, 43]. Kompis and Dillier have also pointed out that

adding adaptive beamforming processing technology to the directional microphones can further improve the de-noising effect [39, 45, 46].

1.2.1 (c) Time reversal

TR is a computation algorithm that does not need to consider the time delay or phases of the signals. The procedure of this method is very simple: reverse the signals measured at each sensor, play it back, and combine the reversed signals at every point in space. In this way, a peak is automatically formed at the source location. The TR method can achieve very high spatial resolution, depending on the sampling rate and scanning step size. Moreover, it is not restricted by SNR and the impacts of a test environment. Accordingly, it is very advantageous to use the TR algorithm to locate a source, especially when SNR is low. TR is effective when the actual sound sources are surrounded by measurement sensors. When the sources are not enclosed by the sensors, errors in source localization can be very large, although the bearing of the source may still be correct [71].

The main disadvantage of TR is that numerical computations are much more time consuming than the other two methodologies. This is because TR needs to scan every point in space to find the source. Nevertheless, TR is still very popular in various applications such as the locating sources underwater [78] or non-homogeneous or layered media [76, 82, 83, 85, 90, 92, 97], detection of electromagnetic, ultrasonic waves, and telecommunication [88, 91, 92]. Moreover, TR is not affected by the interferences of background noise, and the reverberation in the environment can be modeled in its algorithm, it is also widely used in noise and complex environment [80, 81, 89].

1.2.2 Sound sources separation methodologies

Currently, there are several methodologies for the sound separation proposes. One of them is CASA. It mimics the human auditory system and is popular in speech recognition and music segregation [101]. Another family of technologies, known as BSS, enables one to make blind guesses of the original signals with only the mixed signals on hand [105-107]. Several algorithms have been developed for BSS such as Principal Components Analysis (PCA) [129, 155], Independent Component Analysis (ICA) [2, 5], Non-negative Matrix Factorization (NMF) [113, 119, 124, 125, 128], Stationary Subspace Analysis (SSA) [126], etc. These methods employ specified properties of the target signals as the conditions to get the solutions, thus they are only applicable to some particular types of sounds. What is more, all the algorithms mentioned above are limited to separating and estimating the original signals in time domain, but none of them can provide a solution in space domain, namely, distinguish the relative contributions from individual sources located at different positions in space.

1.2.3 Signal decomposition and de-noising

Signal decomposition and de-noising is another challenging topic. Some popular methodologies for signal decomposition and de-noising include the short-time Fourier transforms (STFT) [156, 157] and Wavelet Transform (WT) [158-160]. The basic idea of STFT is to analyze transient or non-stationary signals such as human voices over a very short time period, over which the signals are relatively stationary, and then assemble all individual analyses together to display the time variance of the signals. The concept of STFT is realized by applying a window function on the original signal, thus extracting a short period of time of the signals and then performing the Fourier transform. In this way, STFT provides the frequency contents of the

signals over time history. Note that the resolution in either time or frequency domain may be compromised. In other words, applying a very short window may produce very high resolution in time, but very low resolution in frequency. On the other hand, applying a long window tends to improve the resolution in frequency, but at the cost of decreasing resolution in time.

Unlike STFT, WT enables one to decompose a target signal with vert high accuracy and resolution in both time and frequency domains. Such a goal is accomplished by using a wavelet that enables one to use different scales to best approximate different features of a target signal in both time and frequency domains. There are various types of mother wavelets such as Haar, Mexican hat, Daubechies, etc., from which all other wavelets can be generated by shifting and scaling the mother wavelets. In this way, WT can handle some special cases that STFT cannot, such as analyzing a sharp peak in the time domain signal.

However, both STFT and WT can be used to de-noise a target signal by adding thresholds on the transform coefficients, so that the undesired components in the frequency domain can be removed.

1.2.4 Application to neuron source localization

Tinnitus (i.e., ringing in the ear) is a phantom sound that occurs in the absence of external acoustic stimulations, and is a highly prevalent health problem that affects roughly 10 – 15% of the adult population [161] and 33% of elderly population [162]. Chronic tinnitus has a significant adverse impact on patient quality of life [163-165]. Due to its complex mechanism and origin, tinnitus is still not fully understood and its treatment strategies are very limited. A major contributing factor is the lack of effective tools to measure and analyze the neural network signals such as spontaneous or stimulus-driven signals, which are often contaminated by a variety of interfering signals.

Current technologies for measuring the neural network activities in the brain auditory system are mainly based on multichannel electrode array, together with some rudimentary signal processing techniques such as averaging, low-, high- and band-pass filters, pruning, and thresholding. These methods are *ad hoc* in nature and ineffective in suppressing interfering signals, and background noise. Note that all direct measurements depict the neural activities at electrode tip positions, and the data gathered at individual electrodes are susceptible to interference of the neural activities in the entire neighborhood. Moreover, the neural network activities are contaminated by interfering physiological signals resulting from breathing and blood circulation and non-biological signals from electronic signals instruments and background acoustic signals. These erroneous data may distort the measured pictures of the neural network activities in the brain auditory structure and lead to incorrect conclusions.

For example, when a 32-channel electrode array is used to measure the neural network activities, only the time histories at the tips of individual electrodes are recorded. Since the actual neural network activities may not coincide with the tips of electrodes and since no information is available in between the tips of electrodes, the spatial resolution of localization of the neural activities is limited by the spacing between individual electrodes. Oftentimes, interpolation is used to connect these discrete points to yield tonotopic maps [166-168] to show neural network activities on certain frequency bands. The difficulties with this approach are that: 1) the spatial resolution of such a tonotopic map is limited by shank and electrode spacing; 2) the high intensities measured at any electrode may not represent the true location of an active neuron because the measured signals may be susceptible to interferences by other neurons in the neighborhood; and 3) the input data can be contaminated by a variety of physiology and non-physiology signals. Therefore, one cannot rely on the directly measured data to analyze the

etiology of tinnitus, and a more accurate and reliable methodology is needed to gain a better understanding the fundamental mechanisms underlying tinnitus.

1.3 Goals and objectives of the dissertation

The literature review presented in Section 1.2 indicates that there is a great need to have new technologies that will enable one to locate arbitrary sound sources in 3D space in real time, to extract target information from directly measured data, to separate and locate sound sources in 3D space simultaneously, and to develop more effective methods to diagnose and treat tinnitus. This dissertation aims at addressing these issues by:

1) Developing acoustic modeling based method to locate arbitrarily time-dependent acoustic in 3D space in real time;
2) Developing PSS method to separate target signals from any mixed data, given the source locations;
3) Developing BSLS method to separate source information and locate sources in 3D space simultaneously; and
4) Applying TR algorithm to pinpoint the exact locations of hyper-active neural activities inside the brain auditory structure that are directly correlated to the tinnitus perception.

1.4 Significance and Impact of the dissertation

A successful completion of this dissertation is expected to have significant impacts in a number of fields ranging from the manufacturing industries, homeland security, defense industry, medical diagnosis and applications, etc. Specifically, the proposed technologies will help the Intelligence Community to gather and analyze intelligence, the Homeland Security to monitor

target suspects, the companies that want to identify noise sources and conduct in-line and end-of-line products quality control testing. In particular, the proposed technology for locating hyper-active neurons inside the brain auditory structure may lead to a paradigm shift in diagnosing and treating tinnitus and other neurological disorders.

CHAPTER 2

SOUND SOURCE LOCALIZATION

2.1 Acoustic model based triangulation

Unlike the traditional triangulation, the sources localization method in this dissertation is based on acoustic modeling. Specifically, it assumes that sound is generated by a point source in a free field, and the amplitude of the sound wave follows the law of spherical spreading [171]. Accordingly, TDOAs of the sound signals at the measurement points depend on the relative distances between the source and microphones, and the amplitude of the sound decays inversely proportional to the distance. For the simplest case that only one source is considered, the sound pressure at measurement point can be expressed as:

$$p = \frac{1}{r} f\left(t - c^{-1} r, \theta, \phi\right) \tag{2.1}$$

where the letter p indicates the sound pressure at time t at geometric location (r, θ, ϕ) in polar coordinates, r is the distance between the measurement and the source location, θ and ϕ are the polar and azimuthal angle of the measurement position with regard to the source. The letter c indicates the speed of sound, which is related to the sound travel media and the temperature of the environment. The supposed media in this dissertation is air, thus according to the Laplace's adiabatic assumption for idea gas [172], the speed of sound c can be calculated by the following equation:

$$c = 331 + 0.6T_C \tag{2.2}$$

where T_C is the value of temperature in Celsius.

Assuming that M microphones are employed in the prototype device of sound source localization, one can derive a general equation that governs the distance from the source to the microphone in terms of TOA as follows:

$$r_{is} = c\Delta t_{is} \tag{2.3}$$

where the subscript i indicates the i^{th} microphone, s indicates the source, and r_{is} is the distance between the i^{th} microphone and the sound source. Δt_{is} is the TOA of measurement due to the time concern of the signal traveling in the media.

Similarly, TOA of the j^{th} microphone can be written as:

$$r_{js} = c\Delta t_{js} \tag{2.4}$$

Using the Equation (2.4) minus Equation (2.3), thus

$$r_{js} - r_{is} = c\left(\Delta t_{js} - \Delta t_{is}\right) \tag{2.5}$$

This can be further simplified as:

$$r_{js} - r_{is} = c\Delta t_{ji} \tag{2.6}$$

In Equation (2.6) Δt_{ji} is TDOA between the j^{th} and i^{th} microphone and can be obtained by analyzing the measured signals at these two microphones. The details of the estimation of TDOA are discussed in the next section. The distances r_{js} and r_{is} are in terms of the locations of the sources and the microphones, while the microphone locations are known in advance. Therefore the only unknown in Equation (2.6) is the location of the source.

To determine the position of the point source in 3D space, the values of (x, y, z) in Cartesian coordinates, which are three unknowns, should be calculated. Three equations in the format of Equation (2.6) are required to get a unique solution, where at least four microphones should be employed. As introduced in Chapter 1, this dissertation aims at find the sound source

localization with minimum number of microphones, therefore the number of microphone *M* is firstly set as four in the basic model. The microphone setup is shown in Figure (2.1), the red dot shows the location of the point source, and the four blue rectangles indicates the positions of microphones.

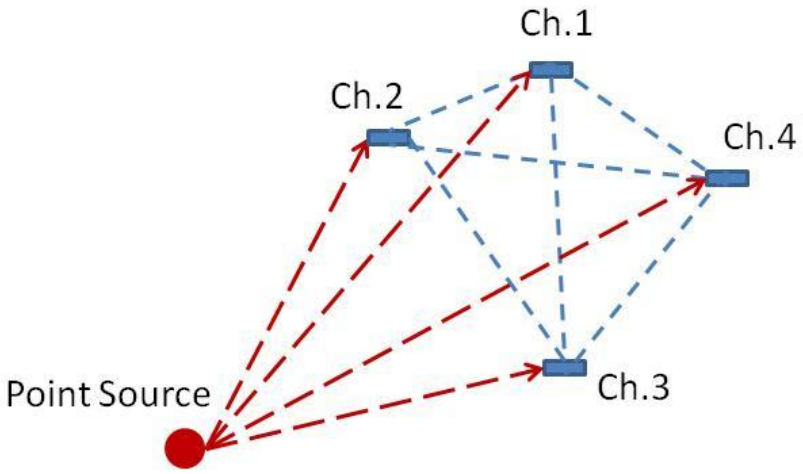


Figure 2.1 Basic modeling of localization method. The red dot indicates the position of point source in space. Four blue rectangles, Channel # 1, 2, 3, and 4, show the positions of microphones.

Note that one can place the microphones anywhere in 3D space as long as not on the same plane. That is because the algorithm of this acoustic model based triangulation has no restrict on the position of measurement; however, if all the four microphones are on the same plane, the microphone setup reduces to a two-dimensional setup, thus the system is lack of the information of the third dimension in space and it cannot successfully give the 3D sound source location. With the microphone setup defined in Figure 2.1, the equation set of solving the location of the source can be written as:

$$\begin{cases} r_{2s} - r_{1s} = c\Delta t_{21} \\ r_{3s} - r_{1s} = c\Delta t_{31} \\ r_{4s} - r_{1s} = c\Delta t_{41} \end{cases} \tag{2.7}$$

It obvious that microphone Channel #1 is involved in all these three equations in Equation (2.7). In this case, it is considered as the reference microphone, and the three equations are built up based on TDOA of microphone Channel #2, 3, and 4 to Channel #1. As the microphone positions are known in advance when the positions of microphone determined, inserting the value of the microphones position into Equation (2.7), the equation set can be rewritten as:

$$\begin{cases} \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2 + (z_2 - z_s)^2} - \sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2 + (z_1 - z_s)^2} = c\Delta t_{21} \\ \sqrt{(x_3 - x_s)^2 + (y_3 - y_s)^2 + (z_3 - z_s)^2} - \sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2 + (z_1 - z_s)^2} = c\Delta t_{31} \\ \sqrt{(x_4 - x_s)^2 + (y_4 - y_s)^2 + (z_4 - z_s)^2} - \sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2 + (z_1 - z_s)^2} = c\Delta t_{41} \end{cases} \quad (2.8)$$

The number subscribes after *x*, *y*, and *z* indicates the index number of the microphones, while the subscribe *s* indicates the source. Solving Equation (2.8), one can get one pair of unique solutions of the (*x, y, z*) value of the sound source. In some of the cases, the two solutions are repeated roots. However, in most of the cases, one of the solutions is the correct answer while the other is a ghost image. Selection of the correct one from the pair of solutions is necessary. The method of selection uses the characteristics of the decaying amplitude of the sound sources, and will be discussed in the next section.

2.2 3D sound source localization algorithm using a four-microphone set

As shown in above section, the basic acoustic model of the newly developed source localization method is simply and easy to understand. However, besides the estimation of TDOAs and the solving of Equation (2.8), some pre-processing and post-analyzing procedures are needed in the methodology, and a flowchart of these procedures is shown in Figure 2.2.

The localization of the microphones are supposed to be known in advance, in the form of (x, y, z) in Cartesian coordinates. The measured time-domain signals at four microphones are firstly be pre-processed thus the value of SNR are increased, followed by estimating of TDOA of the microphones to the reference one, and then contribute all the known values into Equation (2.8) to get a pair of source localization solution. Finally, the correct result of the source location is selected with the help of the amplitude of the measured signals.

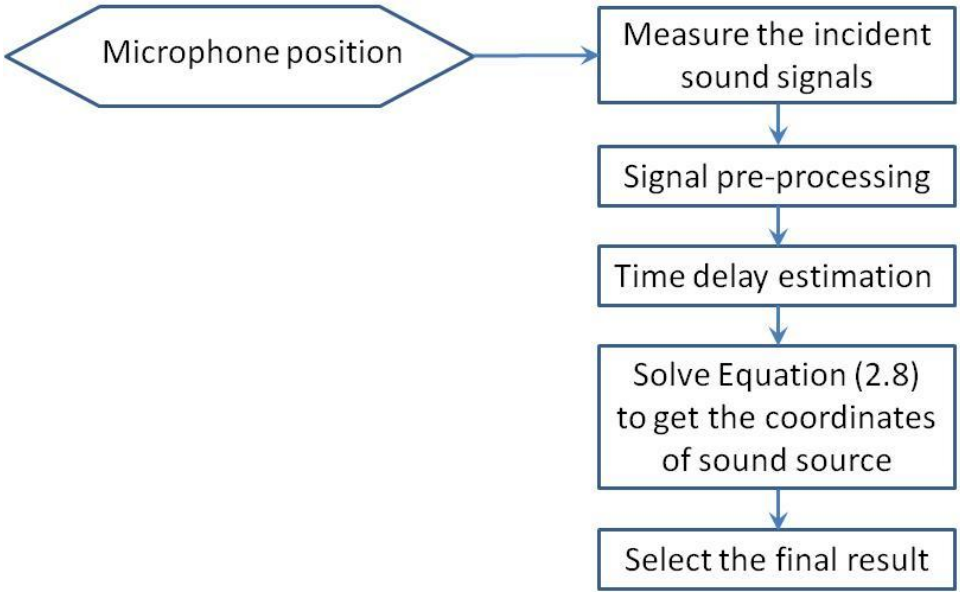


Figure 2.2 Flowchart of the basic sound source localization method.

2.2.1 Signal pre-processing

The first step after the measurement of incident signals is to apply pre-processing to the data at all channels. The goal of signal pre-processing is to increase SNR of the input data, therefore increase the accuracy of the results. As discussed above in the literature review, TDOA estimation is highly depends on the value SNR of the measurements. Therefore, increasing SNR with signal pre-processing methodologies can improve the accuracy of TDOA

estimation, and further has a positive impact on the final sound source localization results. Two methodologies are utilized in this dissertation, namely, windowing and filtering.

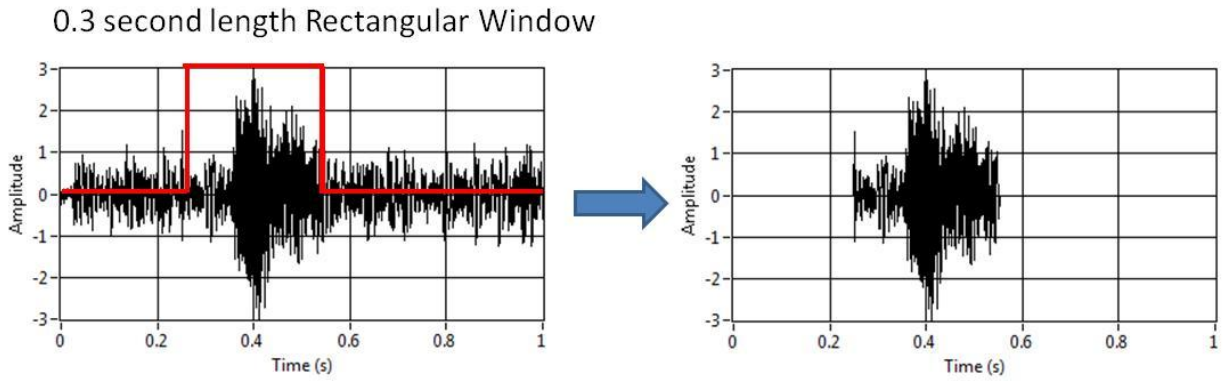
2.2.1 (a) Windowing signal pre-processing

The windowing signal pre-processing is utilized when the target sound sources are non-stationary sources, especially when the amplitude of the target source fluctuating significantly or a peak values happens during a short period of time, for example, impulse signals. In these cases, we assume that the measured signals are the mixtures of the target transient signal and long lasting background noises with relative small amplitude. Therefore the measurements at some of the time segments consist of no contribution of the target signal, but only the backgrounds. These segments are interfering signals, and should be eliminated by window function.

Although the time domain signals at microphones have time delays among each other, the window should be synchronized, and as Channel #1 is usually used as the reference microphone, a rectangular window centered at the peak index of Channel #1 is applied to the time domain signals at all channels. After the windowing processing, the loudest part in the time domain signals are remained, and all the other parts are zeroed. In this way, SNR of the input measurement signals can be effectively increased.

Figure 2.3 (a) shows an example of signal at measurement point which is generated numerically. It the mixture of a man' shouting lasting only about 0.25 second, which is performing as the target signal, and a stationary battle background. The total measurement length is unity second, and SNR in Figure 2.3 (a) is zero, which means that if counting along the whole unity second, the contributions of the target and the background signals are the same. During the windowing procedure, the system automatically detects that the highest peak is at 0.4 second, and then the window function is applied to the input signal centered at the 0.4 second. The

window length is pre-determined by the program and is set as 0.3 second. Figure 2.3 (b) shows the signal after the windowing signal pre-processing. Only the highest part of the signal is remained and all the other indexes are zeroed. SNR of the signal after windowing is increased to 5.0 dB.



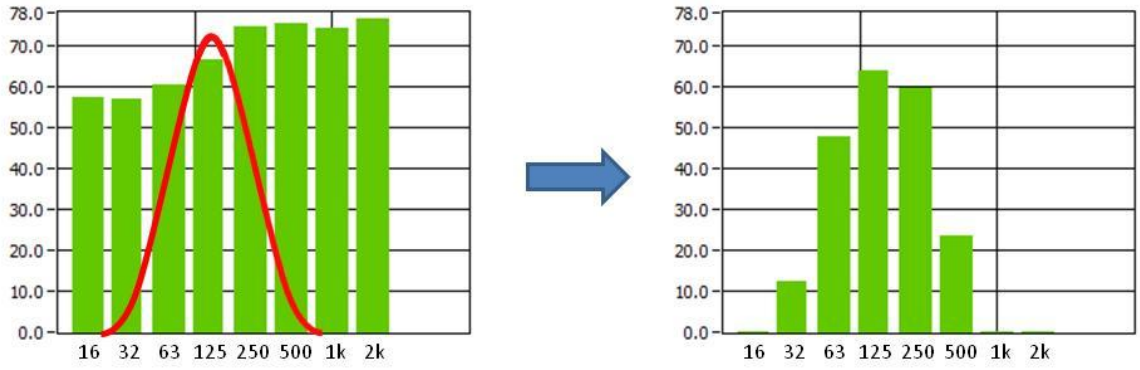
(a) Original signal (b) Windowed signal

Figure 2.3 Time domain signals before and after windowing. (a) Original signal. It is a mixture of a man's shouting happens at around 0.4 second and background noise lasting for whole unity second. SNR of this signal is zero. Red line indicates the shape of rectangular window with 0.3 second length. (b) Windowed signal. All the values outside the window are zeroed, and SNR is increased to 5.0 dB.

2.2.1 (b) Filtering signal pre-processing

Filtering signal pre-processing shares the similar idea with that of the windowing pre-processing though windowing selects the measurement signals on time domain, while the filtering is on frequency domain. Filtering is utilized when the dominated frequency band of the target signal is known in advance. For example, in one case, the octave band spectrum of the input measurement signal is shown in Figure 2.4(a). Assume that we know in advance that the target source is dominated in the frequency band 90 to 180 Hz, thus a filter can be applied to the input signal, select the target frequency band, and eliminate the contribution of other frequencies. SNR of the target source of the particular band is larger than those gained from the signals at

overall frequency bands. Note that the target frequency band is not necessary be the one with the highest amplitude, but should be the one in which the target source is dominant.



(a) Spectrum of original signal (b) Spectrum of filtered signal

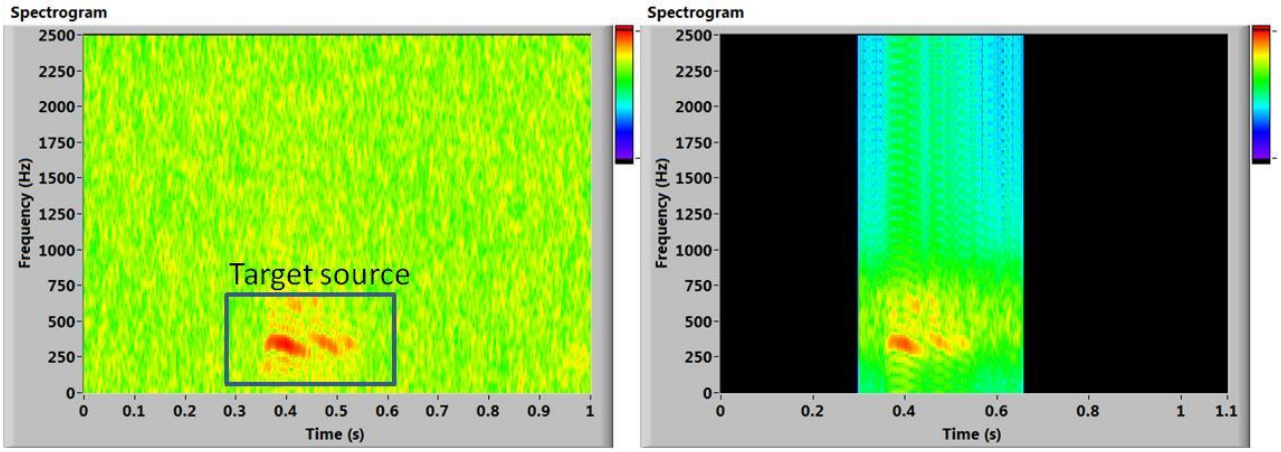
Figure 2.4 Spectrum of signals before and after filtering. (a) Spectrum of original signal. (b) Spectrum of filtered signal. The filtering was applied on the octave band centered at 125 Hz. The contribution of the frequencies beyond the target band is eliminated.

In most cases, the dominate frequency band of the target signal is unknown. Moreover, there might be multiple sound sources in the environment. Therefore, the filtering of the input signals is no longer restricted on one frequency band but multiple bands. One can find out the frequency bands one by one, process the filtered signals in every frequency band, and multiple sound source locations are obtained corresponding to each band. Note that some the localization results at different frequency bands may be the same, or they may be very close, because it is possible that one source is dominate in multiple frequency bands.

2.2.1 (c) Mixture of windowing and filtering

As discussed above, windowing processing can remove the interference of the background noise on time domain, and filtering can reduce the contribution of the un-interested frequency bands. Moreover, the pre-processing of windowing and filtering can be both used during the pre-processing procedure.

Figure 2.5 (a) shows the case when a man's 0.25 second shouting is in a continuous white noise background. To demonstrate the characteristics of the measured signal in time and frequency domain, spectrogram of the signal is shown in Figure 2.5 (a). A blue rectangle is used to point out the area where representing the man's voice. SNR of the measured data is – 3.9 dB. The pre-processing of the input data consists of two steps. Firstly, it is obvious that the target signal is transient, thus the windowing function can be used to select the 0.3 second when the peak happens, and SNR is increased to 0.4 dB. Next, as the spectrogram has shown that the target signal is dominant in the octave band 700 – 1400 Hz, a filter can be applied to the selected signal. In the end, SNR becomes 5.0 dB, and the processed data is shown in Figure 2.5 (b). Note that the black area in Figure 2.5 (b) indicates that the amplitude is zero.



(a) Spectrogram of original signal (b) Spectrogram of pre-processed signals

Figure 2.5 Spectrogram of signals before and after windowing and filtering. (a) Spectrogram of directly original signal. SNR is – 3.9 dB. (b) Spectrogram of pre-processed signal. The signal outside the window function is zeroed, and the frequency distribution beyond the filtering frequency is eliminated. SNR is increased to 5.0 dB.

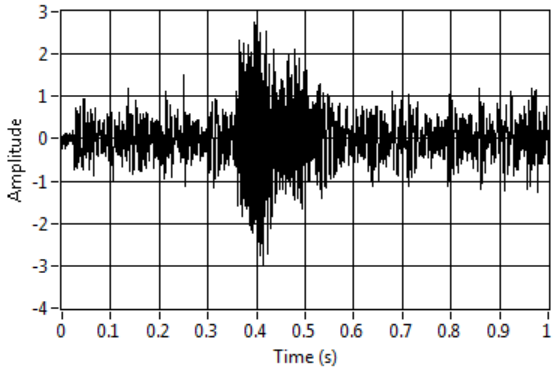
2.2.2 TDOA estimation

As discussed in the literature review, the cross correlation method is commonly used in the estimation of TDOA and TOA. In this dissertation, it is also employed to estimate of TDOA among microphones. Assume time domain signals $x(t)$ and $y(t)$ are observed at two measurement positions. The cross correlation of these two signals $R_{xy}(t)$ can be expressed as:

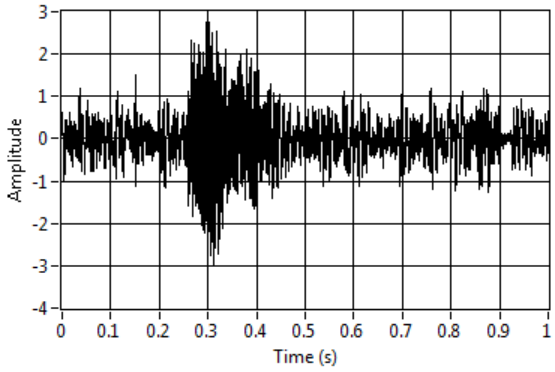
$$R_{xy}(t) = x(t) \otimes y(t) = \int_{-\infty}^{\infty} x^*(\tau) \cdot y(t+\tau) d\tau \tag{2.9}$$

where the symbol " $*$" indicate the complex conjugate.

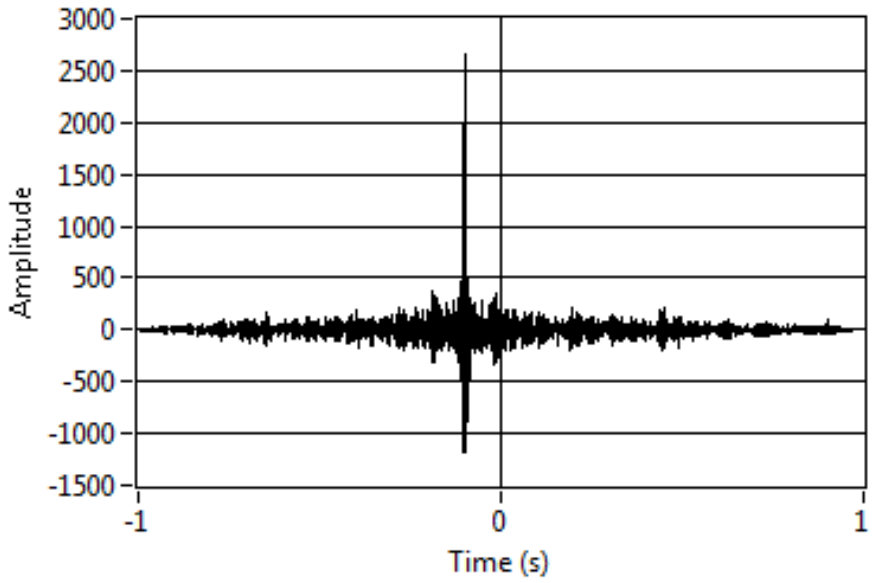
Figure 2.6 gives an example of the cross correlation result among two microphones. Figure 2.6 (a) and (b) are two time domain signals measured at Channel #1 and #2, and Channel #1 is taken as the reference. Figure 2.6(c) demonstrates the cross correlation result. The peak at the time instant – 0.1 second is significant, which shows that TDOA of Channel #2 regarding to Channel #1 is – 0.1 second, in other words, Channel #2 receives original signal from the sound source 0.1 second earlier than Channel #1 does.



(a) Time domain signal at Channel #1



(b) Time domain signal at Channel #2



(c) Cross correlation graph

Figure 2.6 TDOA estimation between two channels. (a) Time domain signal at Channel #1. (b) Time domain signal at Channel #2. (c) Cross correlation graph. The peak in cross correlation results has – 0.1 second offset, which indicates TDOA of Channel #2 regarding to Channel #1 is – 0.1 second.

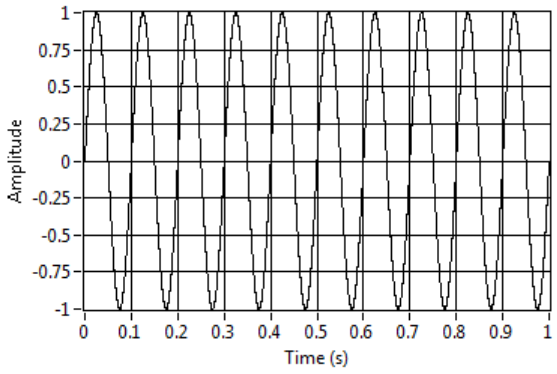
In practice, the environment is non-ideal and inferring background noise can strongly affect the cross correlation results, therefore fluctuation or random peaks may happen in the cross correlation graphs. Thus a window besides zero with a fixed width is usually applied on the cross correlation result to ensure the TDOA estimation is reasonable, and the window width is determined by the microphone spacing as:

$$L = 2\left(\frac{d}{c}\right) \tag{2.10}$$

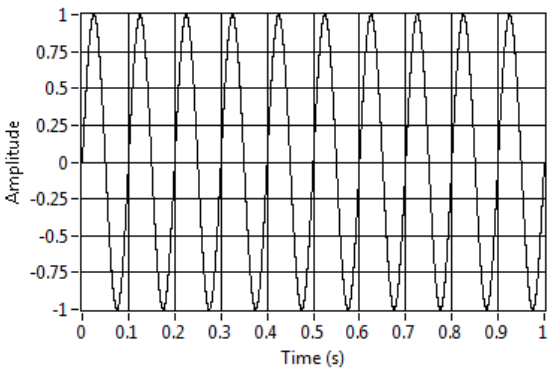
Here d is the microphone spacing, c is speed of sound, and L is the width of the window applied on the time array of the cross correlation result. Only peaks inside the window can be considered as TDOA between the two microphones. For example, if the microphone spacing is 0.7 meter and the speed of sound is 340 m/s, then the window width $L = 2\left(\frac{0.7}{340}\right) = 0.0041$

second. In other word, once the microphone configuration is defined, the range of the possible TDOA is determined. In this case, when microphone spacing is 0.7 meter, the possibly maximal and minimal TDOA is 0.00205 and -0.00205 second, respectively. All the other TDOA result outside this range are caused by non-ideal environment and should be ignored.

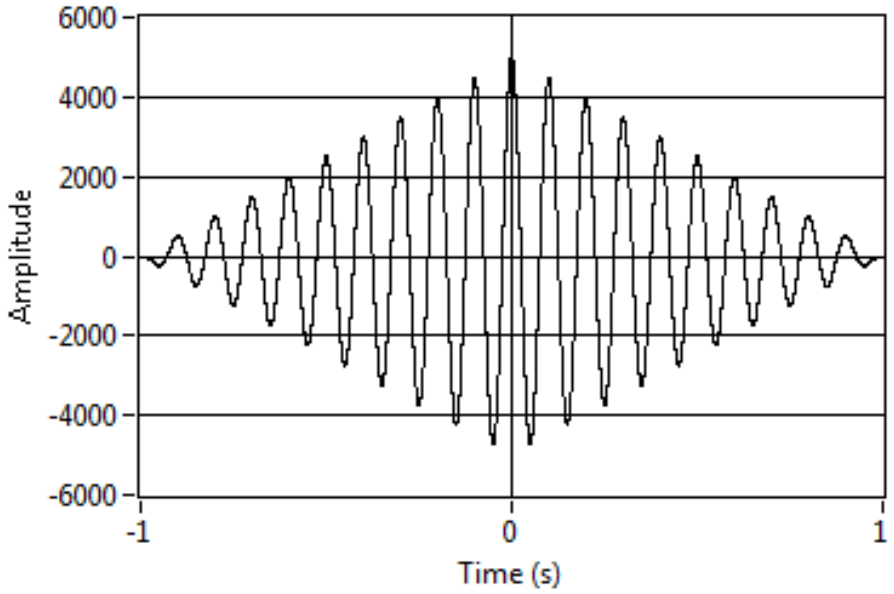
The cross correlation method is applicable to most of the sound types, including transient, continuous, broadband, and narrowband sounds. However, it cannot be used in the cases of a single frequency and its multiple frequencies, because the peaks in cross correlation results for these cases are neither significant nor reliable. One example is a sinusoidal wave with frequency equals to 10 Hz (See Figure 2.7). It can be seen from Figure 2.7 (a) and (b), when the actual TDOA between Channel # 1 and 2 is 0.1 second, which is one period of the sinusoidal wave, the time domain signals at these two microphones are exactly the same. Therefore the TDOA solution by cross correlation method is zero, as Figure 2.7(c) shows. Fortunately, the single frequency's case is very rare in the real world environment.



(a) Time domain signal at Channel #1



(b) Time domain signal at Channel #2



(c) Cross correlation graph

Figure 2.7 TDOA estimation between two sinusoidal signals. (a) Time domain signal at Channel #1. (b) Time domain signal at Channel #2. Channel #2 performs the same as Channel #1, although there is 0.1 second delay at Channel #2. (c) Cross correlation graph. The peak in cross correlation results is at zero.

2.2.3 Triangulation solutions

Inserting the TDOA estimation and positions of four microphones into Equation (8), the double roots solutions for the triangulation equations can be obtained. Rearrange the left and the right part of Equation (2.8), the equation set can be rewritten as:

$$\begin{cases} \sqrt{(x_2-x_s)^2+(y_2-y_s)^2+(z_2-z_s)^2} = c\Delta t_{21} + \sqrt{(x_1-x_s)^2+(y_1-y_s)^2+(z_1-z_s)^2} \\ \sqrt{(x_3-x_s)^2+(y_3-y_s)^2+(z_3-z_s)^2} = c\Delta t_{31} + \sqrt{(x_1-x_s)^2+(y_1-y_s)^2+(z_1-z_s)^2} \\ \sqrt{(x_4-x_s)^2+(y_4-y_s)^2+(z_4-z_s)^2} = c\Delta t_{41} + \sqrt{(x_1-x_s)^2+(y_1-y_s)^2+(z_1-z_s)^2} \end{cases} \quad (2.9)$$

Squaring both side of Equation (2.9) and simplifying the equation, the equations become:

$$\begin{cases} 2(x_1-x_2)x+2(y_1-y_2)y+2(z_1-z_2)z = x_1^2+y_1^2+z_1^2-(x_2^2+y_2^2+z_2^2) \\ \qquad\qquad +c^2\Delta t_{12}^2+2c\cdot\Delta t_{12}\sqrt{(x_1-x)^2+(y_1-y)^2+(z_1-z)^2} \\ 2(x_1-x_3)x+2(y_1-y_3)y+2(z_1-z_3)z = x_1^2+y_1^2+z_1^2-(x_3^2+y_3^2+z_3^2) \\ \qquad\qquad +c^2\Delta t_{13}^2+2c\cdot\Delta t_{13}\sqrt{(x_1-x)^2+(y_1-y)^2+(z_1-z)^2} \\ 2(x_1-x_4)x+2(y_1-y_4)y+2(z_1-z_4)z = x_1^2+y_1^2+z_1^2-(x_4^2+y_4^2+z_4^2) \\ \qquad\qquad +c^2\Delta t_{14}^2+2c\cdot\Delta t_{14}\sqrt{(x_1-x)^2+(y_1-y)^2+(z_1-z)^2} \end{cases}, \tag{2.10}$$

To further simplify the equations above, one can utilize some single variables to indicate a long term polynomials, which are determined in advance as follows:

$$\begin{aligned} &a_1 = 2(x_1-x_2), b_1 = 2(y_1-y_2), c_1 = 2(z_1-z_2), \\ &d_1 = x_1^2+y_1^2+z_1^2-(x_2^2+y_2^2+z_2^2)+c^2\Delta t_{12}^2, \\ &e_1 = 2c\cdot\Delta t_{12}, \\ &a_2 = 2(x_1-x_3), b_2 = 2(y_1-y_3), c_2 = 2(z_1-z_3), \\ &d_2 = x_1^2+y_1^2+z_1^2-(x_3^2+y_3^2+z_3^2)+c^2\Delta t_{13}^2, \\ &e_2 = 2c\cdot\Delta t_{13}, \\ &a_3 = 2(x_1-x_4), b_3 = 2(y_1-y_4), c_3 = 2(z_1-z_4), \\ &d_3 = x_1^2+y_1^2+z_1^2-(x_4^2+y_4^2+z_4^2)+c^2\Delta t_{14}^2, \\ &e_3 = 2c\cdot\Delta t_{14}, \end{aligned} \tag{2.11}$$

Therefore, Equation (2.10) can be rewritten as:

$$\begin{cases} a_1x+b_1y+c_1z = d_1+e_1\sqrt{(x_1-x)^2+(y_1-y)^2+(z_1-z)^2} \\ a_2x+b_2y+c_2z = d_2+e_2\sqrt{(x_1-x)^2+(y_1-y)^2+(z_1-z)^2} \\ a_3x+b_3y+c_3z = d_3+e_3\sqrt{(x_1-x)^2+(y_1-y)^2+(z_1-z)^2} \end{cases}, \tag{2.12}$$

Solving Equation (2.12) yields:

$$x_{1,2} = p_1+\frac{q_1}{2\left(q_3^2+q_1^2+q_2^2-1\right)}\sqrt{2\left(P\pm Q\right)}, \tag{2.13-a}$$

$$y_{1,2} = p_2+\frac{q_2}{2\left(q_3^2+q_1^2+q_2^2-1\right)}\sqrt{2\left(P\pm Q\right)}, \tag{2.13-b}$$

$$z_{1,2} = p_3+\frac{q_3}{2\left(q_3^2+q_1^2+q_2^2-1\right)}\sqrt{2\left(P\pm Q\right)}, \tag{2.13-c}$$

where

$$p_1 = -\frac{\left(-b_1d_3c_2+b_1c_3d_2+c_1d_3b_2-c_1b_3d_2+d_1c_2b_3-d_1b_2c_3\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$

$$p_2 = \frac{\left(-a_1d_3c_2+a_1c_3d_2-c_3d_1a_2-a_3c_1d_2+a_3d_1c_2+d_3a_2c_1\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$

$$p_3 = \frac{\left(d_3a_1b_2+b_3d_1a_2-b_3a_1d_2+a_3d_2b_1-a_3b_2d_1-d_3b_1a_2\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$

$$q_1 = -\frac{\left(-b_1d_3c_2+b_1c_3d_2+e_1c_2b_3-e_1b_2c_3-b_1e_3c_2+b_1c_3e_2+c_1d_3b_2-c_1b_3d_2+d_1c_2b_3-d_1b_2c_3+c_1b_2e_3-c_1e_2b_3\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$
$$+\frac{\left(-b_1d_3c_2+b_1c_3d_2+c_1d_3b_2-c_1b_3d_2+d_1c_2b_3-d_1b_2c_3\right)}{-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3}$$

$$q_2 = \frac{\left(-a_1e_3c_2+a_1c_3e_2-a_1d_3c_2+a_1c_3d_2-a_3c_1e_2-c_3d_1a_2+a_3e_1c_2-a_3c_1d_2+a_3d_1c_2-c_3a_2e_1+e_3a_2c_1+d_3a_2c_1\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$
$$-\frac{\left(-a_1d_3c_2+a_1c_3d_2-c_3d_1a_2-a_3c_1d_2+a_3d_1c_2+d_3a_2c_1\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$

$$q_3 = \frac{\left(d_3a_1b_2+b_3d_1a_2-e_1a_3b_2-b_3a_1d_2+a_1e_3b_2-a_1e_2b_3-b_1e_3a_2+b_1e_2a_3+e_1b_3a_2+a_3d_2b_1-a_3b_2d_1-d_3b_1a_2\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$
$$-\frac{\left(d_3a_1b_2+b_3d_1a_2-b_3a_1d_2+a_3d_2b_1-a_3b_2d_1-d_3b_1a_2\right)}{\left(-b_2a_3c_1+c_2a_3b_1-a_1c_2b_3+a_1b_2c_3-a_2b_1c_3+a_2c_1b_3\right)}$$

$$P = -p_1q_1+y_1q_2-p_2q_2+z_1q_3-p_3q_3+x_1q_1$$

$$\begin{aligned}Q = &-2y_1p_2+x_1^2+y_1^2+z_1^2-p_2^2q_3^2-y_1^2q_1^2-y_1^2q_3^2-p_3^2q_1^2-p_3^2q_2^2-x_1^2q_3^2-x_1^2q_2^2+p_2^2+2y_1p_2q_3^2\\&-p_2^2q_1^2+2x_1p_1q_3^2+2y_1p_2q_1^2-z_1^2q_1^2+2x_1p_1q_2^2-z_1^2q_2^2+2z_1p_3q_2^2+2z_1p_3q_1^2-p_1^2q_2^2-p_1^2q_3^2+p_3^2\\&+2p_1q_1p_3q_3-2p_1q_1z_1q_3+2y_1q_2x_1q_1-2y_1q_2p_3q_3+2y_1q_2z_1q_3-2p_2q_2x_1q_1+2p_2q_2p_3q_3\\&-2p_2q_2z_1q_3+2z_1q_3x_1q_1-2p_3q_3x_1q_1-2z_1p_3-2x_1p_1+2p_1q_1p_2q_2-2p_1q_1y_1q_2+p_1^2\end{aligned}$$

Note that in some cases, the roots of (*x*, *y*, *z*) are complex numbers which have non-zero imaginary parts. That is mainly caused by the error during the estimation of TDOA and the calculation error, thus all the imaginary parts in the roots should be zeroed.

2.2.4 Selection of final results

As seen in Equation (2.13), there are two roots or solutions for the source location, but only one of them leads to the correct location of the source, while the other is a ghost image. There are three possible cases of the roots distribution in space: two roots are repeated; two roots are on opposite direction to the microphone set; two roots are on the same direction but at different ranges.

For the case that two roots are the same, the location of the sound source can be defined by selecting any of the roots.

For the case that the two roots are located at opposite direction regarding the microphone set, the correct solution can be selected by recheck the TDOA estimation between the microphones which are nearest and farthest to the source. The algorithm consists of three steps:

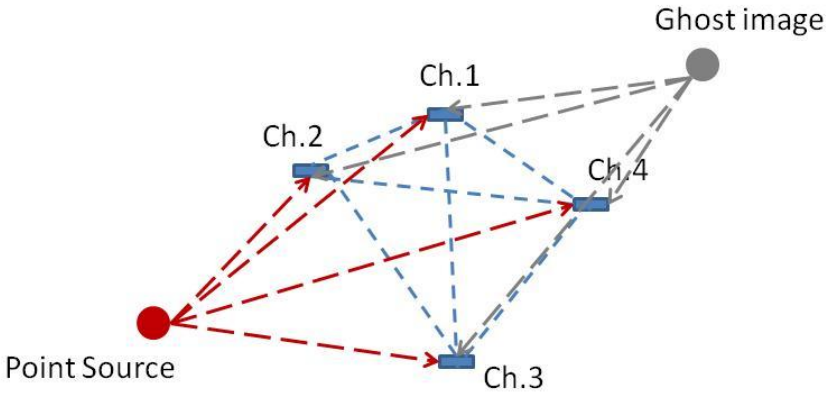
1) Choose one of the roots from the solution set, and find out the nearest and farthest microphone to this root.

2) Take the nearest microphone as the reference, and calculate the TDOA of the farthest microphone regarding to the nearest one.

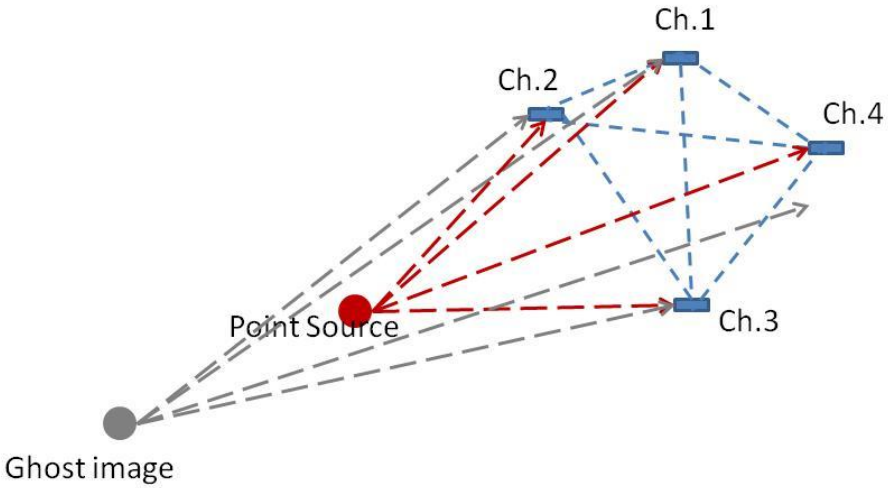
3) Check the value of this TDOA. If the source solution is the true location of the source, the signal should arrive at the nearest microphone earlier than the farthest one, the TDOA value obtained in the 2nd step should be positive. Therefore if the TDOA value is positive, the source root used in the first step is the correct one. On the other hand, if the TDOA is negative, the root is ghost image.

Take Figure 2.8 (a) for an example, TDOA between any two microphones is dependent on the measurement signals, and is fixed. Assume it is known that TDOA of Channel #4 regarding to Channel #2 is positive, namely, $\Delta t_{42} > 0$, and consequently, $\Delta t_{24} < 0$. For the root on

the left, the nearest microphone is Channel #2 and the farthest one is Channel #4, their TDOA Δt_{42} is positive, which means that the source reaches Channel #2 earlier than Channel #4, thus the root on the left of the microphone set is the correct solution of sound source localization. On the other hand, when testing the root on the right, its nearest microphone is Channel #4 and farthest is Channel #2. Their TDOA Δt_{24} is negative. Therefore the root on the right is the ghost image.



(a) Two roots are at the opposite direction to the microphone set



(a) Two roots are at the same direction to the microphone set, but at different range

Figure 2.8 Two possible case of the roots distribution in space when they are not repeated

For the case that two roots are located at the same direction but different ranges, as Figure 2.8 (b) shows, the correct sound source location can be selected with the information from the amplitude of the measured data. An empirical formulation is developed to evaluate error of the amplitude decay rates of the two roots:

$$error_{decay} = \frac{\dfrac{\dfrac{1}{4}\sqrt{\sum_{i=1}^{4} p_{rms,i}^2}}{\sqrt{r_{farthest} \cdot r_{nearest}}} - \dfrac{p_{rms,\,farthest} \cdot p_{rms,nearest}}{r_{farthest} \cdot r_{nearest}}}{\dfrac{p_{rms,\,farthest} \cdot p_{rms,nearest}}{r_{farthest} \cdot r_{nearest}}} \times 100\% \tag{2.14}$$

where $r_{nearest}$ and $r_{farthest}$ are the distances from the root to the nearest and farthest microphones, and $p_{rms,nearest}$ and $p_{rms,farthest}$ are the corresponding root mean square value of the amplitude. The main ideal of Equation (14) is to compare the decay rate of the amplitude calculated by two ways: one is based on the average of distances between the source and the four microphones, the other is based on the difference between the nearest and the farthest microphones. For the correct source location in an ideal environment, the decay rates obtained by these two ways should be the same. However, in the real cases, considering the error during TDOA estimation and calculation, the root with smaller $error_{decay}$ value is selected as the correct solution.

2.3 Impact of various parameters on source localization algorithm

To get a better understanding of the acoustic model in this localization method, impacts of various parameters are tested in to evaluate the characteristics of the acoustic model based sound source localization methodology. During the numerical simulation, different types of sound signals are tested, such as human voices, truck noise, chopper sound, machine noise, etc. The positions of the microphones are fixed at (0.5, 0, 0), (0, 0.5, 0), (0, 0, -0.5), and (0, 0, 0.5).

The positions of the sound sources are chosen arbitrarily, and the mixed signals at four microphones are generated numerically obeying the spherical spreading law. With the mixed signals and the position of the microphones, the computer program follows the flowchart in Figure 2.2 and output the final locations of the sound sources.

To evaluate the accuracy of the sound source localization, error of localization result is defined as follows:

$$error_{localization} = \frac{\left|\vec{r} - \vec{r}_{benchmark}\right|^2}{\left|\vec{r}_{benchmark}\right|^2} \times 100\% \tag{2.15}$$

where r is the calculated vector result the source, and $r_{benchmark}$ is the benchmark position of the sound source.

2.3.1 Impact of frequency

Figure 2.9 shows the impact of frequency on the sound source localization results. The horizontal coordinate indicates the frequency of the sources, and is expressed in octave bands. The vertical coordinate is the error of localization in percentage. It is obvious that the curves of errors at various SNR are almost flat, which means the accuracy of the result is independent of frequency.

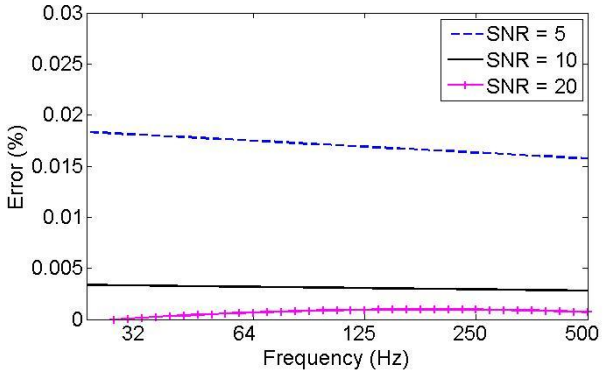


Figure 2.9 Impact of frequency on localization accuracy.

2.3.2 Impact of source range

Figure 2.10 demonstrates the impact of source range. The microphone spacing is fixed as mentioned above, and the source is set at various source ranges. The localization error increases with the source range. However, it can be seen from the figure that even if SNR is as low as zero dB, which means that the amplitude of the target signal is the same as that of the background noise, the error of the source localization can be still controlled less than 1%. Moreover, as the limitation on source range is influenced by the microphone spacing, one can always increase the microphone spacing to gain a large source range with satisfactory accuracy.

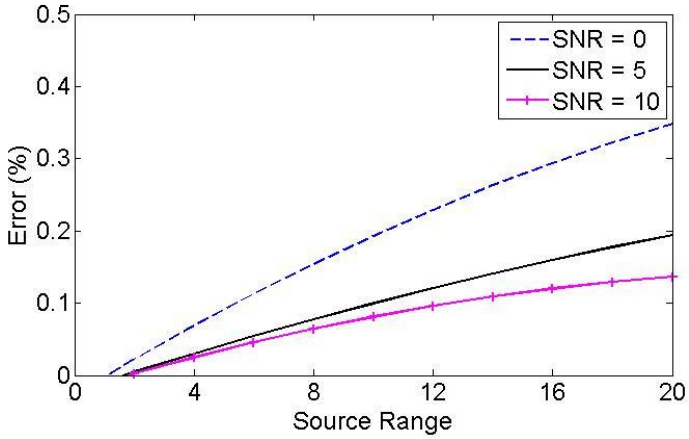


Figure 2.10 Impact of source range on localization accuracy.

2.3.3 Impact of SNR

It is found in the numerical simulation that SNR is the major impact on the accuracy of the sound source localization results. As Figure 2.11 shows, the error reduces when SNR increases.

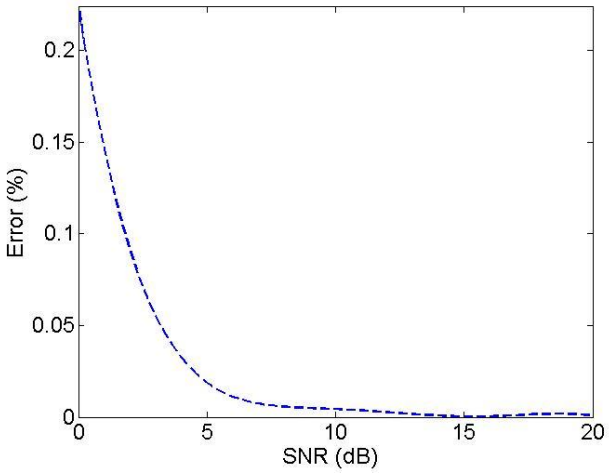


Figure 2.11 Impact of SNR on localization accuracy.

2.3.4 Impact of microphone spacing

To test the impact of microphone spacing, four channels are placed at various distances from 0.1 meter to unity meter. For the cases when SNR is equal to five or ten, the impact of the microphone spacing is small. But when SNR is zero, the impact of microphone spacing is obvious, and error decreases when microphone spacing increases.

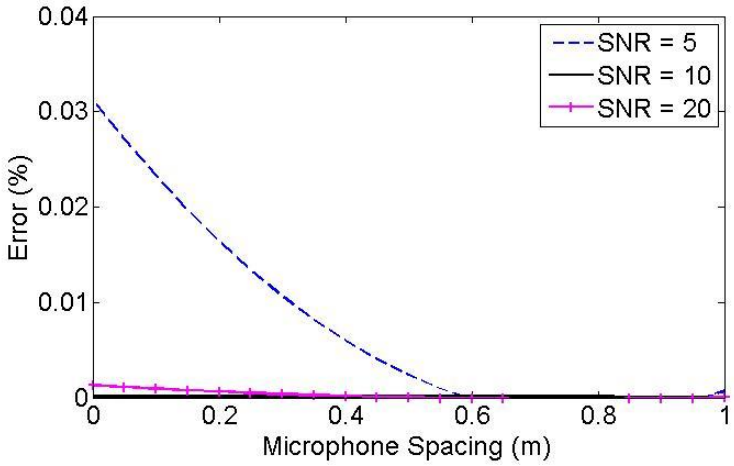
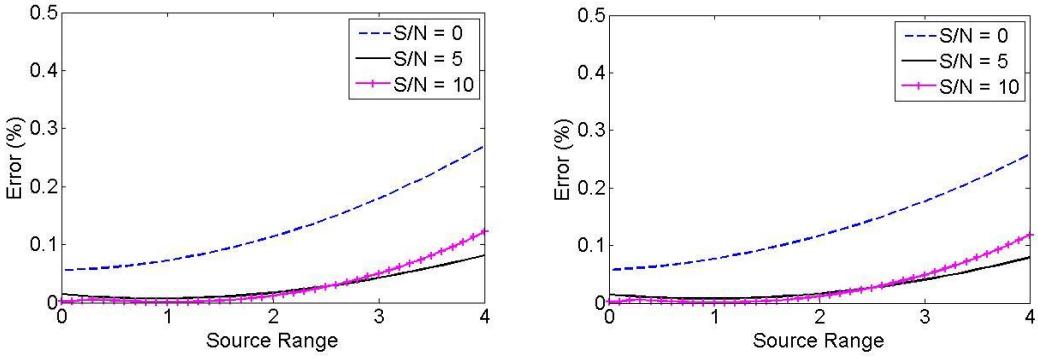


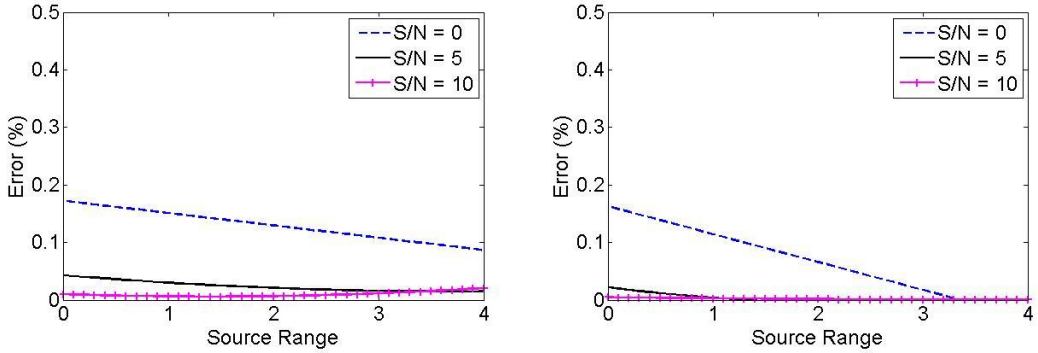
Figure 2.12 Impact of microphone spacing on localization accuracy.

2.3.5 Impact of frequency on spatial resolution

As discussed in the literature review, the spatial resolution of beamforming technology is no better than one wavelength of the sound emitted by the target source, which makes it not applicable for many low-frequency cases. The spatial resolution of the newly proposed method is tested in this section.



(a) 22 – 45 Hz (b) 90 – 180 Hz



(c) 710 – 1400 Hz (b) 1400 – 2800 Hz

Figure 2.13 Impact of spatial resolution on localization accuracy. (a), (b), (c), and (d) show the error distribution of source aiming at various frequency bands. The source is truck pass by noise, and the background is white noise. The errors by the four frequencies perform the same, which indicates that the spatial resolution is independent of the frequency.

Specifically, the target source was placed at (1, 0, 0) and the interfering source was at (1.5, 0, 0) in meters. So the relative distance was 0.5 m. Results show that the errors in source localization using the model based triangulation methodology is below 0.5% for frequencies ranging from 22 Hz to 2,800 Hz. In other words, this method can correctly locate the target source separated from an interfering source at a distance much shorter than the wavelength of the sound signal emitted by the source. This test indicates that the present method is basically independent of the frequency.

2.4 Improved 3D sound source localization algorithm

As discussed above, the error of sound source localization increases with source range. In this section, an improved 3D localization algorithm is introduced, which employs more than four microphones and can achieve higher accuracy and larger source range than the basic microphone set.

Figure 2.14 gives a general idea of the tendency of the impact of the error on TDOA on the final source localization results in the four-microphone set. When the source range is small, in other words, the target source is near to the microphone set, the error on TDOA has a limited impact on the accuracy of localization. However, when the source is far away from the microphone set, a small amount of error on TDOA can cause a significant error on the final localization results. In this way, the error of localization increases with the source range, and the applicable source range coverage for a set of microphones with fixed microphone spacing is limited.

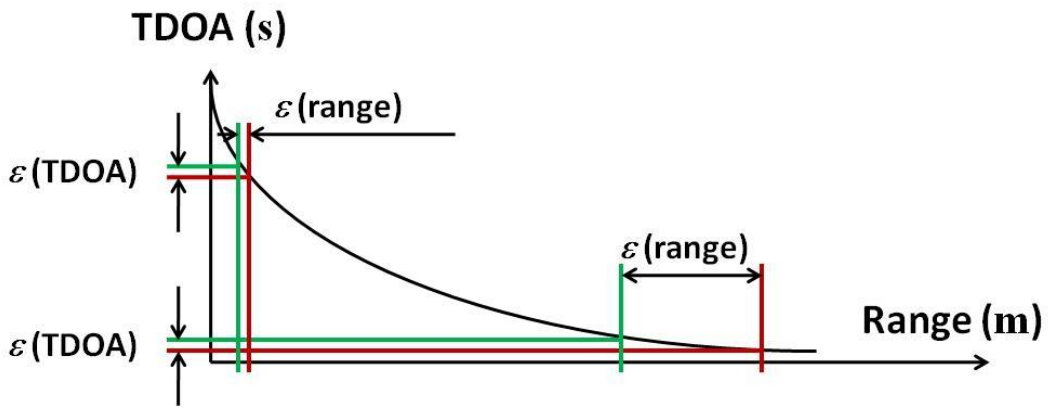


Figure 2.14 Impact of TDOA error on the localization accuracy. With a fixed value of error on TDOA, only a small value of localization happens when the source range is small, while it causes large error when the source range is large.

To improve the accuracy and source range of the 3D sound source localization algorithm, more than four microphones can be utilized. In this way, the accuracy of the final result can be improved in two ways: firstly, when multiple microphones are used during measurements, redundancy check on TDOA estimation can be conducted to reduce the error on TDOA which influences the accuracy of source localization; secondly, more than four microphones can compose multiple four-microphone sets which lead to multiple source location solutions, therefore the final result can be gained by analyzing the distribution of those solutions and gain a better accuracy.

Theoretically, a larger number of microphones can better improve the accuracy of the final results. However, using too many microphones increases the cost of device, and the calculation procedure becomes more complex which is time consuming. To keep the original intention of this dissertation, which is to develop a portable and cost effective technique to locate sound sources in real time, the number of microphone is extended to six, and an improved 3D sound source localization algorithm is developed.

The computer flow chart of six microphones is shown in Figure 2.15. To make the localization methodology flexible and applicable to most of the measurement cases, the designed program can switch between the basic model and the improved model, in which the basic model only uses the data from four microphones, while the improved model employs all the six microphones measurement. At the beginning of the program, the time domain signals are obtained at six microphones, and pre-processed by windowing and filtering procedures. Next, if high accuracy is desired, then data at all the six microphones are used and TDOA estimation between any two microphones is conducted, followed by a redundancy check on TDOAs. Subsequently, the improved TDOA values are submitted into six microphone units, each of them uses four out of the six microphones, therefore six solutions on the source location are obtained. Analyzing the six solutions leads to the final result of the source location. On the other hand, if only a brief view of the location of the source is required, one can select only four of six microphones, skip the redundancy check procedures, and output the source localization solution directly.

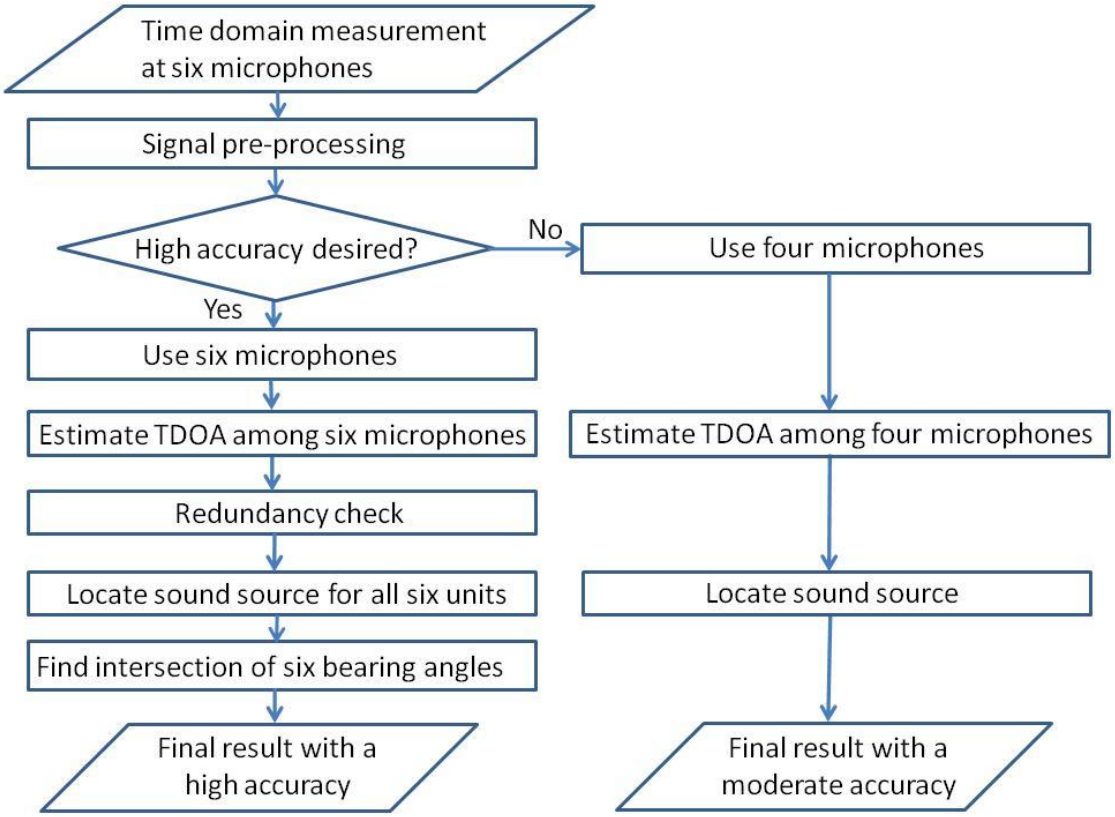


Figure 2.15 Computer flowchart for determining source locations. using a single unit of four microphones and M units that consist of N microphones. A single unit can locate sources in real time, but its accuracy may be compromised when the test environment is non-ideal; whereas using M units can improve the accuracy in source localization even in a non-ideal environment but its speed may be slowed down.

2.4.1 Redundancy check on TDOA estimation

Redundancy check can be conduct when multiple microphones are employed in the device, and it can reduce the random error of TDOA which are caused by the fluctuating backgrounds. Define the number of microphones in the measurement system is *N*, and TDOAs among microphones can form a matrix as:

$$\left[\Delta t_{ij}\right]=\begin{bmatrix} \Delta t_{11} & \Delta t_{12} & \cdots & \Delta t_{1j} & \cdots & \Delta t_{1N} \\ \Delta t_{21} & \Delta t_{22} & \cdots & \Delta t_{2j} & \cdots & \Delta t_{2N} \\ \vdots & \vdots & \ddots & & & \vdots \\ \Delta t_{i1} & \Delta t_{i2} & & \ddots & & \Delta t_{iN} \\ \vdots & \vdots & & & \ddots & \vdots \\ \Delta t_{N1} & \Delta t_{N2} & \cdots & \Delta t_{Nj} & \cdots & \Delta t_{NN} \end{bmatrix} \tag{2.16}$$

It is obvious that TDOA of any microphone with itself is zero, and $\Delta t_{ij} = -\Delta t_{ji}$, where $1 \leq i, j \leq N$. Therefore, TDOA matrix can be simplified as:

$$\left[\Delta t_{ij}\right]=\begin{bmatrix} 0 & \Delta t_{12} & \cdots & \Delta t_{1j} & \cdots & \Delta t_{1N} \\ -\Delta t_{12} & 0 & \cdots & \Delta t_{2j} & \cdots & \Delta t_{2N} \\ \vdots & \vdots & \ddots & & & \vdots \\ -\Delta t_{1i} & -\Delta t_{2i} & & \ddots & & \Delta t_{iN} \\ \vdots & \vdots & & & \ddots & \vdots \\ -\Delta t_{1N} & -\Delta t_{2N} & \cdots & -\Delta t_{jN} & \cdots & 0 \end{bmatrix} \tag{2.17}$$

When multiple microphones are utilized in the improved model, TDOA estimation can be further checked with redundancy procedure; therefore the accuracy of the estimation on TDOA can be improved and has less negative impact on the accuracy of the final source localization results.

The basic idea of the redundancy check on TDOA is that assume any TDOA between two microphones, for example, the i^{th} microphone and the j^{th} microphone, can be recalculated with the time domain signals of i^{th} microphone and the j^{th} microphone together with an additional k^{th} microphone. The equation can be obtained as: $\Delta t_{ij} = t_i - t_j = t_i - t_k + t_k - t_j = \Delta t_{ik} + \Delta t_{kj}$, where $1 \leq i, j, k \leq N$, t_i and t_j are TOAs from the source to the i^{th} and j^{th} microphone, respectively. Note that the third microphone can be any one in the microphone array as long as it is not the i^{th} microphone or the j^{th} microphone, thus for an N-microphone array, the redundancy check on TDOA on one pair of microphones can be hold for $(N - 2)$ times.

For example, N is determined as six in this dissertation. Therefore, when aiming at find the TDOA value between Channel #1 and #2, besides the directly measured TDOA of $\Delta t_{12,\text{ direct}}$ with cross-correlation method on the time domain signals of Channel #1 and #2, the TDOA of Δt_{12} can also be achieved four other approaches, which are:

$\Delta t_{12,\text{ checked by Channel \#3}} = \Delta t_{13} + \Delta t_{32}$,

$\Delta t_{12,\text{ checked by Channel \#4}} = \Delta t_{14} + \Delta t_{42}$,

$\Delta t_{12,\text{ checked by Channel \#5}} = \Delta t_{15} + \Delta t_{52}$, and

$\Delta t_{12,\text{ checked by Channel \#6}} = \Delta t_{16} + \Delta t_{62}$.

Taking average of these five values leads to a more accurate TDOA between Channel #1 and #2, and the final averaged result can be expressed as:

$$\Delta \bar{t}_{12} = \frac{\Delta t_{12,\text{ direct}} + \Delta t_{12,\text{ checked by Channel\#3}} + \Delta t_{12,\text{ checked by Channel\#4}} + \Delta t_{12,\text{ checked by Channel\#5}} + \Delta t_{12,\text{ checked by Channel\#6}}}{5}$$
$$= \frac{\Delta t_{12} + (\Delta t_{13} + \Delta t_{32}) + (\Delta t_{14} + \Delta t_{42}) + (\Delta t_{15} + \Delta t_{52}) + (\Delta t_{16} + \Delta t_{62})}{5} \tag{2.18}$$

The general expression for TDOA between any two microphones in an N-microphone array is:

$$\Delta \bar{t}_{ij} = \frac{\sum_{k=1}^{N} t_i - t_k + t_k - t_j}{N-1} = \frac{\sum_{k=1}^{N} \Delta t_{ik} + \Delta t_{kj}}{N-1}, \quad i, j = 1, 2, \ldots, N, \ i < j, \ k \neq i \tag{2.19}$$

and TDOAs matrix after the redundancy check is:

$$\left[\Delta\tilde{t}_{ij}\right]=\begin{bmatrix} 0 & \frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{k2}}{N-1} & \cdots & \frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{kj}}{N-1} & \frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{k(j+1)}}{N-1} & \cdots & \frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{kN}}{N-1} \\ -\frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{k2}}{N-1} & 0 & \cdots & \frac{\sum_{k=1}^{N}\Delta t_{2k}+\Delta t_{kj}}{N-1} & \frac{\sum_{k=1}^{N}\Delta t_{2k}+\Delta t_{k(j+1)}}{N-1} & \cdots & \frac{\sum_{k=1}^{N}\Delta t_{2k}+\Delta t_{kN}}{N-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ -\frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{ki}}{N-1} & -\frac{\sum_{k=1}^{N}\Delta t_{2k}+\Delta t_{ki}}{N-1} & \cdots & 0 & \vdots & \cdots & \frac{\sum_{k=1}^{N}\Delta t_{ik}+\Delta t_{kN}}{N-1} \\ -\frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{k(i+1)}}{N-1} & -\frac{\sum_{k=1}^{N}\Delta t_{2k}+\Delta t_{k(i+1)}}{N-1} & \cdots & \cdots & 0 & \cdots & \frac{\sum_{k=1}^{N}\Delta t_{(i+1)k}+\Delta t_{kN}}{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \vdots \\ -\frac{\sum_{k=1}^{N}\Delta t_{1k}+\Delta t_{kN}}{N-1} & -\frac{\sum_{k=1}^{N}\Delta t_{2k}+\Delta t_{kN}}{N-1} & \cdots & -\frac{\sum_{k=1}^{N}\Delta t_{jk}+\Delta t_{kN}}{N-1} & -\frac{\sum_{k=1}^{N}\Delta t_{(j+1)k}+\Delta t_{kN}}{N-1} & \cdots & 0 \end{bmatrix} \quad (2.20)$$

where $i, j = 1, 2, \ldots, N$, $i < j$, and $k \neq i$.

The redundancy check on TDOAs estimation matrix can reduce the random error on TDOAs caused by fluctuating background noise, and therefore improve the accuracy of the final source localization results.

2.4.2 Multiple microphone set source localization algorithm

Numerical simulation indicates that the major limitation of the sound source localization algorithm is that the source range coverage is restricted by the microphone spacing. In other word, with a fixed dimension of microphone arrangement, the source range coverage of localization is limited and the accuracy of the source locating results cannot be guaranteed when the source is out of the source range coverage, because the error on source location increases with the source range. This characteristic is caused by the algorithms of the acoustic model. As Figure 2.14 shows, the impact of error on TDOA estimation varies at different source range: when the target signal is at a larger source range, the error of TDOA has a stronger negative impact on the accuracy of source localization.

Another characteristic of the source localization methodology is that the DOA of the source to the four-microphone set is almost independent of the source range, and more reliable than the range detection in the results. Therefore, it is possible to locate a source more accurately by finding the intersection of two DOAs obtained by two microphone sets, as Figure 2.16 shows.

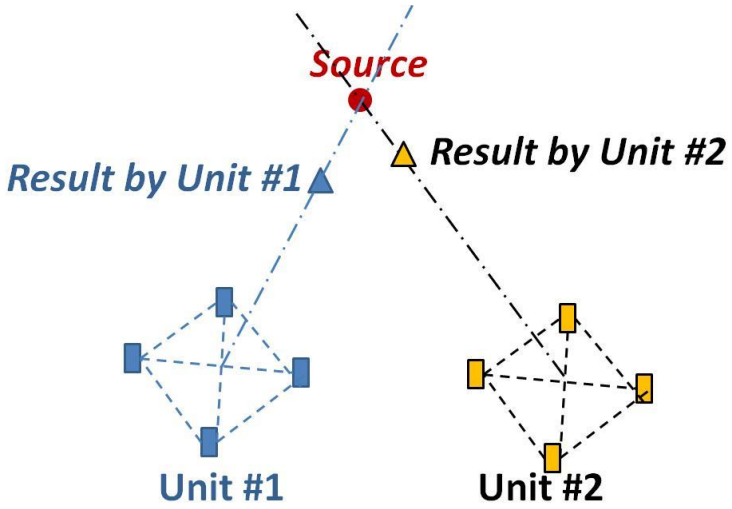


Figure 2.16 Finding the intersection of the localization results by two four-microphone set. The red dot indicates the position of the source, and the blue and yellow rectangles stands for two four-microphone sets, which are Unit #1 and 2, respectively. The blue and yellow triangles are the directly measured source locations by the units, and the dashed dot lines are the DOA detected by them. Intersection of the two dashed dot lines is very near to the original source location.

Theoretically, when there are N microphones in the array, maximal number of $M = C_N^4$ microphone set can be gained. However, as the basic localization program should run once for each of the microphone set, using $M = C_N^4$ microphone units is time consuming. What is more, it is not necessary to employ all the available microphone sets to get an improved source location. One can chose any value of M, as long as $M \le C_N^4$. Six microphone sets are chosen among the six microphones in this dissertation.

The procedure of the N-microphone array source localization algorithm has four steps:

1) Obtain the bearing angle lines of the sources in every four-microphone sets.

2) Find out the intersection points of every two lines. If the two lines are skew lines, the point nearest to both of the lines can be treated as the intersection point. If the two lines are parallel, ignore this pair of lines and move forward to the next pair.

3) Collect all the directly calculated sound source locations and the intersection points as the database of the source location. Analyze the distribution of them, and filter out the singular points which are far away from the convergence of the database and representing abnormal results.

4) Take the special average of the rest points in the database. The averaged point is considered as the final result of the source localization.

2.4.3 Numerical simulation results

In the numerical simulation, the positions of six microphones are set as follows, with unit in meter.

Channel #1: (-1.3856, 0.4000, 0),

Channel #2: (0, 0, 0),

Channel #3: (1.3856, 0.4000, 0),

Channel #4: (-1.3856, 0.4000, 0.8000),

Channel #5: (0, 0, 0.8000), and

Channel #6: (1.3856, 0.4000, 0.8000).

and the microphones chosen in each set are:

Microphone set #1: Channel #2, #1, #3, and #4,

Microphone set #2: Channel #2, #3, #1, and #5,

Microphone set #3: Channel #2, #3, #1, and #6,

Microphone set #4: Channel #5, #1, #4, and #6,

Microphone set #5: Channel #5, #2, #4, and #6, and

Microphone set #6: Channel #5, #3, #4, and #6.

The chosen of the microphones position and sets are based on the consideration of the dimension of the prototype device, and the microphone spacing approximately equal 0.8 meter. However, to get more general characteristics of the methodology, the results are plotted dimensionless. The source range, in terms of (ξ, η, h) are defined in terms of the source coordinates (x, y, z) and the microphone spacing d in meter.

$$\eta = \frac{x}{d},\ \xi = \frac{y}{d},\ h = \frac{z}{d} \tag{2.21}$$

Therefore, when the dimensionless source range (ξ, η, h) are determined, the source localization range in practice with unit meter changes with the microphone spacing.

To estimate the error in cross-correlation, random generated values at a standard deviation 0.0001 with a zero mean is added to the precise value of the actual TDOAs. Next, the modified TDOA matrix, which has certain error, is used as the directly calculated cross-correlation results of the six microphone measurement setup. Comparisons are conducted on the error distribution of the results by three different algorithms procedures, which are:

1) Basic method: Select four out of the six microphones to calculate the sound source location. No redundancy method is employed, and TDOAs used in the localization algorithm are the directly measured ones by cross-correlation.

2) Improved method without redundancy check: No redundancy check is conducted on TDOA matrix. Use all of the six microphones, analyze the calculated source locations gained by the six microphone sets, and get the final result.

3) Improved method with redundancy check: Firstly conduct redundancy check on TDOA matrix to get a new TDOA matrix. Then use the improved TDOAs in the six microphone sets to gain the six source location results. Analyze the six locations to get the final result.

The distributions of the errors on the source localization by these three procedures are shown in Figure 2.17. The source ranges examined are six times of the microphone spacing at three heights: h = –1.25, h = 0 and h = 1.25. And each plot in Figure 2.17 is the average of the error distributions by 500 runs, in each run the error on TDOA is generated randomly.

Figure 2.17 (a) to (c) show the error distribution of the results by the basic model. It is obvious that the localization is more effective when the source is in the front and back of the microphone array than when the source is at the sides of the array. When the source is in the front and back of the array, the error can be controlled less than 10%. However, when the source is located at either side of the array, the error can be as high as 100%; in other words, the source localization result is not reliable. This means that the accuracy of the sound source localization is related to the position of the source. Figure 2.17 (d) to (f) illustrate the error distribution of the results by the improved model without redundancy check. It is obvious that the accuracy of the source localization is highly improved comparing to that by the basic model, though the tendency of the impact of the source position still exists. Figure 2.17 (g) to (i) shows the error distribution of the results by the improved model with redundancy check. It can be seen that the results in this group have the best accuracy.

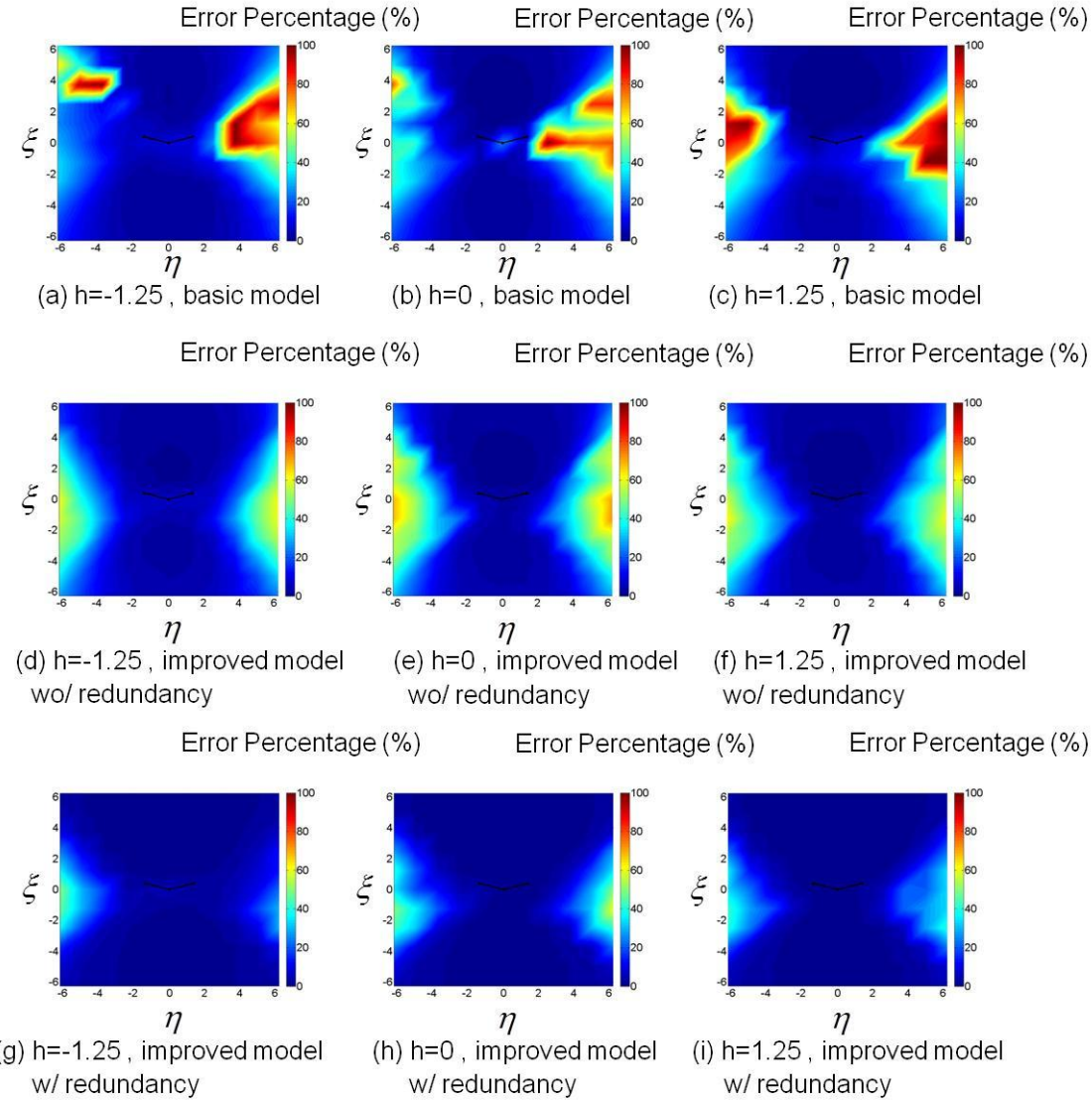


Figure 2.17 Numerical simulations of source localization results subject to random errors in TDOA estimations. The microphone array is indicated by black lines. (a), (b), and (c) show the error distribution results by using the basic model; (d), (e), and (f) indicate the error distribution results by using the improved model without redundancy checks; (g), (h), and (i) represent the error distribution results by using the improved model with redundancy checks. (a), (d), and (g) are over a horizontal plane at a dimensionless height of h = –1.25 respect to the origin of the coordinate system; (b), (e), and (h) are on a horizontal plate at h = 0; and 4(c), 4(f), and 4(i) are on a horizontal plane at h = 1.25.

Table 1 demonstrates the average error of the three programs. Both the redundancy check on TDOA matrix and the multiple-microphone set algorithm can improve the accuracy of the sound source localization method. Moreover, the source range coverage of this six-microphone

setup is up to six times of the microphone spacing, which is much larger than those by any current methodologies mentioned in the literature review [31].

Table 2.1. Average errors of source localizations by using three different algorithms in numerical simulation tests.

Models Used	$h = 1.25$	$h = 0$	$h = -1.25$
Basic Model	21.33%	19.05%	19.32%
Improved Model, without redundancy	16.52%	18.95%	15.73%
Improved Model, with redundancy	6.74%	7.72%	5.75%

Figure 2.17 also shows a tendency of the source localization that the error of the result is minimal when the source is in the front and back of the microphone array. Therefore as the microphone set device is portable designed, the best measurement condition is to let the microphone arrays facing the target signals. Figure 2.18 demonstrates the same data as Figure 2.17 (h), but at a smaller scale, which is from zero to 3%. The errors in the area within the half circle are under 1.5%, and the area can be considered as the optimal measurement zone for this particular microphone array setup. Note that a different microphone configuration may lead to a different optimal zone.

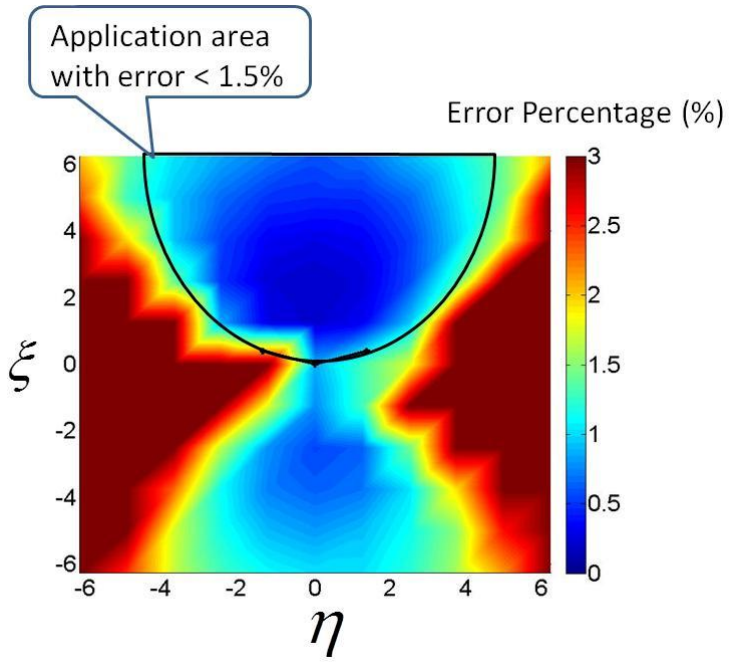


Figure 2.18 Safe zone for locating sound sources using a six-microphone set with redundancy procedure in practice. The scale bar shows the percentage errors in source localization that varies from 0 to 3%. The half circle indicates the safe area within which the error is less than 1.5%.

CHAPTER 3

EXPERIMENTAL VALICATION OF SOUND SOURCE LOCALIZATION

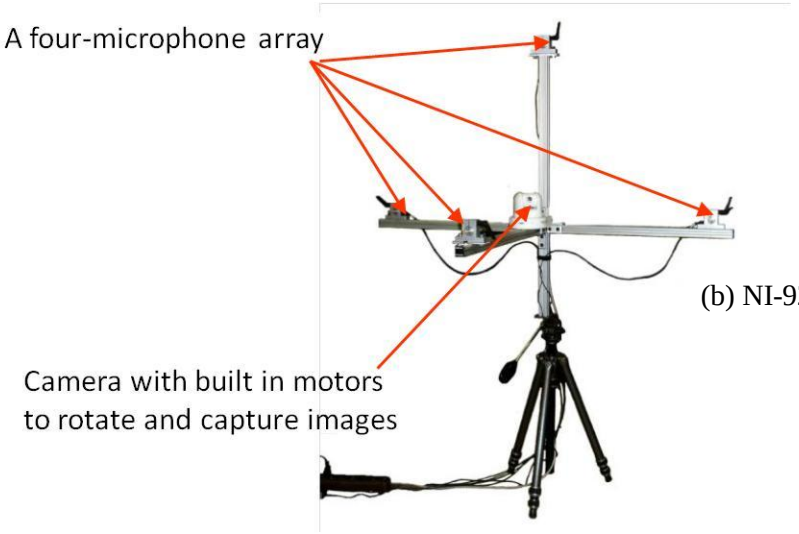
To validate the proposed methodology for sound source localization, experimental validations are conducted with various real world sounds in several non-ideal environments. Three prototype devices are tested: orthogonal four-microphone setup, non-orthogonal four-microphone setup, and six-microphone setup. The devices were tested in various environments, including the Acoustic, Vibration, and Noise Control Laboratory (AVNC Lab), Machine Shop, auditorium room, and the hall in the Building of Engineering, Wayne State University. Multiple types of sounds are used as the target sound sources and background noises in the experiments, some of them are real world sounds played through loudspeakers, such as machine noises, chopper sound, music, radio news, etc. The others are some real sound in the environments, such as people talk, clapping, sport gunshot firing, machine running noises, etc. The experimental validation has shown that both the basic model and the improved model can successfully locate multiple sound sources in none-ideal environments in 3D space, and the improved model with redundancy check has larger source range coverage and higher accuracy than the basic one.

3.1 Experimental validation results for four-microphone set

The four-microphone set devices were firstly validated in various environments with different types of sound sources. Two prototype devices were tested in this section, namely, the orthogonal and non-orthogonal four-microphone set.

3.1.1 Experimental setup

Two microphone configurations are built up to test the basic model of the sound sources localization method. The orthogonal four-microphone device, shown as Figure 3.1, consists of one thermometer, four B&K ½-inch condenser microphones, a NI-9162 carrier with NI-9234 signal acquisition module from the National Instrument, and a web camera. The thermometer is used to measure the temperature in the environment, with which the speed of sound can be estimated in real time during sound source localization. The positions of the four microphones in Cartesian coordinates are (0.5, 0, 0), (0, 0.5, 0), (0, 0, -0.5), and (0, 0, 0.5) in meter, respectively. The web camera is attached at the origin of the coordinates, and it has two built-in motors thus it can pan and tilt to face the target source and capture the image of the sound sources in 3D space. During the source localization measurement, sound signals in time domain at four microphones are recorded and updated continuously. The program goes through the computer flow chart shown in Figure 2.2, and find out the source location. The Cartesian coordinates of the results is then transferred to the spherical coordinate to get the azimuth and inclination angles which corresponding to the panning and tilting angle of the built-in motors of the camera. Therefore the motors can rotate at the calculated panning and tilting angles, and make sure the target sound source is covered inside the image of the camera.



(a) Prototype device for orthogonal four-microphone set

(b) NI-9234 signal acquisition module

(c) NI-9162 carrier

Figure 3.1 Prototype device for the orthogonal four-microphone set model. (a) This device consists of four microphones, a web camera, a thermometer, and the data acquisition systems. (b) NI-9234 signal acquisition module. (c) NI-9162 carrier.

The non-orthogonal four-microphone device (Figure 3.2) uses the same thermometer, microphones, and signal conditioner, but utilized another microphone stand up and changes the positions of the microphones to approximately at four vertexes of a tetrahedron, which are (0, 0.73, 0), (-0.2, 0.1 -0.35), (-0.2, 0.1, 0.35), and (0.4, 0.1, 0) in meter in Cartesian coordinates, respectively. The microphone spacing of the second device is determined as 0.8 meter. The reason of changing the positions of the microphones is to build a tight frame thus the error of sound source localization is minimal and balanced in any direction. Moreover, the camera in this device is fixed at the origin and doesn't move. This is based on the consideration that the movement of camera may cause the shake of the microphone stand up and change the position of the microphones during the measurements thus have negative impact on the accuracy of the final source localization results. This device can gain more reliable results of source locations though the camera cannot catch the image of the source when the source falls outside the viewing angle of the camera. Moreover, the information of the source locations can always be found in the top

view and front view in the programming screen, and the precise locations are updated as well. The computer flowchart of this device is the same as the first device, which are shown in Figure 2.2.

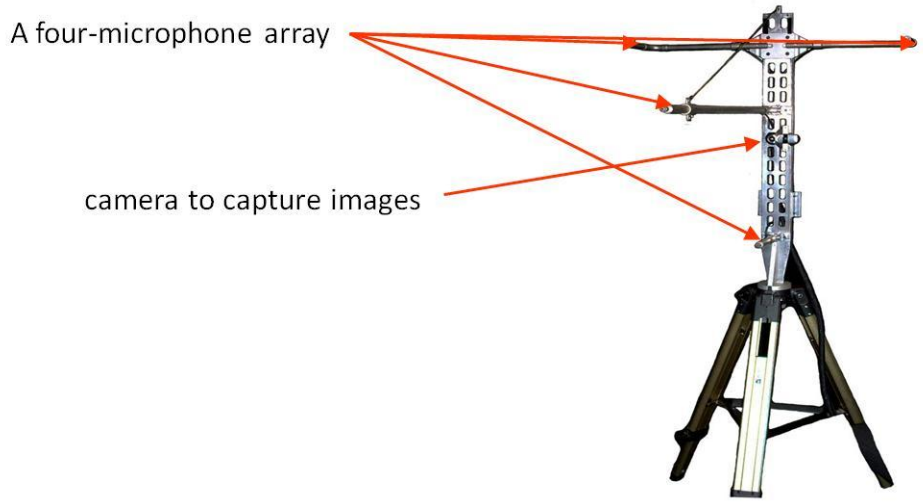


Figure 3.2 Prototype device for the non-orthogonal four-microphone set model. This device consists of four microphones, a web camera, and the data acquisition systems which is the same as Figure 3.1 (b) and (c).

The following experimental results in section 3.1.2 and 3.1.3 are from the screen captured by the orthogonal four-microphone device, while all the other results are from the non-orthogonal four-microphone device.

3.1.2 Case 1: Locating one sound source

Figure 3.3 demonstrates the screen captured during the real time sound source localization. A yellow crosshair is overlaid on the image captured by the camera to show the location of the target sound source. A precise location of the source in terms of Cartesian coordinates (x, y, z) is updated automatically in real time above the camera image. The two graphs below the camera image show the time domain signals and sound pressure level by

octave frequency bands at Channel #1. Two color maps on the right side of the image, namely, the front view and the top view, show the precise location of the source in 3D space and the sound pressure level in the environment. In Figure 3.3, the loudspeaker on the table is playing a typical party crowd sound, the time domain signals are recorded and updated every unity second, and overall frequency bands are considered. The four-microphone set can successfully locate the position of the sound source, which is the dominate sound in the environment, and the precise location of the loudspeaker is given above the image in the screen, as (4.87, – 0.97, –0.97) in meter.

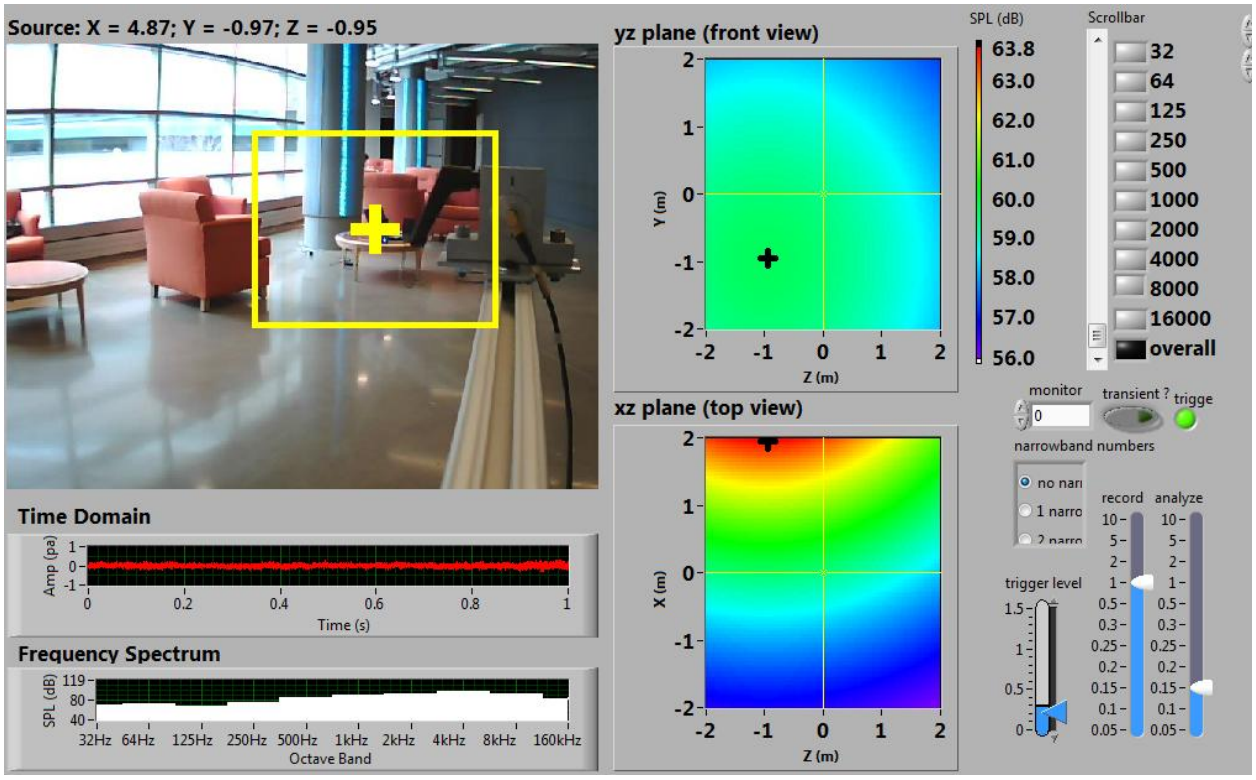
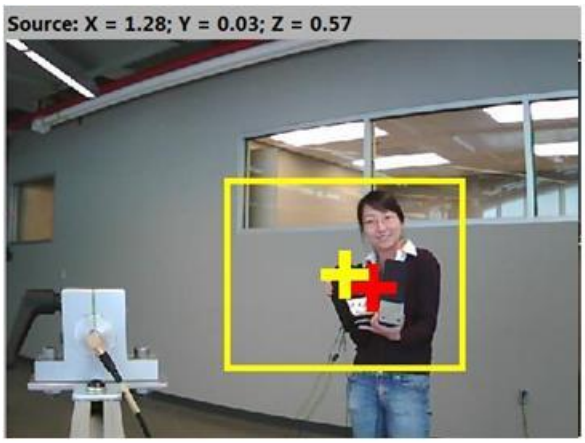


Figure 3.3 Experimental validations of the four-microphone set model. Experiments were conducted inside the hall way of Engineering Building. The yellow crosshair indicates the source location, and the precise location in the Cartesian coordinates (x, y, z) in meters that are shown above the image.

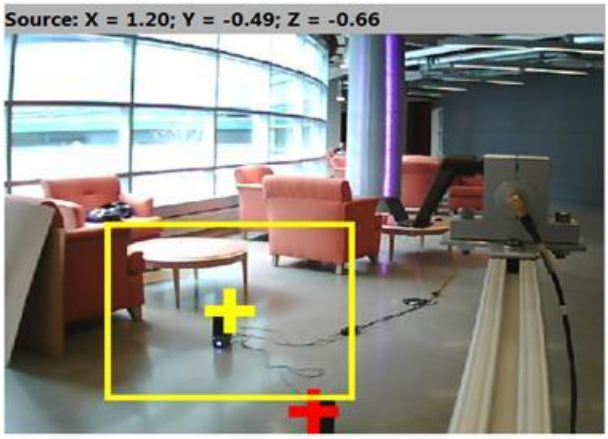
3.1.3 Case 2: Locating multiple incoherent sound sources

Figure 3.4 demonstrates the cases of locating multiple incoherent sound sources. The yellow crosshair shows the position of the target source, and the red crosshair indicates another source existing in the environment. Although only the coordinates of the target source indicated by yellow crosshairs are shown in the text bar above the image of the camera, the precise locations of the two sources are obtained, and the crosshairs point out the location of both sources simultaneously.

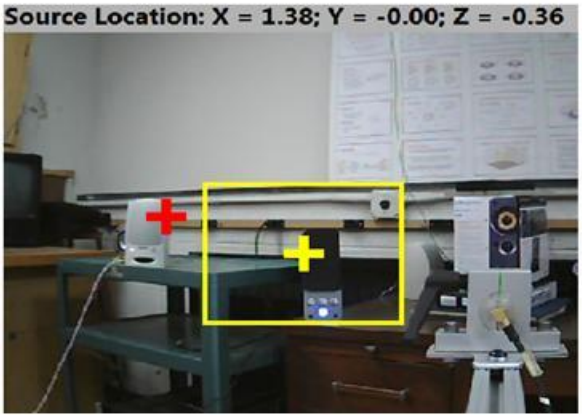
Figure 3.4 (a) and (b) shows the experimental results in the hall in the Engineering Building. One of the loudspeakers is considered as the target source and playing party crowd sound, while the other loudspeaker playing white noise as background noise. Figure 3.4 (a) shows the case when the two sources are next to each other, and Figure 3.4 (b) shows the case that the two sources have off-set in source range. The four-microphone device can successfully locate both sources. Figure 3.4 (c) and (d) are the localization results of the same sound sources but in a different environment in the AVNC Lab. The device can also locate both of the sources, which means that the device is applicable for various none-ideal environments.



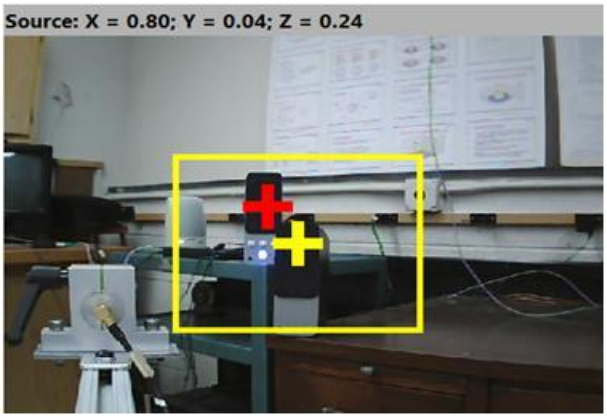
(a) Locate two sources in a hall



(b) Locate two sources in a hall



(c) Locate two sources in a small room



(d) Locate two tonal sounds in a small room

Figure 3.4 Experimental validations of locating multiple sound sources by the four-microphone set. Experiments were conducted inside the hall way and AVNC Lab in the Engineering Building. The yellow and red crosshairs indicate the source locations.

3.2 Error analysis and Empirical modeling for source localization

To further understand the underlying characteristics of the acoustic model based localization algorithm, error analysis was conducted on the localization experimental results by the non-orthogonal four-microphone set. The error of the results consists of two parts, the biased and random error. The error distribution is discussed in this section, and two semi-empirical models are developed to improve the accuracy of the results.

3.2.1 Error analysis of experimental results

The error in source localization is defined in Equation (2.15). A systematic test was conducted in the auditorium room in the Building of Engineering. Various sound sources were placed at 200 different positions around the four-microphone set. The distribution of the source positions covers 360 degree horizontally, three different heights, and the source range is up to five meters range.

Benchmark locations of the sources were measured by 3D sonic digitizer model 5230XL, which is a localization device employing the ultrasonic technologies. The 3D sonic digitizer has an ultrasonic gun and a receiver set. One can pinpoint the target with the ultrasonic gun and generate ultrasonic sound with it. The receiver recognizes the sounds generated by its producer thus can find out the geometry position in 3D space of the target. During the measurement of the benchmark location, the ultrasonic gun pinpointed at the center of the loudspeakers or other sound sources, generated ultrasonic sound, and the position can be obtained by the 3D sonic digitizer program. For an object within a radius of 4 meters, the error margin of this 3D sonic digitizer is ±2.5mm. As the localization error of the 3D digitizer is much less than the expected error of the present approach, the location gained by 3D digitizer can be considered as benchmark position.

Once the benchmark locations are obtained, the localization errors of the four-microphone can be calculated using Equation (2.15). The results of the systematic test have shown that the accuracy of the localization result is almost independent of the sound type of the source, and the error of the source localization can be considered as two parts: the random and biased error.

Figure 3.5 illustrates the 100 times calculated results by the localization methodology when the loudspeaker is placed at one of the contour positions in the systematic test. In this case, the position of the loudspeaker is at (1.88, – 0.20, 0.15) in Cartesian coordinates in meter, called benchmark position, which is indicated by red dot in Figure 3.5, and it was playing white noise continuously. The four-microphone system gather the time domain signals measured at each channel every 0.5 second, take the 0.5 second data as the input of the source localization algorithm, and therefore calculate the precise location of the target sound source. The position of the 100 estimated results by the localization algorithm are indicated by blue circles in Figure 3.5, and a mean value of the 100 results was calculated and expressed by a black triangle. It can be seen from Figure 3.5 that the distribution of the calculated results concentrates tightly around the mean value, and the distance from any single calculated result to the mean value is much smaller than that from the mean value to the benchmark position.

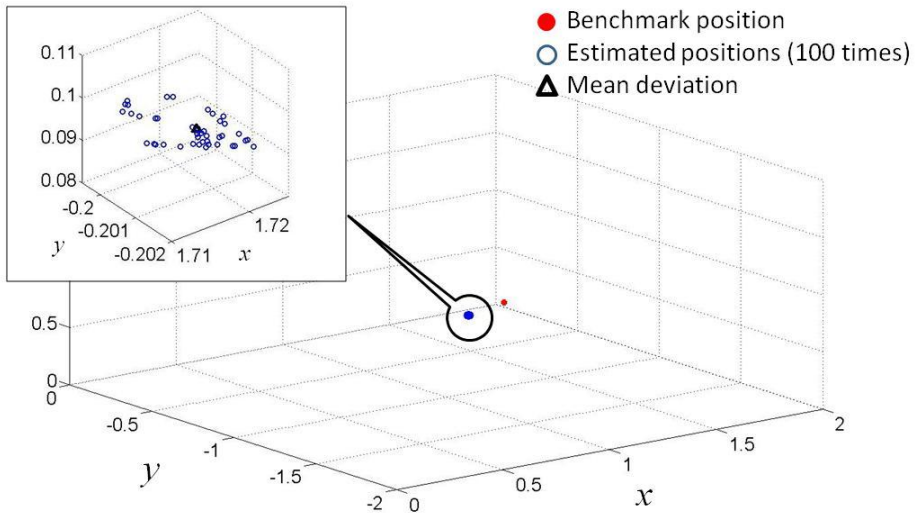


Figure 3.5 Localization results on one of the points. The source was placed at (1.88, –0.20, 0.15) in meter. Circles indicate 100 calculated source localization results, red dot indicates the benchmark location obtained by the 3D sonic digitizer, and the triangular is the mean of 100 calculated results.

The distribution of the sound source localization in this case can be demonstrated more clearly in the PDF contour of the source range distribution in Figure 3.6. The benchmark source range is indicated by the blue line, which is 1.8992 *m* from the origin. The mean value is at 1.7322 *m*, and the error of the mean value respect to the benchmark position is 8.8%, which indicates the biased error. On the other hand, the random error can be presented by the standard deviation of the source localization results, which is $\sigma^2 = 6.76 \times 10^{-6}\ m^2$. It is obvious that the biased error is the dominate one, and the value of the random error is much smaller than the biased error and can be ignored.

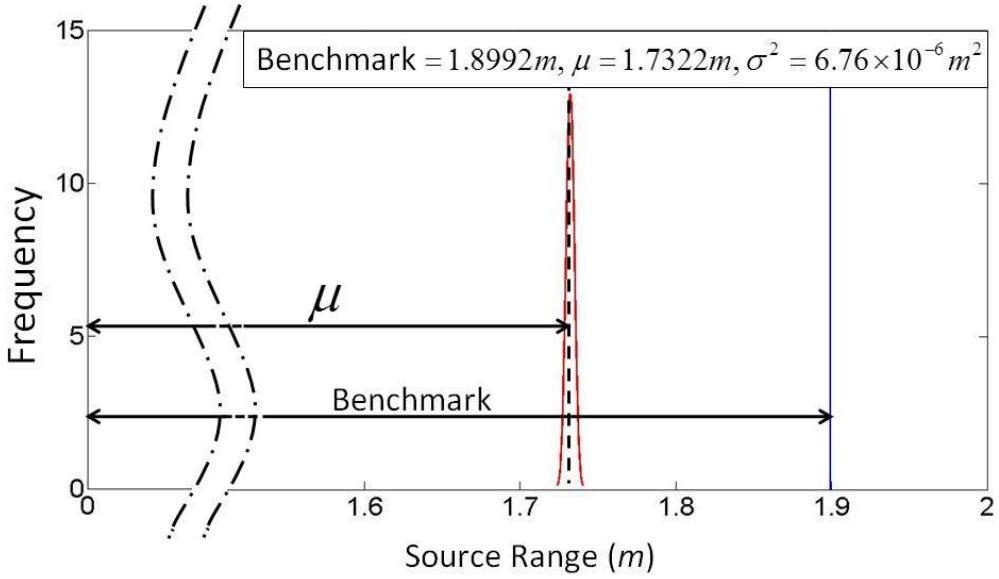


Figure 3.6 Normal distribution of the calculated source localization results at an arbitrarily selected point in space. A loudspeaker playing white noise was placed at (1.88, -0.20, 0.15) meter. The benchmark source range is 1.8992 meter, the mean source range based on 100 calculations is 1.7322 meter; and the variance of the calculated results is $6.76 \times 10^{-6}\ m^2$.

3.2.1 (a) Biased error

The error distributions of the 200 points in the systematic test show the same tendency as the particular case demonstrated in Figure 3.5 and 3.6. Therefore it can be concluded that the

biased error is the dominate one in the localization method, and the effort on analyzing and reducing the biased error can significantly improve the accuracy of the final results.

Comparing to the numerical simulation errors which are very small, the errors in the experimental validation are relatively high. This is reasonable because there are sound reverberation, reflection, and interferences of background noises in the none-ideal environment in the real world tests, while the numerical simulation is generated in an ideal environment, in which only the direct sound waves radiated from the sources is considered in free field.

The none-ideal environment has a negative impact on the accuracy of the TDOA estimation, because the sound reverberation, reflection, and background noise distort the waveforms of the sounds radiating in the field, thus the measured time domain signals has certain error comparing to the numerically mixed signals. The distortion of the waves brings the error in the cross-correlation calculation, thus reduce the accuracy of the TDOA estimation. As discussed in Chapter 2, the final result of the sound source localization is in terms of the TDOA, thus the error on TDOA has a direct effect on the accuracy of the localization results. The experimental results have shown that for the cases in which the error on TDOA is smaller than the others, the accuracy of the source localization result is relatively higher than the others. As the numerical model of sound source localization doesn't account the sound reverberation, reflection, and background noise, the environment in real world is different from the one in the modeling. Therefore the errors are biased.

3.2.1 (b) Random error

The random error are basically caused by the fluctuation of the sources and background noises, and the experimental results of the 200 points have shown the same tendency that the random error in the source localization methodology is much smaller than to the biased error.

More experiments have conducted with various measurement lengths, and the results show that the random error reduces when the length of the time domain signals increases. This is because the TDOA results is obtained by the cross-correlation graph, and the influences of the fluctuating in time domain signals on the cross-correlation results can be averaged out and eliminate if longer signals are accounted.

Figure 3.7 demonstrates the relation of the signal measurement length and the random error. The record length of the signal is varying from 0.1 to unity second, and the sound source location is calculated 100 times with each record length. The random error is represented by the variance σ^2 of the 100 calculated source ranges in m^2. It is obvious from Figure 3.7 that the variance σ^2 reduces linearly with the increasing of the record length. Therefore, increasing the record length of the measurement signals can effectively reduce the random error of the localization results, and location results gained by longer record length are more stable than those by short record length. However, the real-time running characteristics of the localization algorithm may not be guaranteed if the record length is long, and the cross-correlation calculation of longer recorded signals may cause computer system problem due to the memory limitation of the computer. Therefore, a middle value of the record length should be selected to maintain both the efficiency and the accuracy of sound source localization. Based on the results in the experimental validation, a record length 0.5 second is selected for the sound source localization.

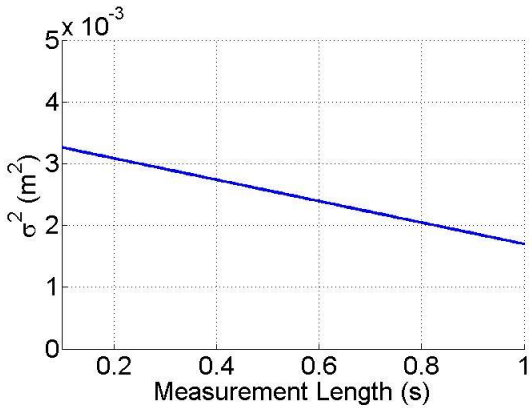


Figure 3.7 Impact of the measurement record length on the random error. The measurement record length was chosen from 0.1 second to one second. Results show that the random error decreases when the measurement record length increases.

3.2.2 Empirical modeling

As discussed in the Section 3.2.1, the error of the sound source localization consists of two parts, the biased error and the random error. The biased error is the dominate one, and it is caused by the sound reverberation, reflection, and background noises in the non-ideal environment. However, it is extremely difficult to improve the numerical modeling of sound source localization by accounting the sound reverberation, reflection and background noises in the environment, because the environment for each particular case is different and the background noise is unpredictable.

However, on the other hand, it is possible to generate empirical modeling to modify the localization results and eliminate the error. Therefore, in this section, we aim at use two different methodologies to build up semi-empirical models for particular non-ideal environment, and therefore reduce the error. In other words, when facing a new environment, one can always use a group of directly measured source locations and their corresponding benchmark positions to find a set of semi-empirical model, which concludes the information of this particular environment. Once the semi-empirical model is built up, one can add the model to the original sound source

localization algorithm, and use them to locate any source in this environment with a more accurate result than before. The procedure of finding the semi-empirical model is called “calibration” of this particular environment. The calibration procedure is simple and practical, and it can be used without any prior knowledge of the background noise and the environment, such as the room size, furniture, materials of the wall, etc., which may affect the sound reverberation and reflection.

As discussed in Section 3.2.1, the random error are much smaller than the biased one and can be ignored, the calibration of the environment only focus on eliminate the biased error. For each of the 200 points in the auditorium room during the experimental validation, the random error are eliminated by taking averaging of 100 times results along time domain signals, and every set of signals is 0.5 second long. Next, the 200 points results are used to evaluate the feasibility of the procedure of the environment calibration and evaluate the source localization results calibrated by the semi-empirical models.

The 200 points are separated randomly to two groups of data, each has 100 points. The first group is called the training group, and it is used to calibrate the environment of auditorium room and build the semi-empirical model; the second group is called testing group. Once the semi-empirical model is obtained, they are applied to the testing group, in which the 100 points are totally different from the 100 points in the training group. The calibrated locations with the semi-empirical model in the testing group are compared with the directly measured source locations as well as the benchmark positions; therefore the efficiency of the empirical modeling is evaluated.

In this dissertation, the semi-empirical model is generated by two different methodologies, one is Least Square (LS) method, and the other is ANFIS method.

3.2.2 (a) Empirical modeling using Least Square (LS) method

LS method attempts to compute the optimal polynomial curve which is best fitting with the input data and the benchmark data. It is a commonly used methodology for the development of empirical simulation [171-176]. In the case in this dissertation, the input data of the LS method are the directly calculated sound source locations in the training group, while the benchmark data are their corresponding benchmark positions. The goal of the LS method is to find the optimal polynomials which are in terms of the input data, and the values of the polynomials are nearest to the benchmark data.

For a general LS method, assume there are N set of input data, and N set of corresponding benchmark output. The order of the polynomials is j, and the position in 3D space concludes three variables in the Cartesian coordinates, which are the value on the x, y, and z axes. The non-zero terms in the polynomials in the LS model can be calculated as $N_j = (j + 1)^3$. The LS formulation is expressed as:

$$[R]_{3\times N} = [A]_{3\times N_j}[r]_{N_j\times N} \tag{3.1}$$

The matrix [r] consists of the N set of input data which are the three variables directly measured in the source localization method and the products with up to the j^{th} order of them, thus it has N_j rows and N columns. The matrix [R] is the output data of the LS method, which are N set of data, and each set has three variables indicating the values on the x, y, and z axes, thus it consists of three rows and N columns. The matrix [A] is the parameter matrix to be determined in the calibration procedure.

As mentioned above, the goal of the LS method is to find the best fitting matrix [A] which can produce the matrix [R] optimal to the benchmark output. Therefore the Equation (3.1) can specified to used in the training group,

$$\left[R^{tr}\right]_{3\times N} = \left[A\right]_{3\times N_j}\left[r^{tr}\right]_{N_j\times N} \tag{3.2}$$

Here the superscript "*tr*" means that the data is from the training group. The benchmark output [$R^{tr}{}_B$] is known, and it can be formatted as the same dimension as the matrix [R^{tr}]. Thus one can find the matrix [*A*] by attempting the minimal value of the L2 norm of the two matrixes [R^{tr}] and [$R^{tr}{}_B$]:

$$\mathbf{min}\left\|\left[R_B^{tr}\right]_{3\times N} - \left[R^{tr}\right]_{3\times N}\right\|^2 \tag{3.3}$$

Insert Equation (3.1) into Equation (3.2), thus

$$\mathbf{min}\left\|\left[R_B^{tr}\right]_{3\times N} - \left[A\right]_{3\times N_j}\left[r^{tr}\right]_{N_j\times N}\right\|^2 \tag{3.4}$$

Take the partial differential of Equation (3.3) respect to [*A*]:

$$\frac{\partial}{\partial A_{ij}}\left\|\left[R_B^{tr}\right]_{3\times N} - \left[A\right]_{3\times N_j}\left[r^{tr}\right]_{N_j\times N}\right\|^2 = 0 \tag{3.5}$$

and

$$\left[R_B^{tr}\right]_{3\times N} = \left[A\right]_{3\times N_j}\left[r^{tr}\right]_{N_j\times N} \tag{3.6}$$

Thus [*A*] can be calculated as follows:

$$\left[A\right]_{3\times N_j} = \left[R_B^{tr}\right]_{3\times N}\left[r^{tr}\right]^T_{N\times N_j}\left(\left[r^{tr}\right]_{N_j\times N}\left[r^{tr}\right]^T_{N\times N_j}\right)^{-1} \tag{3.7}$$

Once the matrix [A] is determined, it can be used in the general equation of the LS method shown in Equation (3.1) to calibrate any input data and a set of optimal output data can be obtained.

Here matrix [*A*] is gained in Equation (3.5), [*r*] can be any set of input data, and the [*R*] is the corresponding output which is calibrated by the LS method. In this dissertation, Equation

(3.6) is further applied to the testing group to validate the efficiency of the empirical modeling, and the calibrated results are compared with the benchmark positions in the next section.

Another parameter need to be defined in this model is the order of the polynomials. As the relation between the input and output of the model is unknown, the order can be freely chosen in the semi-empirical model. A low order number may not be enough to describe the relation between the input and output thus the optimal fitting curve cannot be achieved. After the observation of the data gathered in the experimental validation, it can be found that the relation between the input and benchmark in this case is more complex than a linear function, thus the order number should be larger than one. On the other hand, a high order can find the parameter matrix [*A*] which fits the training group perfectly, but fail to calibrate the localization results in the testing group. As a compromise, the 2nd order is chosen in this semi-empirical model. Hence j=2, and $N_j = (2 + 1)^3 = 27$. The Equation (3.5) can be rewritten as:

$$[A]_{3\times 27} = [R_B^{tr}]_{3\times N}[r^{tr}]_{N\times 27}^T\left([r^{tr}]_{27\times N}[r^{tr}]_{N\times 27}^T\right)^{-1} \tag{3.8}$$

where

$$[r^{tr}]_{N\times 27}^T = \begin{bmatrix} \alpha_{1,1} & \alpha_{2,1} & \cdots & \alpha_{27,1} \\ \alpha_{1,2} & \alpha_{2,2} & \cdots & \alpha_{27,2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,N} & \alpha_{2,N} & \cdots & \alpha_{27,N} \end{bmatrix}_{N\times 27} \tag{3.9}$$

Each row in the matrix [r^{tr}] gives the coordinates value and their products up to the 2nd order of one point in the training group. For example, the i^{th} row in matrix includes 27 items, namely, $\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$,…, $\alpha_{27,i}$, and they indicate the information of the i^{th} point (x_i, y_i, z_i) in the database.

$$\begin{aligned} &\alpha_{1,i}=1,\ \alpha_{2,i}=x_i,\ \alpha_{3,i}=y_i,\ \alpha_{4,i}=z_i,\ \alpha_{5,i}=\alpha_{2,i}\alpha_{3,i},\ \alpha_{6,i}=\alpha_{2,i}\alpha_{4,i},\ \alpha_{7,i}=\alpha_{3,i}\alpha_{4,i},\ \alpha_{8,i}=\alpha_{2,i}^2,\ \alpha_{9,i}=\alpha_{3,i}^2,\ \alpha_{10,i}=\alpha_{4,i}^2, \\ &\alpha_{11,i}=\alpha_{2,i}\alpha_{3,i}\alpha_{4,i},\ \alpha_{12,i}=\alpha_{2,i}^2\alpha_{3,i},\ \alpha_{13,i}=\alpha_{2,i}^2\alpha_{4,i},\ \alpha_{14,i}=\alpha_{2,i}\alpha_{3,i}^2,\ \alpha_{15,i}=\alpha_{3,i}^2\alpha_{4,i},\ \alpha_{16,i}=\alpha_{2,i}\alpha_{4,i}^2,\ \alpha_{17,i}=\alpha_{3,i}\alpha_{4,i}^2, \\ &\alpha_{18,i}=\alpha_{2,i}^2\alpha_{3,i}^2,\ \alpha_{19,i}=\alpha_{2,i}^2\alpha_{4,i}^2,\ \alpha_{20,i}=\alpha_{3,i}^2\alpha_{4,i}^2,\ \alpha_{21,i}=\alpha_{2,i}^2\alpha_{3,i}\alpha_{4,i},\ \alpha_{22,i}=\alpha_{2,i}\alpha_{3,i}^2\alpha_{4,i},\ \alpha_{23,i}=\alpha_{2,i}\alpha_{3,i}\alpha_{4,i}^2, \\ &\alpha_{24,i}=\alpha_{2,i}\alpha_{3,i}^2\alpha_{4,i}^2,\ \alpha_{25,i}=\alpha_{2,i}^2\alpha_{3,i}\alpha_{4,i}^2,\ \alpha_{26,i}=\alpha_{2,i}^2\alpha_{3,i}^2\alpha_{4,i},\ \alpha_{27,i}=\alpha_{2,i}^2\alpha_{3,i}^2\alpha_{4,i}^2 \end{aligned} \tag{3.10}$$

Similarly, $\left[R_B^{tr}\right]$ can be written as

$$\left[R_B^{tr}\right]_{3\times N} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,N} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,N} \\ \beta_{3,1} & \beta_{3,2} & \cdots & \beta_{3,N} \end{bmatrix}_{3\times N} \tag{3.11}$$

and the i^{th} row in matrix $\left[R_B^{tr}\right]$ indicates the three coordinates of benchmark source location.

$$\beta_{1,i} = x_i^B,\ \beta_{2,i} = y_i^B,\ \beta_{3,i} = z_i^B \tag{3.12}$$

3.2.2 (b) Empirical modeling using ANFIS

Another method used to generate the semi-empirical model is the ANFIS system. The ANFIS method employs the concept of fuzzy logic, which is widely used in machine controls and other applicable problems [177-191]. It estimates the judgment of human operator, thus no longer only use “true” or “false”, but “partially true” to define the contribution of any input data. And it shares the idea with the LS method that each of the output elements is affected by all the input data, but the underlying algorithm is more complex than the polynomial functions used in the LS method.

Similar to the environment calibration procedure by the LS method, the ANFIS use the three variables (x, y, z) of the directly calculated positions in the training group as the input of the model, and their corresponding benchmark positions are considered as the target of the output.

As we assume that the values on the three coordinates influence each other, therefore totally three ANFIS models are generated, the inputs of the three models are all the directly calculated three coordinates of the source position, while the outputs of are the calibrated x, y, and z value respectively. The procedures of generating any of these three semi-empirical model and find out the parameters have several steps:

1) First, define the structure of the ANFIS system, which consists of the input membership functions (MF), type of rules, and output MFs (shown in Figure 3.8). The variables at the input MFs are the directly measured data by the original source localization algorithm, the values of these membership functions are forwarded to the rules. There are three types of rules: “and”, “not”, and “or”. All the rules in this ANFIS system are set as “and” rules. The output MFs are in terms of the results from the rules, and combining of their value leads to the final result of the calibration model, which is desired to be the benchmark position.

2) Substitute the directly measured data in the training group to the input MFs and take one of the coordinates of benchmark position (for example, the *x* axis) as the target output value of the ANFIS system.

3) Conduct the Least Square as well as the back propagation algorithm to estimate all the unknown parameters in the MFs.

4) Package the ANFIS system structure and all the value of parameters determined in the last step as the semi-empirical model for further use.

In practice, the ANFIS structure and the determination of the parameters can be conducted with the help of Matlab Fuzzy Toolbox. [192]. Figure 3.8 gives an example of the ANFIS structure of the semi-empirical model for calibrating the x-axis of the sound source location. The inputs of this structure are the Cartesian coordinates (*x*, *y*, *z*) of the directly calculated sound source location, and are indicated by three black dots on the left column. Totally nine input MFs are employed, three for each input, and they are expressed by white dots, listing at the second column on the left. 27 rules were utilized, and they are indicated by the blue dots in Figure 3.8. The output MFs are located right to the rules, and combination of the output

of these MFs leads to the final calibrated result value of the x-axis, which are suppose to be as close as possible to the benchmark.

All the MFs selected in this structure are the generalized bell-shaped MF (GBELL MF), and relations of its input variable and output value can be expressed as [192]:

$$output_{GBELLMF} = f\left(input_{GBELLMF}; a_1, a_2, a_3\right) = \frac{1}{1+\left|\frac{input_{GBELLMF} - a_3}{a_1}\right|^{2a_2}} \tag{3.13}$$

where $input_{GBELLMF}$ is the value of input variable which is obtained from the last step in the flowchart, and $output_{GBELLMF}$ is the output value. a_1, a_2 and a_3 are three parameters and need to be determined in the training procedure.

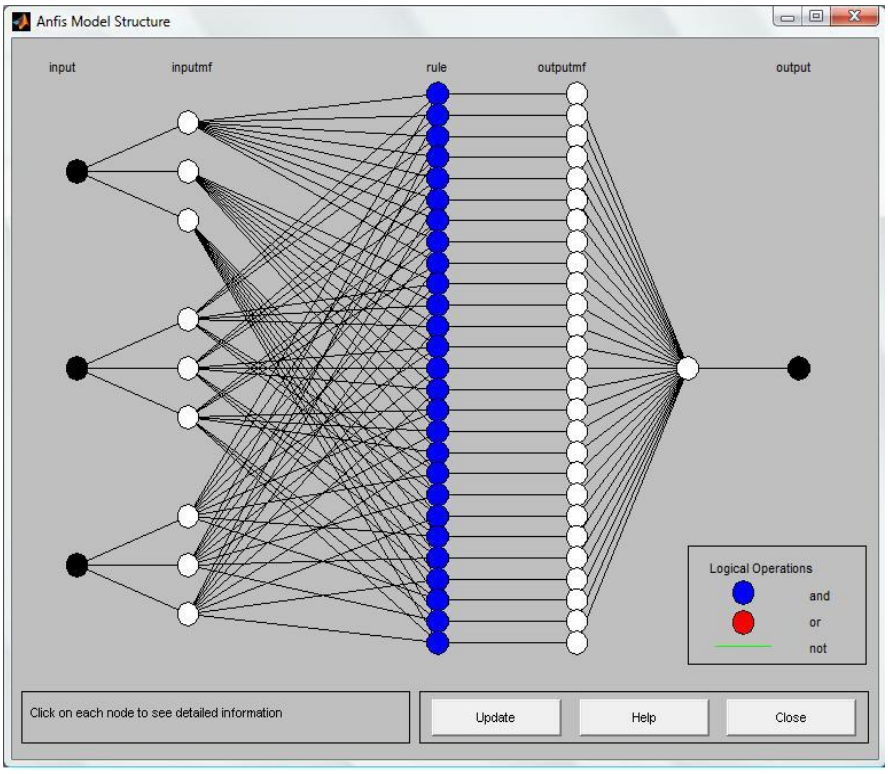


Figure 3.8 ANFIS Model Structure for calibrate x-axis of the source location. The input of this ANFIS model consists of three element, which are *x*, *y* and *z*-axis of the directly measure results. The output of this model is the calibrated *x*-axis value.

The ANFIS structures for the values in *y* and *z* coordinates are similar to the x one. And once all the parameters of the *x*, *y*, and *z* ANFIS structures are determined, the training procedure is finished and the semi-empirical model is ready to be used, and it can calibrate any directly measured sound source location and obtain the calibrated location with its error reduced.

Both of the LS and ANFIS semi-empirical models are developed to calibrate the directly measured sound source location, and aim at eliminate the bias error occurs on the locations, which are caused by the non-ideal environment. The parameters in these two models are determined using a large amount of data in the training group, including the directly measured sound source locations and their benchmark positions. The goal of training procedure is to find the parameters with which the models can calibrate the sound sources and the outputs are closest to the benchmark position. Once the parameters are obtained, they are packaged into the models thus the semi-empirical models can be further used at any directly measured location and gives a calibrated one with better accuracy. To check the effectiveness of the semi-empirical models, a different group of data, called the testing group, is applied to the models, and the directly measured locations and the calibrated locations are compared to the benchmark in this group to see if the calibrated ones have higher accuracy than the directly measured ones.

3.2.3 Experimental validations of empirical models

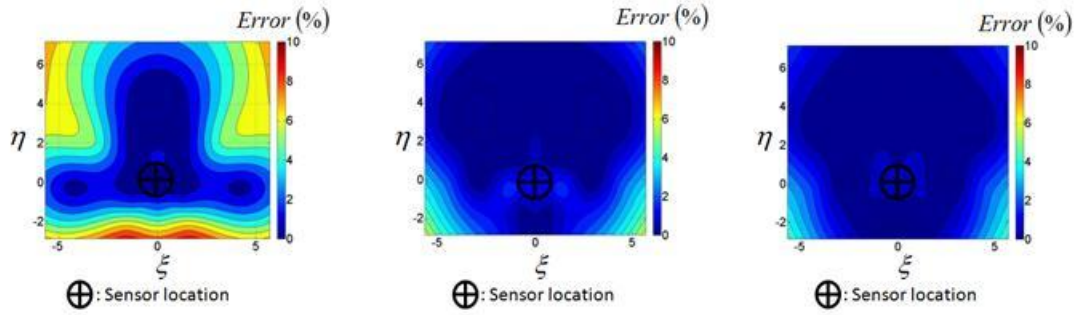
Figure 3.9 illustrates the comparisons of three types of results in the testing group: the directly measured sound source locations and the calibrated locations by LS and ANFIS method, respectively. As introduced in Chapter 2, the error and source range is influenced by the microphone spacing. The larger microphone spacing is, the more accurate source location is gained and larger source range can be reached. This is reasonable because when the microphone spacing increase, the TDOAs increases accordingly, thus when a same amount of error value

happens in the TDOA estimation, its contribution to the TDOA decreases, thus the impact on the source localization results is eliminated. In this section, the ratio of the source range and microphone spacing is also used in Figure 3.9, namely, dimensionless parameters η, ξ, and h are presented instead of the Cartesian coordinates x, y, and z in meter, where $\eta = x/d$, $\xi = y/d$, and $h = z/d$, d indicates the microphone spacing. Once the dimensionless source range is analyzed, one can always change the microphone spacing to get a matching source range as desired.

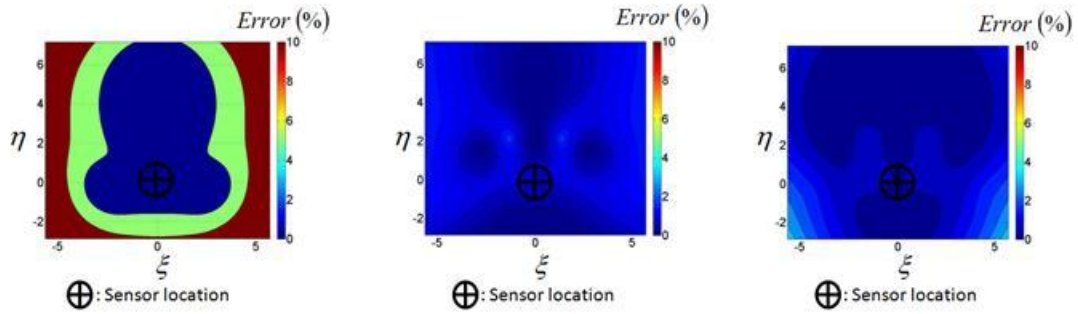
As it is difficult to demonstrate the error in a three-dimensional space, Figure 3.9 chooses to plot the top view of the error in layers. Each layer has a specific height h, and covers a $12\xi \times 10\eta$ area in horizontal plane. Figure 3.9 (a), (b), and (c) show the results at $h = -0.36$; Figure 3.9 (d), (e), and (f) are at $h = 0.21$; and Figure 3.9 (g), (h), and (i) are at h=0.7.

Results show that the errors of the directly measured source locations can be as high as 10% inside the testing area, and only about half area can keep the error under 1%, as shown in Figure 3.9. Errors of calibrated locations by either of the two methods are reduced as expected. In particular, the errors of the calibrated locations can be controlled under 2% when $h = -0.36$ and 0.21, and under 3% when $h = 0.7$. Moreover, for $h = -0.36$, 0.21, and 0.7, the percentage of the calibrated point with an error less than 1% in the testing group are 78%, 85%, and 60%.

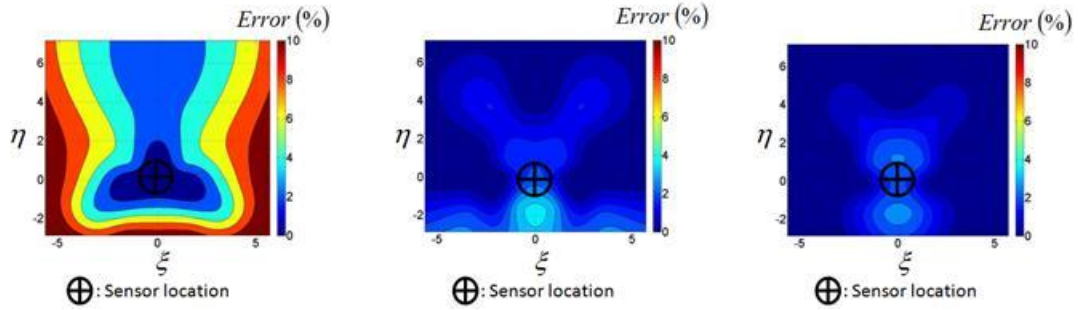
What is more, the calibrated results by the LS and ANFIS methods are similar to each other. The error distributions in three heights look the same, and the statistical similarity of the results provided by LS and ANFIS is 99.57%. This indicates that the two semi-empirical models have the same effect on the directly measured source locations, though their underlying algorithms are different.



(a) h=-0.36, w/o calibration (b) h= − 0.36, Calibrated by LS (c) h= − 0.36, Calibrated by ANFIS



(d) h=0.21, w/o calibration (e) h=0.21, Calibrated by LS (f) h=0.21, Calibrated by ANFIS



(g) h=0.7, w/o calibration (h) h=0.7, Calibrated by LS (i) h=0.7, Calibrated by ANFIS

Figure 3.9 Comparison of the results without calibration, with LS calibration and ANFIS calibration. The source was places on a plane at h = –0.36 in (a), (b), and (c); at h = 0.21 in (d), (e), and (f); and h = 0.7 in (g), (h), and (i). (a), (d) and (g) show the error of results without calibration, the errors in some of the area are as high as 10%. (b), (e), and (h) show the error of the results after LS calibration. (c), (f), and (i) show the error of the results after ANFIS calibration.

Table 3.1 gives the error averages and the programing time for the original approach and the calibrated results by two methods. The average error of the directly measured results are as high as 2.77%, and those by LS and ANFIS calibration are reduced to 0.59% and 0.45%, respectively.

Table 3.1. Summary of the spatial-averaged errors for the original approach vs. those of the LS and ANFIS based semi-empirical models.

	Original Approach	After LS Correction	After ANFIS Correction
Spatial-averaged Error	2.77%	0.59%	0.45%
Model build-up time	0	0.003443 (second)	3 (seconds) by using Matlab Fuzzy Toolbox
Processing time	T	T+3.7273e-007 (second)	T+6.7127e-004 (second)

When with the data of the training group, the time cost for building up the semi-empirical modal by LS is 0.003443 second, and that for ANFIS is about three seconds. Numerically, the model build-up time of building ANFIS semi-empirical model is 1000 times longer than that of LS. However, both of the build-up time is short and acceptable.

Assume T is the processing time of localization algorithm, which includes signal pre-processing, the estimation of TDOAs, solving equations to get a pair of solutions, and selection of the correct solution as the directly measured source location. The time T is dependent on the type and length of the sounds, the signal pre-processing setup, speed of the computer processor, etc. As T only counts the time consuming of the localization algorithm, thus is unconcerned in the calibration procedure, and the three columns share the same value of T. The processing time to gain a LS calibrated location is T+3.7273e-007 second, which means that the time cost to get a calibrated result with a directly measured location in hand is 3.7273e-007 second, while that for the ANFIS is 6.7127e-004 second, which is 1000 times longer than that by LS. Nevertheless, both processing time mentioned above are short enough to be ignored, and adding the calibration procedure doesn't influence the real-time characteristic of the localization method.

The information of the source range coverage can also be observed in Figure 3.9. For the localization results without calibration, the source range can cover η from negative unity to seven, and ξ from negative three to three, any source range extends this coverage cannot have its

accuracy guaranteed, namely, the results are not reliable. The source range coverage after calibration are increased to an area of $12\xi \times 10\eta$, where any source falls inside negative six to six in ξ and negative three to seven in η can be correctly located with the error controlled under 2%.

It is emphasized that these two semi-empirical models are designed to calibrate the sound source locations measured in the auditorium room in the Engineering Building in Wayne State University, where the data of training group were gathered. They may not be applicable to other environment. When the environment is changed, it is better to conduct the experiment first to get a group of training data thus the parameters in the semi-empirical models can be optimal determined. For the tests conducted in this example, the procedure of gathering data, which insists of the training and testing group, cost about four to five hours, though one can take less points in the training group in various cases thus the time of experiment can be shorted.

3.3 Experimental validation for six-microphone set

The improved model with redundancy check is also tested in the experiments. The experimental setup consists of one thermometer, six B&K quarter-inch condenser microphones, 8-Channel Dynamic Signal Acquisition Module NI-4472 inside Carrier PXI-1033 from the National Instrument, and a web camera. Similar to the four microphone setup, the thermometer is used to measure the environment temperature thus the speed of sound can be estimated, and the web camera captures the images in real time and can give the operator the information of the sound sources. The six microphones are located at the same positions introduced in the numerical simulation in Chapter 2, which are (-1.3856, 0.4000, 0), (0, 0, 0), (1.3856, 0.4000, 0), (-1.3856, 0.4000, 0.8000), (0, 0, 0.8000), and (1.3856, 0.4000, 0.8000) in meter in the Cartesian coordinates.

(a) Prototype device for six-microphone set

(b) NI-4472 signal acquisition module

(c) PXI-1033 carrier

Figure 3.10 Prototype device for the six-microphone set model. (a) This device consists of six microphones, a web camera, a thermometer, and the data acquisition systems. (b) NI-4472 signal acquisition module. (c) PXI-1033 carrier.

Multiple types of sound sources and environment were tested to validate the localization by the six microphone setup. And the experimental results have shown that the characteristics of the results by this setup coincide with the numerical simulation. The capability of localization is stronger when the sources is located in the optimal zone, the accuracy is better than that of four microphone set. Most importantly, the source range coverage of the six microphone setup is larger than that by four microphone set; therefore it is not necessarily conduct the calibration procedure, but still can locate a sound source which is far away from the device.

Figure 3.11 demonstrates a group of screen prints in the test. In Figure 3.11 (a) and (b), two loudspeakers were playing noises dominating at different frequency bands in the hall of the Engineering Building in Wayne State University. One source is lower than 2000 Hz, while the other is higher than 5000 Hz. The locations of the two sources are indicated by yellow and red

crosshairs, respectively. The six microphone set can locate both of them even they are next to each other.

(a) Sounds from two speakers separated at certain distance

(b) Sounds from two speakers next to each other

(c) Airflow noise from a hose

(d) Noise from an operating cutting machine

Figure 3.11 Experimental validations of locating multiple sound sources by the six-microphone set. Experiments were conducted inside the hall way and Machine Shop in the Engineering Building. The yellow and red crosshairs indicate the source locations.

Figure 3.11 (c) and (d) show the sound source localization in the Machine Shop in the Engineering Building, which is a noisy environment. Figure 3.11 (c) illustrate the case of locating the airflow noise from a hose, and Figure 3.11 (d) shows the program locating the noise from a cutting machine. Sources can also be detected and located in Machine Shop, as indicated by the yellow crosshairs and the (*x*, *y*, *z*) coordinates above them.

Figure 3.12 and 3.13 are also snapshots from the running program. The images at the top right corner of the figures are recorded by another camera as a benchmark. In this test, the program is designed to automatically switch between the image overlapped by the crosshair indicating the location, as shown in Figure 3.12, and the top view of the source path in space, as shown in Figure 3.13. The effectiveness of the localization is demonstrated not only the direction by overlapping the crosshair on the camera image, but also the source ranges by the top view. The yellow dots in the top view in Figure 3.13 tell the walking path. Some of the yellow dots are at the bottom of the top view, which has source ranges about six times as the microphone spacing. Note that the in this test no calibrated procedures were hold, and the program can still cover a large source range.

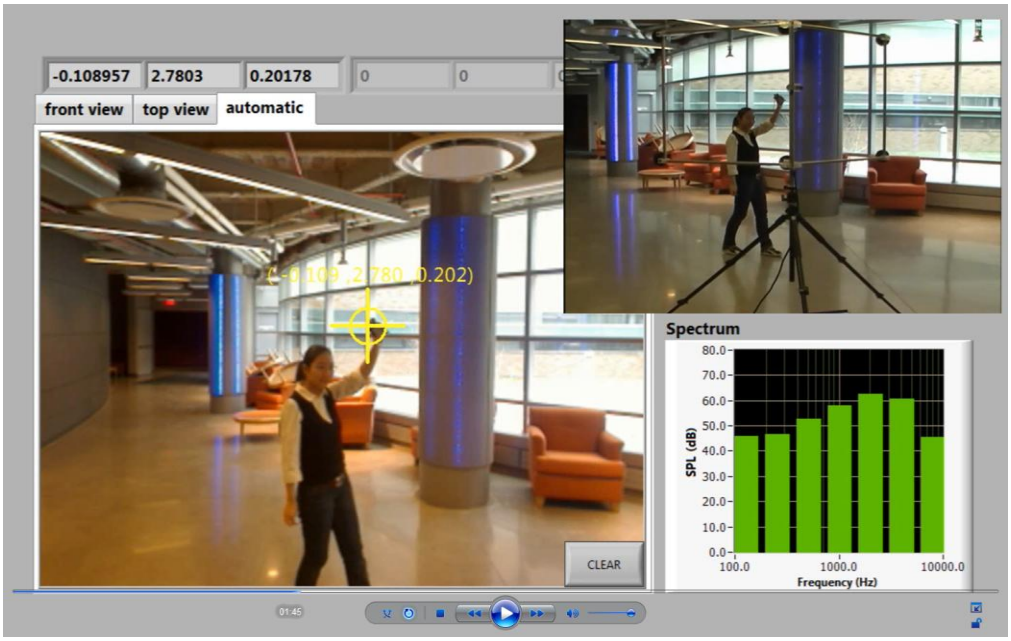


Figure 3.12 Experimental validations of tracking and tracing sound source by the six-microphone set (Camera View). Experiments were conducted inside the hall way in the Engineering Building. The yellow crosshair indicates the source location.

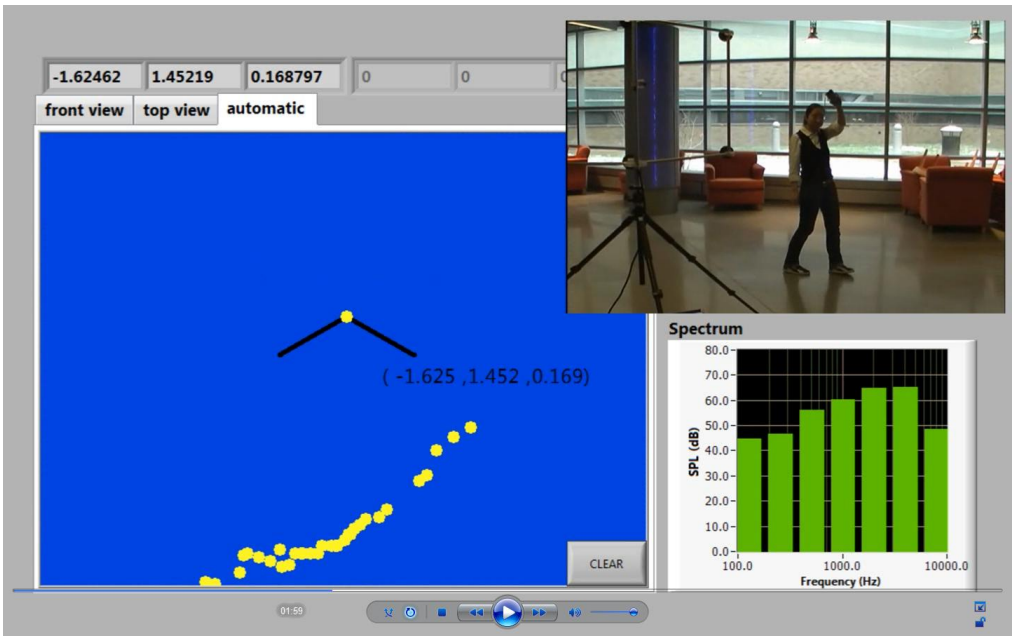


Figure 3.13 Experimental validations of tracking and tracing sound source by the six-microphone set (Top View). Experiments were conducted inside the hall way in the Engineering Building. The yellow dots indicate the location history of the sources, which reconstruct the path of the travelling source.

CHAPTER 4

SOUND SOURCES SEPARATION

This chapter is concerned with the development and validation of PSS for sound sources separation.

4.1 Theory of Point Source Separation (PSS)

Similar to the sound source localization algorithm introduced in Chapter 2, PSS assumes that sounds are radiated from point sources in free field. In addition, PSS requires the information of time domain signals at each microphone as well as the locations of the sources and microphones as its input. With this information, PSS can separate the signals at individual sources, regardless of their types. Theoretically, PSS can handle any mixed sounds as long as the number of measurement channels is equal to or larger than that of sound sources. However, an ideal free field does not exist in practice and sound reflection and reverberation may affect the accuracy of PSS results.

In this Chapter, N point sources and M microphones are used in numerical simulations. The vectors $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$ are used to denote the positions of the n^{th} source and m^{th} microphone in 3D space, respectively, where $1 \le m \le M$, $1 \le n \le N$, and $N \le M$. R_{nm} indicates the distance from the n^{th} source to the m^{th} microphone. Therefore, the acoustic pressure at the m^{th} microphone is the sum of the acoustic pressures emitted by N sources that can be expressed as:

$$p(x_m, t) = \sum_{n=1}^{N} \frac{s\left(x_n, t - \frac{R_{nm}}{c}\right)}{R_{nm}} \tag{4.1}$$

where $s(\boldsymbol{x}_n, t - R_{nm}/c)$ represents the source strength of the n^{th} source and $t - R_{nm}/c$ implies the retarded time. Note that PSS is based on an assumption of a causal process. So p is identically zero when $t < 0$.

On the other hand, the Fourier transform of the time domain signal at the m^{th} microphone $p(x_m,t)$ can be expressed as:

$$P(x_m,\omega) = \int_0^\infty p(x_m,t) e^{i\omega t} dt, \tag{4.2}$$

and its inverse Fourier transform is given by:

$$p(x_m,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(x_m,\omega) e^{-i\omega t} d\omega \tag{4.3}$$

Substituting Equation (4.1) into (4.2) yields the Fourier transform of the signal at the m^{th} measurement point:

$$P(x_m,\omega) = \int_0^\infty \sum_{n=1}^{N} \frac{s\left(x_n, t - \frac{R_{nm}}{c}\right)}{R_{nm}} e^{i\omega t} dt = \sum_{n=1}^{N} \frac{e^{jkR_{nm}}}{R_{nm}} \int_0^\infty s(x_n,t) e^{i\omega t} dt, \tag{4.4}$$

where $G(x_m|x_n;\omega) = \dfrac{e^{jkR_{nm}}}{R_{nm}}$ is known as the free-space Green's function [169], and the Fourier transformation of the n^{th} source is $S(x_n,\omega) = \int_0^\infty s(x_n,t) e^{i\omega t} dt$, ($1 \le n \le N$). Thus Equation (4.4) can be rewritten as

$$P(x_m,\omega) = \sum_{n=1}^{N} G(x_m|x_n;\omega) S(x_n,\omega). \tag{4.5}$$

The positions of the sources and microphones are known, thus $P(x_m,\omega)$ and $G(x_m \mid x_n; \omega)$ can be calculated, and the unknown source strength $S(x_n,\omega)$ can be determined. Once this is done,

the desired source signal $s(\mathbf{x}_n, t)$ can be obtained by taking an inverse Fourier transformation (see Equation (4.3)).

Equation (4.5) can be generalized using the following matrix form:

$$\left\{P\left(x_m,\omega\right)\right\}_{M\times 1}=\left[G\left(x_m\middle|x_n;\omega\right)\right]_{M\times N}\left\{S\left(x_n,\omega\right)\right\}_{N\times 1} \tag{4.6}$$

Note that the transfer matrix $\left[G\left(x_m\middle|x_n;\omega\right)\right]_{M\times N}$ is not a square matrix. Accordingly, $\{S(x_n,\omega)\}_{N\times 1}$ must be solved by taking a pseudo inversion of Equation (4.6),

$$\left\{S\left(x_n,\omega\right)\right\}_{N\times 1}=\left[G\left(x_n\middle|x_m;\omega\right)\right]^{\dagger}_{N\times M}\left\{P\left(x_m,\omega\right)\right\}_{M\times 1} \tag{4.7}$$

where the symbol "†" indicates the pseudo inversion of a matrix, and

$$\left[G\left(x_n\middle|x_m;\omega\right)\right]^{\dagger}_{N\times M}=\left(\left[G\left(x_n\middle|x_m;\omega\right)\right]^{*}_{N\times M}\left[G\left(x_m\middle|x_n;\omega\right)\right]_{M\times N}\right)^{-1}\left[G\left(x_n\middle|x_m;\omega\right)\right]^{*}_{N\times M} \tag{4.8}$$

where the symbol "*" in Equation (4.8) means the conjugate transposition of a matrix.

Note that Equation (4.7) is an explicit solution when the number of sources N is equal to the number of microphones M; and a least-square solution when $M > N$. The n^{th} row on the left side of Equation (4.7) is the Fourier transformation of the time domain signal at the n^{th} source, which can be expressed as:

$$S\left(x_n,\omega\right)=\sum_{n=1}^{N}G^{\dagger}\left(x_n\middle|x_m;\omega\right)P\left(x_m,\omega\right) \tag{4.9}$$

The acoustic signal of the n^{th} source can now be obtained by taking the inverse Fourier transform of Equation (4.9).

$$s\left(x_n,t\right)=\sum_{n=1}^{N}\int_{-\infty}^{\infty}g^{\dagger}\left(x_n\middle|x_m;t'\right)p\left(x_m,t-t'\right)dt'=\sum_{n=1}^{N}g^{\dagger}\left(x_n\middle|x_m;t'\right)*\,p\left(x_m,t-t'\right),\ m=1,2,\cdots,M \tag{4.10}$$

where the temporal free-space Green's function $g^{\dagger}\left(x_n \middle| x_m; t'\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G^{\dagger}\left(x_n \middle| x_m; \omega\right) e^{-i\omega t'} d\omega$.

For example, consider the case where there are two sources and four microphones in a free field, thus $N = 2$ and $M = 4$. The source strengths are $\alpha(\boldsymbol{x}_{\text{source1}})$ and $\beta(\boldsymbol{x}_{\text{source2}})$, respectively. Then the acoustic pressures at the four channels are:

$$p\left(x_{\text{ch}\,1}, t\right) = \frac{\alpha\left(x_{source1}\right)\delta\left(t - \frac{R_{\text{ch}\,1\sim\text{source1}}}{c}\right)}{R_{\text{ch}\,1\sim\text{source1}}} + \frac{\beta\left(x_{source2}\right)\delta\left(t - \frac{R_{\text{ch}\,1\sim\text{source2}}}{c}\right)}{R_{\text{ch}\,1\sim\text{source2}}} \tag{4.11a}$$

$$p\left(x_{\text{ch}\,2}, t\right) = \frac{\alpha\left(x_{source1}\right)\delta\left(t - \frac{R_{\text{ch}\,2\sim\text{source1}}}{c}\right)}{R_{\text{ch}\,2\sim\text{source1}}} + \frac{\beta\left(x_{source2}\right)\delta\left(t - \frac{R_{\text{ch}\,2\sim\text{source2}}}{c}\right)}{R_{\text{ch}\,2\sim\text{source2}}} \tag{4.11b}$$

$$p\left(x_{\text{ch}\,3}, t\right) = \frac{\alpha\left(x_{source1}\right)\delta\left(t - \frac{R_{\text{ch}\,3\sim\text{source1}}}{c}\right)}{R_{\text{ch}\,3\sim\text{source1}}} + \frac{\beta\left(x_{source2}\right)\delta\left(t - \frac{R_{\text{ch}\,3\sim\text{source2}}}{c}\right)}{R_{\text{ch}\,3\sim\text{source2}}} \tag{4.11c}$$

$$p\left(x_{\text{ch}\,4}, t\right) = \frac{\alpha\left(x_{source1}\right)\delta\left(t - \frac{R_{\text{ch}\,4\sim\text{source1}}}{c}\right)}{R_{\text{ch}\,4\sim\text{source1}}} + \frac{\beta\left(x_{source2}\right)\delta\left(t - \frac{R_{\text{ch}\,4\sim\text{source2}}}{c}\right)}{R_{\text{ch}\,4\sim\text{source2}}} \tag{4.11d}$$

Given the time domain signals at the four microphones and the positions of the sources and microphones, we can use PSS to separate individual source signals. The procedures of PSS are as follows.

1) Calculate the Fourier transformed matrix of the signals at four microphones:

$$\{P(x_m,\omega)\}_{4\times1}=\begin{bmatrix}P(x_{\text{ch}\,1},\omega)\\P(x_{\text{ch}\,2},\omega)\\P(x_{\text{ch}\,3},\omega)\\P(x_{\text{ch}\,4},\omega)\end{bmatrix}=\begin{bmatrix}\dfrac{\alpha(x_{source1})e^{ikR_{\text{ch}\,1\sim\text{source1}}}}{R_{\text{ch}\,1\sim\text{source1}}}+\dfrac{\beta(x_{source2})e^{ikR_{\text{ch}\,1\sim\text{source2}}}}{R_{\text{ch}\,1\sim\text{source2}}}\\\dfrac{\alpha(x_{source1})e^{ikR_{\text{ch}\,2\sim\text{source1}}}}{R_{\text{ch}\,2\sim\text{source1}}}+\dfrac{\beta(x_{source2})e^{ikR_{\text{ch}\,2\sim\text{source2}}}}{R_{\text{ch}\,2\sim\text{source2}}}\\\dfrac{\alpha(x_{source1})e^{ikR_{\text{ch}\,3\sim\text{source1}}}}{R_{\text{ch}\,3\sim\text{source1}}}+\dfrac{\beta(x_{source2})e^{ikR_{\text{ch}\,3\sim\text{source2}}}}{R_{\text{ch}\,3\sim\text{source2}}}\\\dfrac{\alpha(x_{source1})e^{ikR_{\text{ch}\,4\sim\text{source1}}}}{R_{\text{ch}\,4\sim\text{source1}}}+\dfrac{\beta(x_{source2})e^{ikR_{\text{ch}\,4\sim\text{source2}}}}{R_{\text{ch}\,4\sim\text{source2}}}\end{bmatrix} \tag{4.12}$$

2) Find the free-space Green's functions needed for PSS:

$$\left[G(x_m|x_n;\omega)\right]_{4\times2}=\begin{bmatrix}G(x_{\text{ch}\,1}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,1}|x_{\text{source2}};\omega)\\G(x_{\text{ch}\,2}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,2}|x_{\text{source2}};\omega)\\G(x_{\text{ch}\,3}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,3}|x_{\text{source2}};\omega)\\G(x_{\text{ch}\,4}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,4}|x_{\text{source2}};\omega)\end{bmatrix}=\begin{bmatrix}\dfrac{e^{ikR_{\text{ch}\,1\sim\text{source1}}}}{R_{\text{ch}\,1\sim\text{source1}}} & \dfrac{e^{ikR_{\text{ch}\,1\sim\text{source2}}}}{R_{\text{ch}\,1\sim\text{source2}}}\\\dfrac{e^{ikR_{\text{ch2}\sim\text{source1}}}}{R_{\text{ch}\,2\sim\text{source1}}} & \dfrac{e^{ikR_{\text{ch}\,2\sim\text{source2}}}}{R_{\text{ch}\,2\sim\text{source2}}}\\\dfrac{e^{ikR_{\text{ch}\,3\sim\text{source1}}}}{R_{\text{ch}\,3\sim\text{source1}}} & \dfrac{e^{ikR_{\text{ch}\,3\sim\text{source2}}}}{R_{\text{ch}\,3\sim\text{source2}}}\\\dfrac{e^{ikR_{\text{ch}\,4\sim\text{source1}}}}{R_{\text{ch}\,4\sim\text{source1}}} & \dfrac{e^{ikR_{\text{ch}\,4\sim\text{source2}}}}{R_{\text{ch4}\sim\text{source2}}}\end{bmatrix} \tag{4.13}$$

3) Since $N = 2$ and $M = 4$, Equation (4.7) can be simplified to:

$$\{S(x_n)\}_{2\times1}=\left[G(x_n|x_m;\omega)\right]^{\dagger}_{2\times4}\{P(x_m,\omega)\}_{4\times1} \tag{4.14}$$

$$\begin{Bmatrix}S(x_{\text{source1}})\\S(x_{\text{source2}})\end{Bmatrix}=\begin{bmatrix}G(x_{\text{ch}\,1}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,2}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,3}|x_{\text{source1}};\omega) & G(x_{\text{ch}\,4}|x_{\text{source1}};\omega)\\G(x_{\text{ch}\,1}|x_{\text{source2}};\omega) & G(x_{\text{ch}\,2}|x_{\text{source2}};\omega) & G(x_{\text{ch}\,3}|x_{\text{source2}};\omega) & G(x_{\text{ch}\,4}|x_{\text{source2}};\omega)\end{bmatrix}^{\dagger}\begin{bmatrix}P(x_{\text{ch}\,1},\omega)\\P(x_{\text{ch}\,2},\omega)\\P(x_{\text{ch}\,3},\omega)\\P(x_{\text{ch}\,4},\omega)\end{bmatrix} \tag{4.15}$$

where the Fourier transformed measurement signals and the Green function matrix are from the first and second steps from Equation (4.12) and (4.13).

The solution of Equation (4.15) is:

$$\begin{Bmatrix}S(x_{\text{source1}})\\S(x_{\text{source2}})\end{Bmatrix}=\begin{Bmatrix}\alpha(x_{\text{source1}})\\\beta(x_{\text{source2}})\end{Bmatrix} \tag{4.16}$$

4) Applying the inverse Fourier transform to each row of Equation (4.16), we can reconstruct the individual source signals as:

$$s(x_{\text{source1}},t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(x_{\text{source1}})e^{-i\omega t}d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}\alpha(x_{\text{source1}})e^{-i\omega t}d\omega=\frac{\alpha(x_{\text{source1}})}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}d\omega=\alpha(x_{\text{source1}})\delta(t) \quad (4.17a)$$

$$s(x_{\text{source2}},t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(x_{\text{source2}})e^{-i\omega t}d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}\beta(x_{\text{source2}})e^{-i\omega t}d\omega=\frac{\beta(x_{\text{source2}})}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}d\omega=\beta(x_{\text{source2}})\delta(t) \quad (4.17b)$$

Theoretically, PSS can reconstruct the exact acoustic source signals, and separation is independent of the types of signals. However, the number of the microphones should be equal to or larger than the number of sources. Note that there is no limitation on the configuration of microphones.

4.2 Theory of ICA

To examine the effectiveness of source separation using the PSS algorithm, the separated signals are compared with those provided by a completely different blind source separation algorithm known as ICA. The reason for choosing ICA as the benchmark algorithm is because its code is available on line [193]. ICA is an existing methodology employing the concept of statistics analyzing and used to solve the BSS problems. It supposes that any observation is a linearly mixture of the independent variables. In the particular case of sound source separation, the measured signals are considered as the mixtures of the sources, which are independent of each other. Moreover, the time delay caused by the wave propagation is assumed small enough to be ignored. Therefore a matrix can be used to describe the relation of the source signals and the measurements.

$$\begin{Bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_m(t) \\ \vdots \\ p_M(t) \end{Bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2n} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & & & \vdots \\ a_{m1} & a_{m2} & & a_{mn} & & a_{mN} \\ \vdots & \vdots & & & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{Mn} & \cdots & a_{MN} \end{bmatrix} \begin{Bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \\ \vdots \\ s_N(t) \end{Bmatrix} \tag{4.18}$$

which can be rewritten as:

$$\{p_m(t)\}_{M\times 1} = [a_{mn}(t)]_{M\times N} \{s_n(t)\}_{N\times 1} \tag{4.19}$$

in which $1 \le m \le M$, $1 \le n \le N$, and $N \le M$.

In Equation (4.19), $\{p_m(t)\}_{M\times 1}$ indicates the measured signals at the M microphones, with m=1, 2, …, M, $\{s_n(t)\}_{N\times 1}$ is the source signals at N sources, and the matrix $[a_{mn}(t)]_{M\times N}$ is the weighting matrix which specifies the contribution parameters of each source to the measurement points. BSS assumes that neither the information of the individual sources nor the weighting matrix is available, thus there are two unknowns in Equation (4.19). Theoretically, a unique solution cannot be obtained when two unknowns exist in one equation, and infinite sets of the solutions of the signal arrays $\{s_n(t)\}_{N\times 1}$ and the weighting matrix $[a_{mn}(t)]_{M\times N}$ exist. However, the ICA algorithm assumes that the target signals are non-Gaussian statistically independent sounds, and utilizes this characteristic to attempt the most reasonable solution of Equation (4.19). In other words, the programming of ICA is actually the procedure of finding a set of solution where the solved target signals have the maximum non-Gaussianity. In this way, ICA can estimate the best solution of both the single arrays $\{s_n(t)\}_{N\times 1}$ and the weighting matrix $[a_{mn}(t)]_{M\times N}$, when the information of the mixed signals $\{p_m(t)\}_{M\times 1}$ is the only input information.

Because of the concept of ICA mentioned above, one of the characteristics of the ICA results is that the separation results may vary every time the program runs on the same input data.

The estimated source signals and their orders are non-predictable and different in every calculation. And if only one time calculation is applied on the mixed signals, the separated results are not guaranteed be the optimal one, therefore the ICA program is desired to run multiple times, and one need to pick up the best solution among all the separation results. Another characteristic of the ICA results is that the amplitudes of the separated signals are not inevitably be the same as the original source signals though their shapes are similar. This is because ICA only tries to attempt the best set of non-Gaussian and statistically independent solutions, which may not necessarily be the actual sound sources. Thus ICA is not reconstructing the original sources, but to estimate a reasonable solution which may represent the sources. In this way the amplitude of the ICA results and the original sources are different. However, the shape of the time domain signals remains similar.

The ICA algorithm used in this paper is the fast ICA program package "FastICA for Matlab 7.x and 6.x Version 2.5" [193].

The only input information for ICA is the measurement at the microphones, while PSS requires the relative positions of the microphones and the sources as well as the measured mixed signals. However, the PSS can work along with the sound source localization program discussed in Chapter 2. Therefore the source position can be obtained by processing the mixed signals and taken as the input for PSS. The details of the combination of sound source localization and PSS will be demonstrated in next chapter.

In this section, numerical simulation is conducted to evaluate the separation effects of PSS and ICA. The mixed signals were firstly estimated at each microphones, and the estimation procedure take the time delay and amplitude decay into account. Next, the mixed signals and the

relative position of the sources regarding to the microphones are considered as known input data for both PSS and ICA. And the separated signals by both methods were compared.

4.3 Comparisons of PSS and ICA

Various real world signals were used as the source signals in the numerical simulation, such as human voices, harp music, chopper sounds, battle field sounds, etc. The mixed signals were estimated at the microphone position, and taken as the input for PSS and ICA, and the separation results were compared to the original sound sources.

To valid the effectiveness of separation, similarity of the separated and the original signals are studied by calculating the correlation coefficient $\gamma_{s,s_0}(x_n)$ between the original sound source $s_0(x_n,t_j)$ and the separated ones $s(x_n,t_j)$ at the n^{th} sound source. The correlation coefficient of two signals can be calculated with the time domain arrays, with

$$\gamma_{s,s_0}(x_n)=\frac{\sum_{j=1}^{J}\left[s(x_n,t_j)-\bar{s}(x_n,t_j)\right]\left[s_0(x_n,t_j)-\bar{s}_0(x_n,t_j)\right]}{(J-1)St_s(x_n)St_{s_0}(x_n)} \tag{4.20}$$

where the subscript *n* specifies the order of the sound source, *j* is the index number of the array in time domain, $1 \le j \le J$. A bar over $s(x_n,t_j)$ or $s_0(x_n,t_j)$indicates the mean value of the array is used. $St_s(x_n)$ and $St_{s_0}(x_n)$ are the standard deviation of $s(x_n,t_j)$ and $s_0(x_n,t_j)$, respectively, and can be expressed as:

$$St_s(x_n)=\sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left[s(x_n,t_j)-\bar{s}(x_n,t_j)\right]^2} \tag{4.21a}$$

$$St_{s_0}(x_n)=\sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left[s_0(x_n,t_j)-\bar{s}_0(x_n,t_j)\right]^2} \tag{4.21b}$$

The value of correlation coefficient between two arrays can be from −100% to +100%, with −100% indicating that the two arrays have the same amplitude but opposite phase, and +100% for two arrays which are exactly the same. In this dissertation, the correlation coefficient is utilized to evaluate the similarities between the separated signal and the original one, therefore the range of the correlation coefficient $\gamma_{s,s_0}(x_n)$ is from zero to +100%. A relatively larger correlation coefficient indicates that the separated signal is more close to the original sound source signal, and the separation is more effective. A correlation coefficient approaching to unity means that the separated signal is the same as the original signal.

As introduced above, the ICA algorithm ignores the time delay at the measurement point and the separated signals are estimated with the amplitudes not necessarily equal to the original signals. Therefore a pre-processing procedure on the separated results of ICA is added before the calculation of the correlation coefficient. The procedure consists of three steps. First, find out the time delay between the separated signal and the original signal by using the cross-correlation method. Second, synchronize these two arrays with zero padding. Third, normalized these two arrays thus their amplitudes are at the same level. After the pre-processing procedure, the contributions in the correlation coefficient caused by the time delay problem and the different level of amplitude of the two signals no long exist. And Equation (4.20) can be applied to the two signal arrays.

On the other hand, the PSS algorithm obtains the unique solution of the sound sources, thus the separated signals by PSS have the same time instance and amplitude as the original signals. Equation (4.20) can be directly applied to the source signal and the separated signal.

There are no restriction on the positions of the sources, and in this dissertation, two sources were set at (0, 0, 1) and (2, 0, 0) in meter in the Cartesian coordinate, respectively. Four

microphones were located at (0.50, 0, 0), (0, 0.50, 0), (0, 0, –0.50) and (0, 0, 0.50) in meter, and this set of microphones are considered as orthogonal configuration. Another non-orthogonal microphone configuration was also tested in the simulation, and the impact of the microphone positions on the separation will be discussed later in this chapter.

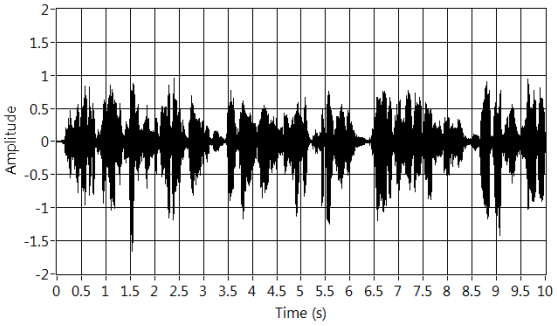
The measurements were estimated at the microphone positions, and ICA and PSS were used to separate sound signals, respectively. Three cases were tested in the simulation, and in each case two real world signals with same amplitude were employed as the sound sources:

1) Case 1: woman's and man's voice,
2) Case 2: woman's voice and chopper sound, and
3) Case 3: woman's voice and harp music.

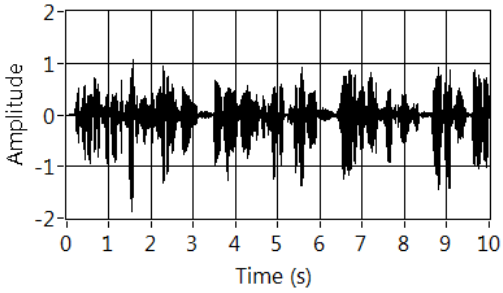
4.3.1 Case 1: Separation of mixed woman's and man's voices

Figure 4.1 shows a sample of the mixed signals in Case 1 and the separation results by PSS and ICA. Figure 4.1 (a) demonstrates the measurement at the Channel #1 among the four microphones, which is a mixed signal containing the woman's and man's voice. Figure 4.1 (b) and (c) show the separated woman's and man's voice, respectively. Figure 4.1 (d) and (c) show the results obtained by ICA. Note that ICA cannot achieve a specific order of the separated signals. In various running of the ICA program, some of the solutions have the woman's voice come out first, while the others take the man's voice as the first signal. The order of the separated signals by ICA is reordered manually thus it matches that of the original signal, therefore the comparison of the results can be easily conducted. From the information shown in Figure 4.1, both of the methods can successfully separate the two sound signals. Moreover, the correlation coefficients of these two set of results were calculated. The separated woman's and man's voice by PSS have the correlation coefficients equal to 99% and 98.29% corresponding to their

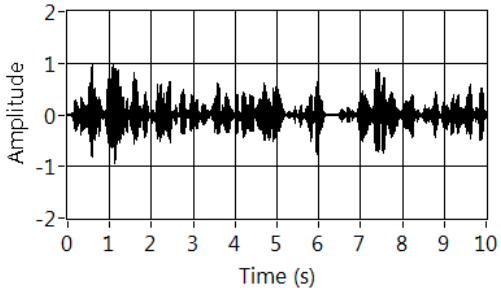
benchmark signals, respectively; and those for the results of ICA are 91.8% and 82.24%, respectively.



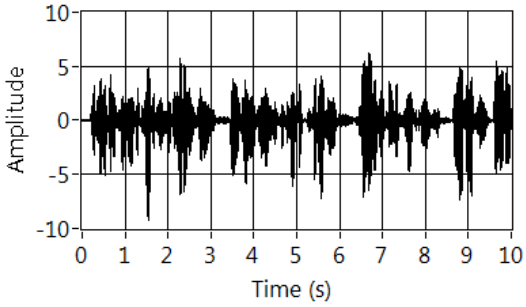
(a) Mixed signals (woman's and man's voices)



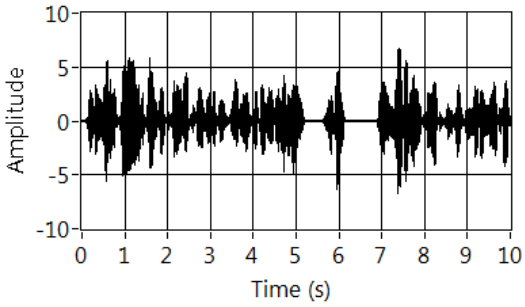
(b) Separated woman's voice using PSS



(c) Separated man's voice using PSS



(d) Separated woman's voice using ICA

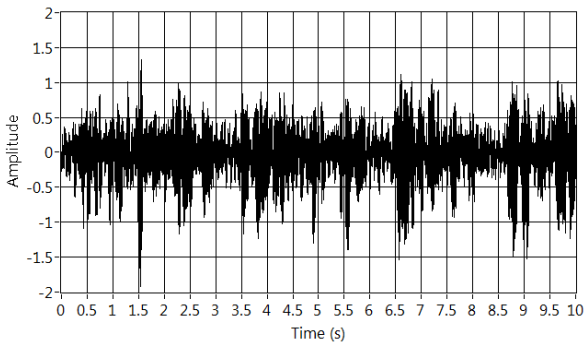


(e) Separated man's voice using ICA

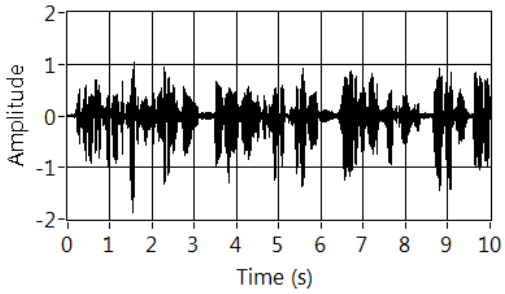
Figure 4.1 Separation of woman's and man's voices with an orthogonal array of microphones. (a) Mixed signals of woman's and man's voices measured at Mic. #1. (b) Separated woman's voice using PSS; (c) Separated man's voice using PSS; (d) Separated woman's voice using ICA; and (e) Separated man's voice using by ICA.

4.3.2 Case 2: Separation of mixed woman's voice and chopper sound

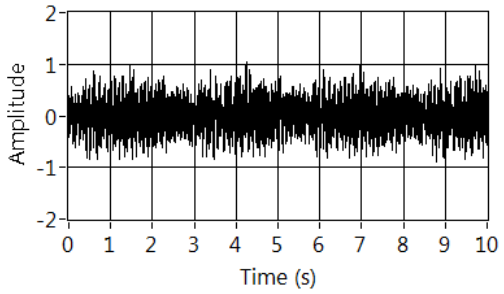
Figure 4.2 illustrates the separation in Case 2, where the woman's voice and chopper sound were used as the sound source signals. In this case, both methods can also separate the two signals. The correlation coefficients for the separated woman's voice and chopper sound by PSS are 98.97% and 99.25%, respectively; and those obtained by ICA are 93.2% and 81.58%, respectively.



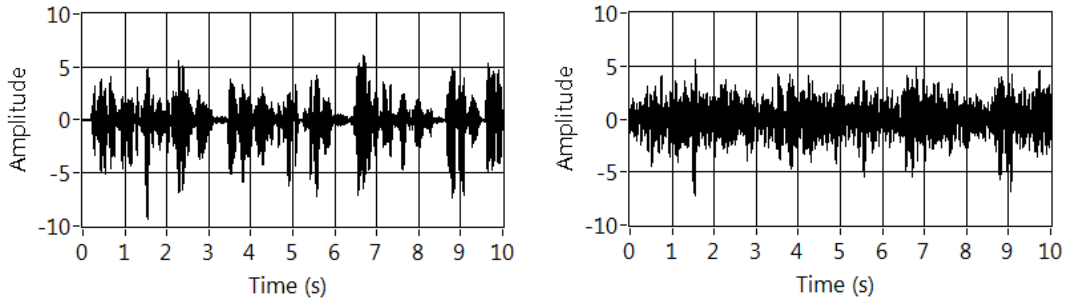
(a) Mixed signals (woman's voice and chopper sound)



(b) Separated woman's voice using PSS



(c) Separated chopper sound using PSS

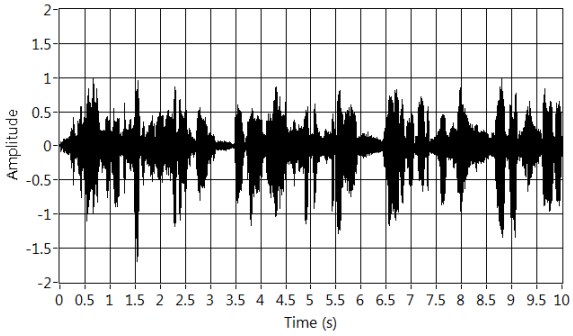


(d) Separated woman's voice using ICA (e) Separated chopper sound using ICA

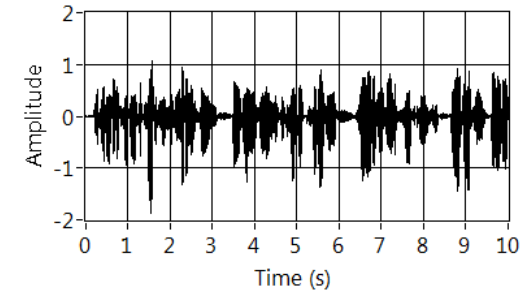
Figure 4.2 Separation of woman voice and chopper sound with an orthogonal array of microphones. (a) Mixed signal of woman's voice and chopper sound measured at Mic. #1. (b) Separated woman's voice by using PSS; (c) Separated chopper sound by using PSS; (d) Separated woman's voice by using ICA; and (e) Separated chopper sound by using ICA.

4.3.3 Case 3: Separation of mixed woman's voice and harp music

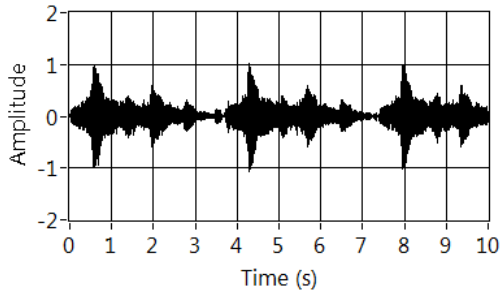
Figure 4.3 shows the results in Case 3. Similarly, the correlation coefficients of the separated woman's voice and harp music by using PSS are 99.01% and 98.74%, respectively; and those by ICA are 92.73% and 69.83%, respectively.



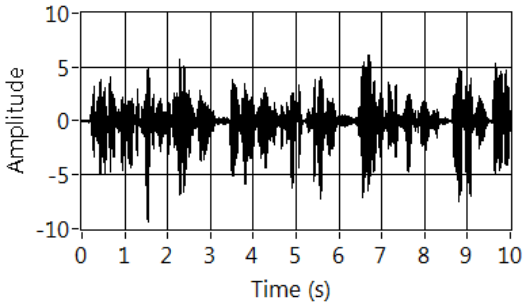
(a) Mixed signals (woman's voice and harp music sound)



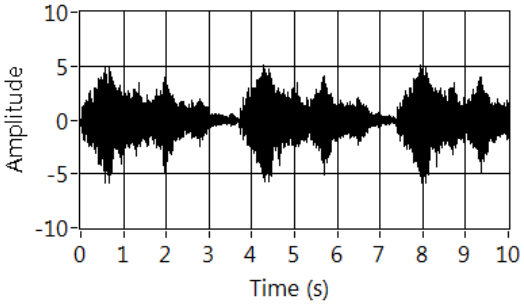
(b) Separated woman's voice using PSS



(c) Separated harp music sound using PSS



(d) Separated woman's voice using ICA



(e) Separated harp music sound using ICA

Figure 4.3 Separation of woman's voice and harp music sound with an orthogonal array of micro-phones. (a) Mixed signals of woman's voice and harp sound measured at Mic. #1. (b) Separated woman's voice by using PSS; (c) Separated harp sound by using PSS; (d) Separated woman's voice by using ICA; and (e) Separated harp sound by using ICA.

Comparing the separated results by PSS and ICA, the ICA results have distortions in amplitudes as well as phases. This is more obvious in the separated results of chopper sound and the harp music, and the correlation coefficients of these two results are also lower than the others. In contrast, the results by PSS have no distortions on amplitude or phase, and have higher correlation coefficients than those by ICA, which means PSS can better separate the mixed signals.

The results are understandable based on underlying principles of PSS and ICA. PSS employs the mixed signals as well as a known transfer matrix, which is composed by free-space Green's function working on each source and microphones. Equation (4.7) only has one unknown, thus the source signals can be uniquely determined. The numerical model of PSS

considered the time delay and amplitude decay happens during the wave propagation, and the separated signals are reconstructed and theoretically should be the same as the original source signals. On the other hand, ICA has two unknowns in its numerical model, as Equation (4.19) shows, and it is impossible to get a unique solution of both the source signals and the weighting matrix, but rather estimate an optimal solution which has the ratio or relative amplitudes of the source signals. The solution of ICA has not the same amplitude or phase as the original signal, though its amplitudes may indicate the general idea of the source signals.

The advantage of PSS is that it can reconstruct the original sound signals in both the amplitude and the phase at each individual source location as long as their positions are known in advance. However, the locations of the sources may not be available or be determined correctly in practice. In these cases ICA can be used to separate the mixed signals and give an estimation of the source signals, though there is no information about how these sources are distributed in space, the order of the sources may vary every time the ICA program runs, and even sometimes the separation may not be successful.

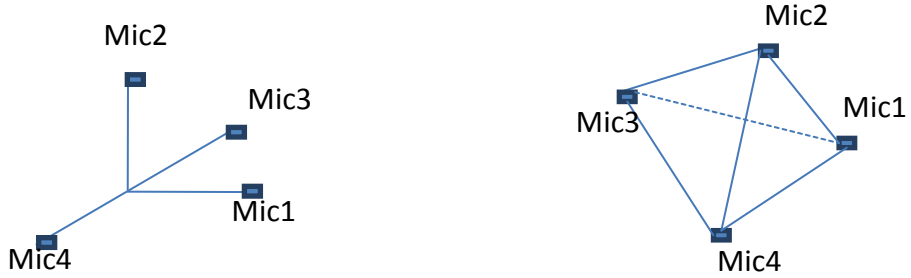
Besides the separation effect, the processing time of these two methods was also tested in simulation. PSS operates faster than ICA in general. Take the processing time in Case 1 for an example, PSS used 3.29 second while ICA requires 5.81 second to get the results. The time cost in the other cases shows the similar tendency as Case 1.

4.4 Impact of various parameters on sources separation using PSS and ICA

The impacts of various parameters, including the microphone configuration, the number of microphones, sound source type, and the SNR, are discussed in this section for a better understanding of PSS and ICA.

4.4.1 Impact of microphone configuration

In last section, an orthogonal microphone set (See Figure 4.4 (a)) was used in space collecting the mixed signals to validate the separation effectiveness of PSS and ICA. In this section, another microphone configuration, which put the microphones at four arbitrarily selected positions in space, is employed and called non-orthogonal set. The Cartesian coordinates of the non-orthogonal microphone set are (0.3, 0.73, 0.02), (0.09, 0.16, –0.33), (–0.33, 0.14, 0.14), (0.34, 0.34, 0) in meter, as shown in Figure 4.4 (b). The separation results by two methods on two microphone sets were compared and the impact of the positions of microphones is discussed.



(a) Orthogonal array of microphones (b) Non-orthogonal array of microphones

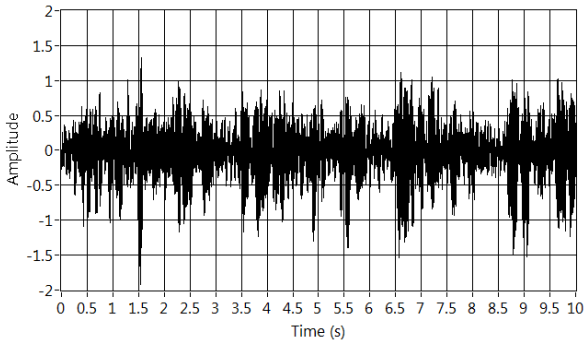
Figure 4.4 Microphone configurations. (a) Orthogonal configuration with microphones mounted at (0.50, 0, 0), (0, 0.50, 0), (0, 0, –0.50) and (0, 0, 0.50) in meter. (b) Non-orthogonal configuration with microphones placed at (0.3, 0.73, 0.02), (0.09, 0.16, –0.33), (–0.33, 0.14, 0.14), (0.34, 0.34, 0) in meter.

Similar to the last section, three cases of mixed signals were chosen in the test of the non-orthogonal configuration. The mixed signals were estimated at the positions of the microphones, and then be used as the input for the PSS and ICA algorithm.

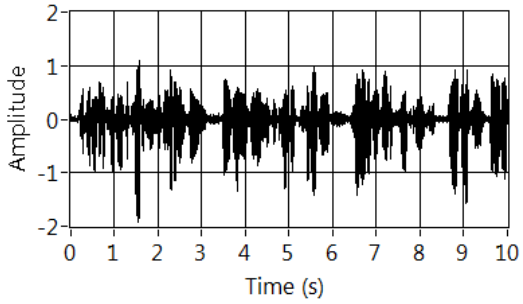
Figure 4.5 illustrates the separation results in Case 2 of both methods with a non-orthogonal microphone configuration. Figure 4.5 (a) shows mixed signal at the Channel #1. Figure 4.5 (b) and (c) show the separated woman's voice and chopper sound obtained by PSS, which have correlation coefficients equal to 94.94% and 96.15%, respectively. The separation by

PSS with non-orthogonal set is successful though the correlation coefficients are a little less than those obtained by orthogonal set. Figure 4.5 (d) and (e) demonstrate the separation results with by ICA. It is obvious that the separation is not complete, especially for the separated woman's voice shown in Figure 4.5 (d), the background chopper can be also seen along the time domain signal. The correlation coefficients for the results by ICA are 76.8% and 76.9%, respectively, which are much less than those by the orthogonal set. The separation by ICA in this case is successful yet not clear.

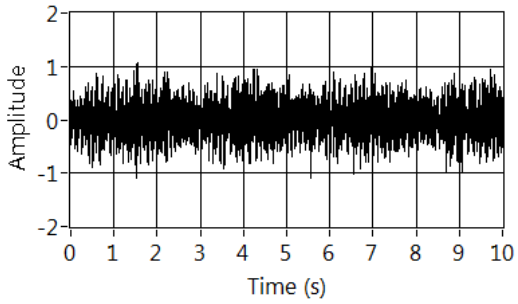
Table 4.1 gives a summarization of the values of correlation coefficients in the three signal mixture case with orthogonal and non-orthogonal microphone sets. The separation effectiveness of PSS is better than ICA in general. The PSS can separate all the mixed signals listed in the table and keep the correlation coefficients above 90%. However, the results by PSS also indicate that the separation on the mixed signals at non-orthogonal set is less effective than those on orthogonal set. This tendency is much more significant when ICA is utilized for sound separation. ICA can successfully separate signals when orthogonal microphone set is used. However, when non-orthogonal set is used, for the case that woman's and chopper sound are mixed, the separation by ICA is not clear or complete (See Figure 4.5). For the cases of the woman's and man's voice mixture and the woman's voice and harp music mixture, the correlation coefficients of the results by ICA are even below 50%, which mean the separations are not successful.



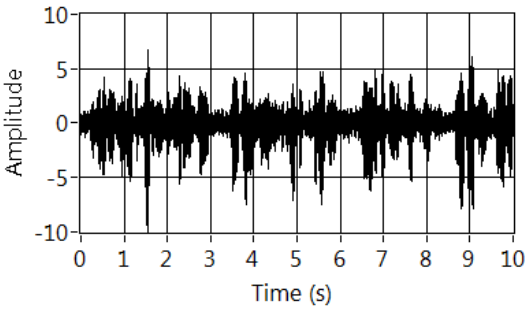
(a) Mixed signals (woman's voice and chopper sound)



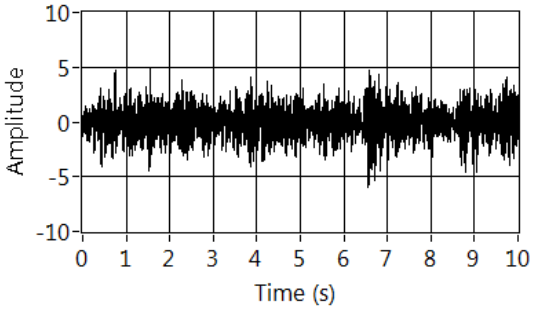
(b) Separated woman's voice using PSS



(c) Separated chopper sound using PSS



(d) Separated woman's voice using ICA



(e) Separated chopper sound using ICA

Figure 4.5 Separation of woman's voice and chopper sound with a non-orthogonal array of micro-phones. (a) Mixed signals of woman's voice and chopper sound measured at Mic. #1. (b) Separated woman's voice by using PSS; (c) Separated chopper sound by using PSS; (d) Separated woman's voice by using ICA; and (e) Separated chopper sound by using ICA.

Table 4.1. Comparisons of the correlation coefficients under different microphone configurations.

		$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **PSS**		$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **ICA**	
		1st Source	2nd Source	1st Source	2nd Source
Orthogonal Microphone Array	Woman's voice vs. Man's voice	99.00%	98.29%	91.80%	82.24%
	Woman's voice vs. Chopper sound	98.97%	99.25%	93.20%	81.58%
	Woman's voice vs. Music sound	99.01%	98.74%	92.73%	69.83%
Non-orthogonal Microphone Array	Woman's voice vs. Man's voice	95.17%	90.09%	48.12%	24.37%
	Woman's voice vs. Chopper sound	94.94%	96.15%	76.80%	76.90%
	Woman's voice vs. Music sound	95.19%	92.54%	44.28%	42.64%

4.4.2 Impact of number of microphones

The impact of microphone number is also interested in the sound separation problems. Because of the underlying theory of PSS and ICA, both methods require that the number of microphone be equal or larger than the number of sources. In the study in this dissertation, two sources are mixed in each test case, thus at least two microphones are needed to separate the sounds. In this section, the impact of the microphone number on the separation effectiveness is tested by using the measurement of mixed woman's voice and chopper sound at two to four microphones in the orthogonal set. When the number of microphone is defined as two, only the mixed signals at Channel #1 and #2 are used; if one decides to use three microphones, measurements at Channel #1, #2, and #3 are selected; and all the four channels are used when the microphone number is set as four.

Table 4.2 illustrates the correlation coefficients of the separated signals using different number of microphones by two methods. The results show that the separation effectiveness is independent of the number of microphones, and there is no advantageous of using more microphones than only employing the minimum number of microphones. Hence there is no need

to increase the number of microphones, and the microphone number equal to the source number is enough for sound separation by PSS and ICA.

Table 4.2. Correlation coefficients of sources separation using different numbers of microphones.

	$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **PSS**		$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **ICA**	
	1st Source	2nd Source	1st Source	2nd Source
Two Microphones	99.04%	96.98%	96.1%	82.06%
Three Microphones	99.65%	98.40%	95.79%	84.92%
Four Microphones	99.00%	98.29%	91.80%	82.24%

4.4.3 Impact of source type

The type of source is also an important factor in sound separation, and it is desired that a separation method be applicable to all types of sources, because the type of sound sources are unknown in advance in most separation cases. In this dissertation, four different sound signals, the woman's and man's voice, chopper sound, and harp music are used in the three mixture cases tested.

The correlation coefficients in Table 4.1 have shown that PSS can separate any kind of sources, and the effectiveness of separation is independent of the source type. However, the effectiveness of separation is different on the source type when ICA is used. This is because that the ICA algorithm consists of whitening and dimensionality reduction, which is based on the assumption that the sound sources are statistical independent and non-Gaussian. In other words, ICA employs the non-Gaussian and statistical independent characteristics of the sources to solve the sound separation problem. Therefore, it is designed to separate the signals which are non-Gaussian, and when the source has a high Gaussianity, the assumption of the ICA algorithm is not fully satisfied, and the separation of the sound sources can no longer be guaranteed. In the

case when the signals are mixture of non-Gaussian sources, such as the woman's and man's voice, ICA can separate the sound source successfully; while when some more Gaussian sources involved in the mixtures, such as the chopper sound, the effectiveness of separation by ICA decreases.

PSS employs the idea that sound are radiated from point sources, and transfer function between the sources and the measurements are defined by Green's function in free field. It doesn't have any restrictions on the type of sources in its algorithm, and the correlation coefficients shown in Table 1 also demonstrate that its separation is available for all types of sources.

4.4.4 Impact of SNR

In all the results given above, the two sources have the same amplitude, thus their SNR are zero. However, this may not always be true in reality. It is highly possible that the level of one source is higher than the other in many cases, and the separation of the sources is still needed. To validate the effectiveness of separation by both methods, the mixture of woman's and man's voice are used in the test, the woman's voice is consider as the target source, while the man's voice is an interfering background noise. The sound level of the man's voice was increased step by step until it can totally mask the woman's voice. At the beginning, the man's voice has the same level as the woman's voice, and SNR is zero dB; next, SNR is decreased to −5 dB, −10 dB, and −15 dB, which means the man's voice is three times, ten times, and 32 times greater than the woman's voice, respectively. And when the SNR = −15 dB, the woman's voice is totally covered by the man's voice, and can be hardly heard at the measurement points.

The mixed signals were taken as the input to PSS and ICA, respectively, and the correlation coefficient of the results and the original signals are shown in Table 4.3. The results by ICA are independent of SNR, as long as the type of sound sources satisfies the non-Gaussian and statistically independent characteristics. On the other hand, separation results of PSS are influence by SNR. The correlation coefficient of target source decreases from 99.00% to 82.9% while SNR decreases from zero dB to −15 dB. However, as discussed above, the separation effectiveness of PSS is better than ICA in general. Thus when the SNR is above −10 dB, the correlation coefficients by PSS are still higher than those by ICA though they are influenced by SNR.

Table 4.3. Correlation coefficients of sources separation under different SNR.

	$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **PSS**		$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **ICA**	
	Woman's voice	Man's voice	Woman's voice	Man's voice
SNR = 0 dB	99.00%	98.29%	91.8%	82.24%
SNR = –5 dB	98.81%	99.56%	93.18%	93.73%
SNR = –10 dB	97.17%	99.67%	89.08%	86.52%
SNR = –15 dB	82.90%	99.68%	93.45%	86.70%

CHAPTER 5

BLIND SOUND SOURCES LOCALIZATION AND SEPARATION

After the validation of the sound source localization in Chapter 2 and PSS in Chapter 3, it is highly desirable to combine these two algorithms together. Therefore, once the directly measured signals at each microphone are obtained, the combined program can find the source location as well as reconstruct the source signals. In this chapter, the blind sources localization and separation (BSLS) is introduced and validated numerically, and impact of the accuracy of localization on the effectiveness of separation is discussed.

5.1 Combined source localization and source separation

The sound source localization and PSS algorithm share the same assumption that the sound is radiated from point sources in free field, and its travelling obeys the Green's function in free field. Localization algorithm requires at least four microphones in order to find the sources positions in three-dimensional space; and PSS requires the number of microphone equal to or larger than that of the sources. Moreover, as the results shown in Chapter 4, the orthogonal microphone configuration has better separation effectiveness than the non-orthogonal one, and the orthogonal set is also validated in Chapter 2 and 3 that it can successfully locate multiple sound sources. Therefore the orthogonal microphone set, which consists of four microphones located at (0.5, 0, 0), (0, 0.5, 0), (0, 0, − 0.5), and (0, 0, 0.5) in meter in Cartesian coordinates, are utilized in the validation of BSLS.

Sound source localization requires the input of the microphone positions and the measurement at each microphone; while PSS needs the microphone measurements and the relative positions of microphone and source. In other words, the inputs required by PSS are from

both the input and output of the sound source localization. Therefore, BSLS consists of two steps: First, use the localization algorithms to determine the source positions; second, take the calculated source positions and the measured signals as the input of PSS, the separated sound signals pairing with their locations can be obtained.

Figure 5.1 demonstrates the detailed flowchart of BSLS. The positions of the microphones are fixed and their coordinates are stored in computer program in advance. Once the incident sound sources are measured by the microphones, the program automatically acquires the spectrogram of the measurement at Channel #1. One of the significantly high peaks on the spectrogram is found and its corresponding time domain signals at each microphone are selected to be the input of the sound source localization program. Note that the position of a peak in a spectrogram indicates both the time instance and frequency band information of the signal; therefore the selection of the corresponding time domain signals requires the windowing and filtering pre-processing introduced in Chapter 2. With the selected time domain signals at each microphone, the source localization algorithm can obtain the location of the dominant signal at this specific time and frequency band. As there can be multiple peaks in the spectrogram, the procedure of selecting time domain signals and finding source location should be repeated until all the peaks in spectrogram are evaluated or a satisfactory result of localization is obtain. Whenever a source location is found, its coordinates should be saved in the database for further use in the PSS program. However, as the localization procedure is conducted in iteration, and the peaks in the spectrogram may representing the same dominant source signal, it is possible that the newly calculated one coincides with a location already been saved in the database. At this point, if the location is already been stored, it doesn’t need to be saved again, thus the latest source localization result should be discarded; otherwise, it should be saved to the database of

the source locations. This procedure of searching sources is looped until all the sound sources are located and their positions are saved in the database. Next, all the source locations, microphone positions, and the directly measured incident sound signals are used as the input of the PSS program. PSS separate and reconstruct the original sound sources, and output them pairing with their locations.

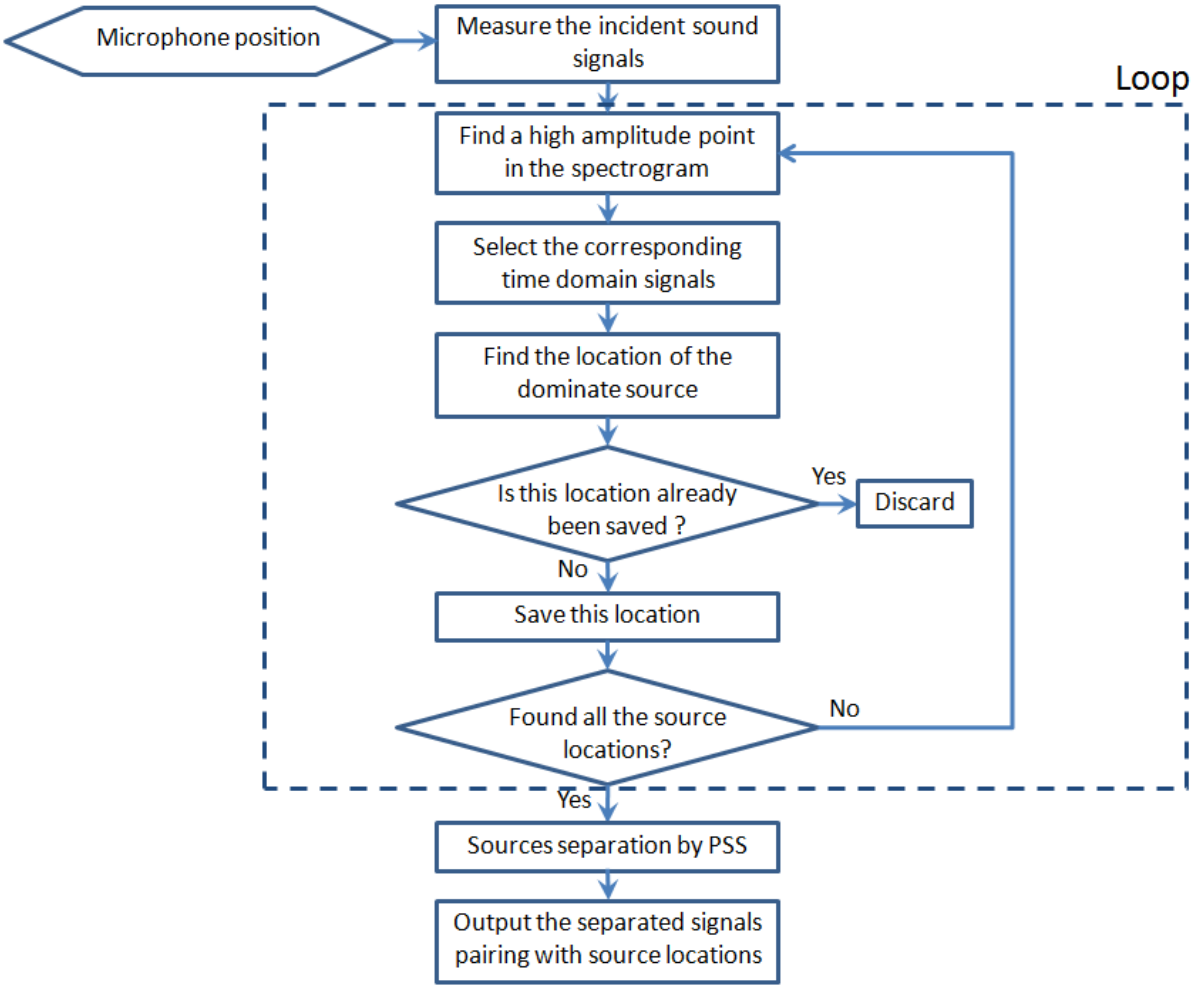
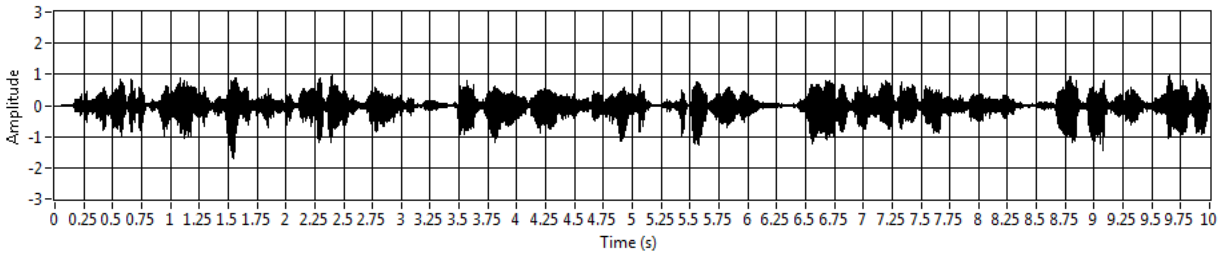


Figure 5.1 Flow chart of BSLS

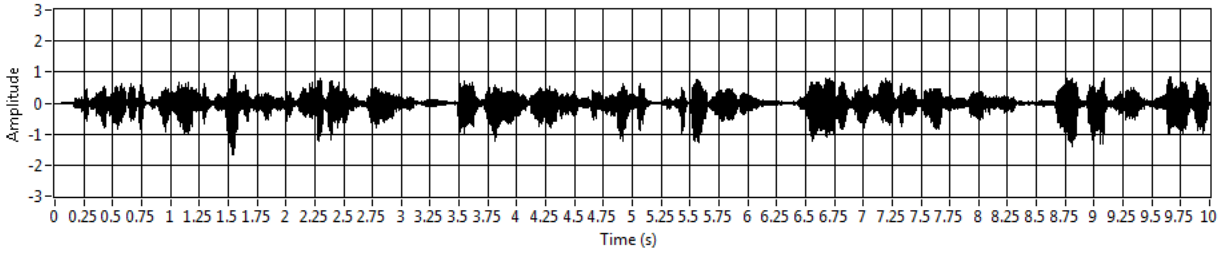
Take one case for an example, assume that woman's and man's voices exist in space, and they are located at (0, 0, 1) and (2, 0, 0) in meter in Cartesian coordinates, respectively. The microphone positions, as mentioned above, are at (0.5, 0, 0), (0, 0.5, 0), (0, 0, − 0.5), and (0, 0,

0.5) in meter. The measurements at four microphones can be generated, and they are applied to the programming following the flowchart in Figure 5.1 to validate the blind sound sources localization and separation.

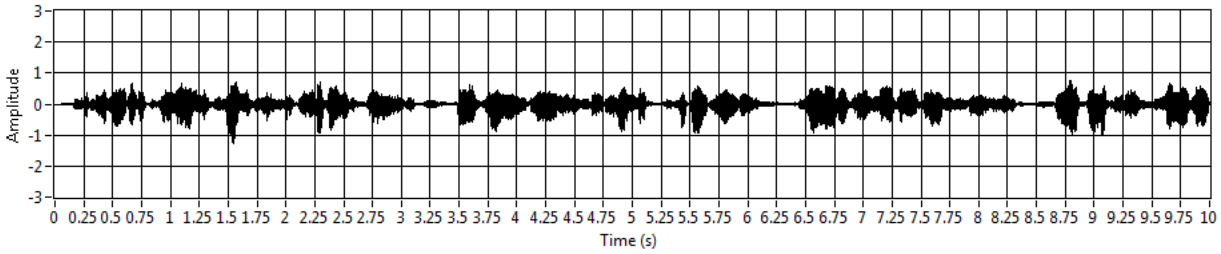
In this example, the only inputs for the program are the positions of microphones and the mixed time domain signals at the four channels, as shown in Figure 5.2.



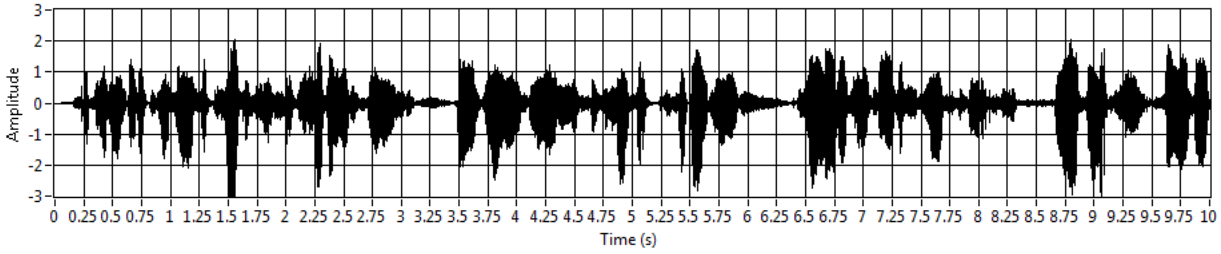
(a) Sound signal measured at Channel #1



(b) Sound signal measured at Channel #2



(c) Sound signal measured at Channel #3



(d) Sound signal measured at Channel #4

Figure 5.2 Incident sound signals measured at four microphones

BSLS consists of following steps:

1) Generate the spectrogram of the signal measured at Channel #1, as shown in Figure 5.3.

2) Detect one of the amplitude peaks in the spectrogram, and find the time instance and frequency band of it. In this example, as the black crosshair shown in Figure 5.3, the peak is located at between 7 to 7.5 second in time domain, and about 400 Hz in frequency domain. Therefore a 0.5 second width window is applied to the signals at each channel, from 7 to 7.5 second, followed by a filter, which has the low and high pass frequency equal to 300 Hz and 500 Hz, respectively. The selected signals at four microphones are shown in Figure 5.4.

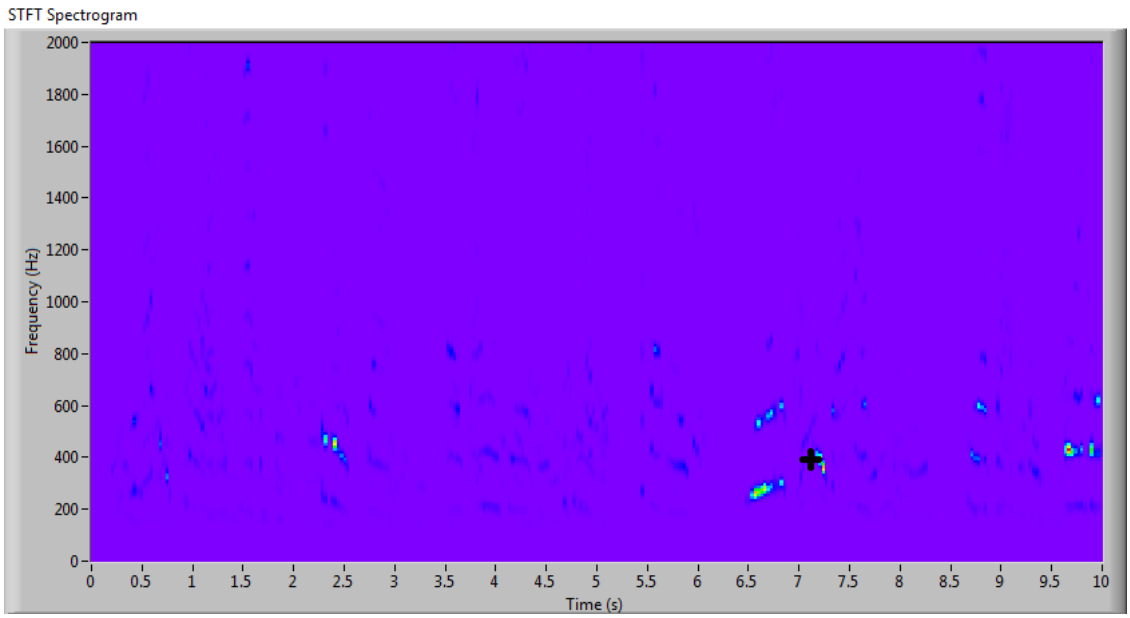
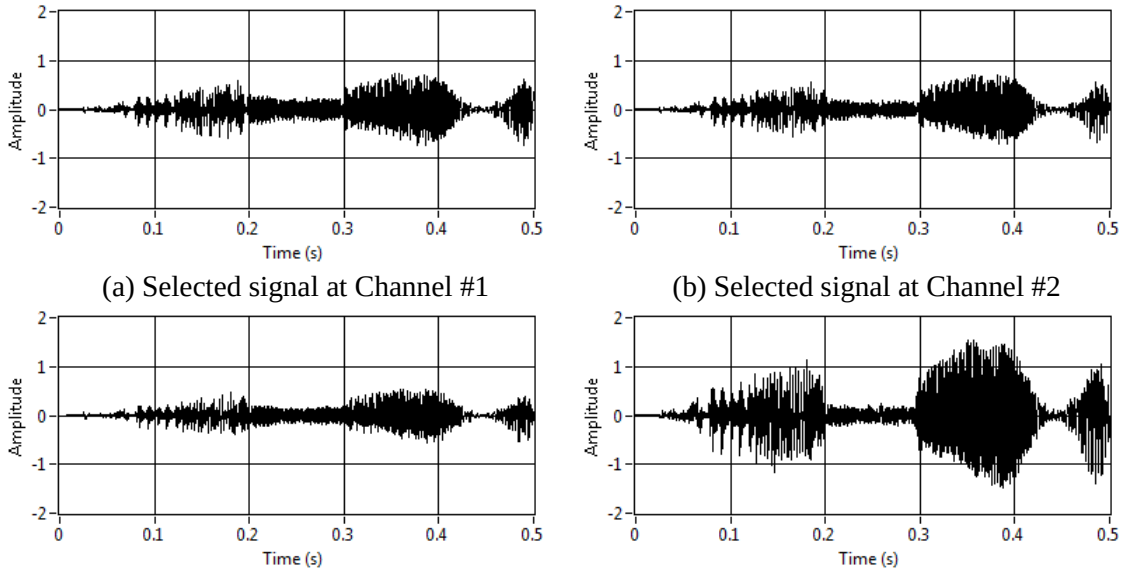


Figure 5.3 Spectrogram of measured signal at Channel #1. The signal length is 10 second, and the frequency domain shows the information up to 2000 Hz. The black crosshair indicates one of the peak amplitude position inside this spectrogram, which happens between 7 to 7.5 second, and at frequency of 300 to 500 Hz.



(a) Selected signal at Channel #1

(b) Selected signal at Channel #2

(c) Selected signal at Channel #3

(d) Selected signal at Channel #4

Figure 5.4 Corresponding time domain signals at four microphones. The data displayed in this figure are those after windowing and filtering pre-process.

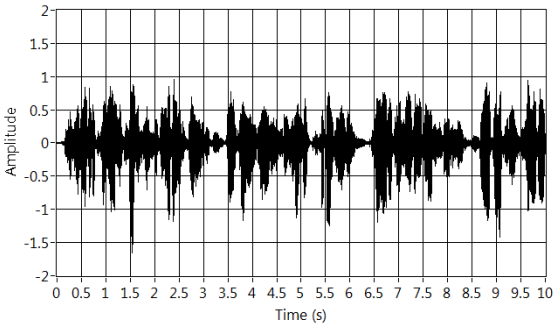
3) Use the selected signals as the input of the basic sound source localization algorithm. As the signal pre-processing procedure is already done in the last step, all one need to do is to estimate the time delays between microphones, submit them to Equation (2.8) and get the solution. In this example, the calculated source location corresponding to this particular point in the spectrogram is (0.0815947, 0.0815947, 0.840506) in meter.

4) Save the position into the database. If the new source location is the same as anyone in the database, discard it; otherwise save it in the database.

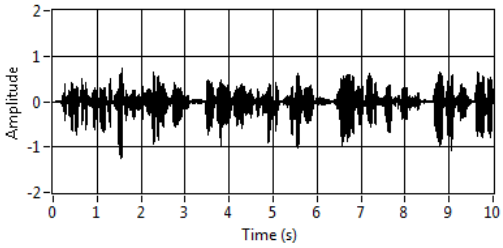
5) Go back to the spectrogram and find another amplitude peak, and then go through Step two to four. This step should be conducted multiple times until all the sources are found, in other words, no more new source location comes out in Step three. In this example, a second source location obtained is obtained at (1.56356, 0, 0).

6) Take the directly measured signals of the sound sources and the sound source locations into the PSS program and reconstruct the original sound sources. In this

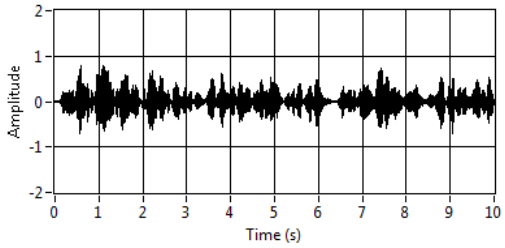
example, the two sources are at (0.0815947, 0.0815947, 0.840506) and (1.56356, 0, 0). The reconstructed sound signals represent the woman's and man's voice, respectively, as shown in Figure (b) and (c).



(a) Mixed signals at Channel #1 (woman's and man's voices)



(b) Separated woman's voice using PSS



(c) Separated man's voice using PSS

Figure 5.5 . Separation of woman's and man's voices with an orthogonal array of microphones. (a) Mixed signals of woman's and man's voices measured at Mic. #1. (b) Separated woman's voice using PSS; (c) Separated man's voice using PSS; (d) Separated woman's voice using ICA; and (e) Separated man's voice using by ICA.

5.2 Numerical simulation results

Numerical simulation of BSLS was conducted in this section. Two sound sources were set at (0, 0, 1) and (2, 0, 0) in meter, respectively, and the orthogonal microphones configuration were chosen as (0.50, 0, 0), (0, 0.50, 0), (0, 0, – 0.50) and (0, 0, 0.50) in meter. Real world sounds were utilized as the sound source signals, and three cases were tested, which are:

1) Case 1: woman's and man's voice,

2) Case 2: woman's voice and chopper sound, and

3) Case 3: woman's voice and harp music.

The estimation of the mixed signals at microphones position is the same as those in Chapter 4. However, Chapter 4 only aims at the validation of PSS, and the source locations were considered as known input in Chapter 4. In this section, the only inputs for the program are the mixed signals and the positions of microphones. The source locations are firstly determined by the sound source localization program and then taken as the input for PSS.

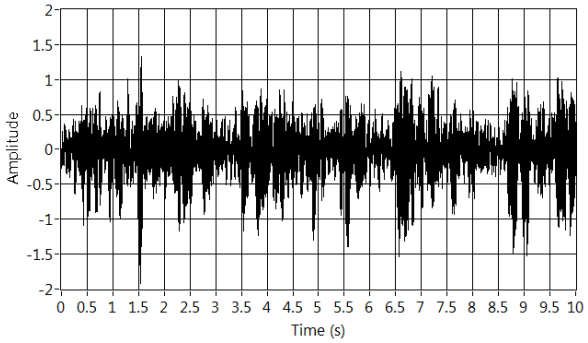
5.2.1 Case 1: Separation of mixed woman's and man's voices

Figure 5.5 shows the BSLS results for the case that woman's and man's voice are mixed in space. Figure 5.5 (a) illustrates the mixed signal measured at Channel #1. After the procedure of finding the sound sources locations introduced above in Section 5.1, the two sound sources are located at (0.0815947, 0.0815947, 0.840506) and (1.56356, 0, 0) in meter, respectively. As mentioned above, the benchmark locations of the two sources are (0, 0, 1) and (2, 0, 0) in meter. The errors of sound localization are therefore 3.88% and 4.76%, respectively, which are calculated based on Equation (2.15) in Chapter 2. Take the calculated sound sources and the directly measurements at the four microphones as the input for PSS, the sound sources are separated as shown in Figure 5.5 (c). The separated results are clear and completed, with the correlation coefficient with the original signals equal to 96.27% and 87.12%, respectively.

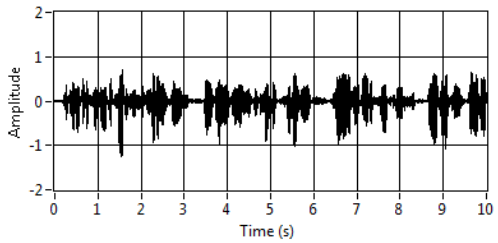
5.2.2 Case 2: Separation of mixed woman's voice and chopper sound

In the second case, the woman's voice and chopper sound were mixed. The results are shown in Figure 5.6. Figure 5.6 (a) indicates the mixed signal measured at Channel #1. The program follows the flowchart in Figure 5.1 and two locations were detected at (0.0815947, 0.0815947, 0.840506) and (1.98627, – 0.0697182, – 0.0697182) in meter, respectively. The

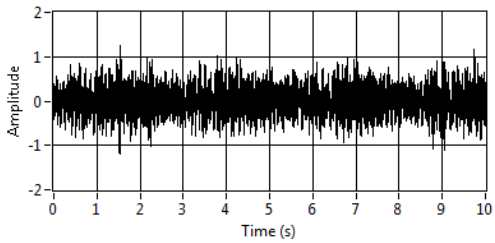
errors for the localization results in this case are 3.88% and 0.25%, respectively. Next, PSS was conducted and the separated signals are shown in Figure 5.6 (b) and (c), and they are paired with the sound source locations, respectively. The correlation coefficient on the woman's voice result is 96.26%, and that for chopper sound is 94.59%, which illustrate that BSLS in this case is successful.



(a) Mixed signals at Channel #1 (woman's voice and chopper sound)



(b) Separated woman's voice using PSS



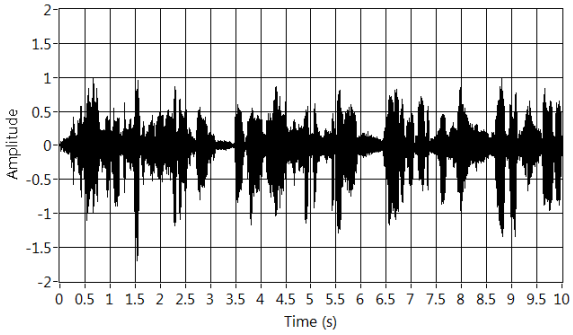
(c) Separated chopper sound using PSS

Figure 5.6. Separation of woman voice and chopper sound with an orthogonal array of microphones. (a) Mixed signal of woman's voice and chopper sound measured at Mic. #1. (b) Separated woman's voice by using PSS; (c) Separated chopper sound by using PSS; (d) Separated woman's voice by using ICA; and (e) Separated chopper sound by using ICA.

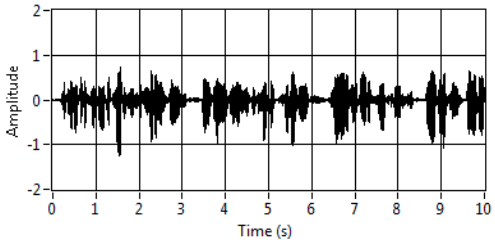
5.2.3 Case 3: Separation of mixed woman's voice and harp music

Figure 5.7 demonstrate the third case where woman's voice was mixed with harp music. Similar to the last two cases, the sound sources were located at (0.0815947, 0.0815947, 0.840506) and (1.56356, 0, 0) in meter, and their errors are 3.88% and 4.76%, respectively. The separated sound signals are shown in Figure 5.7 (b) and (c), and their correlation coefficients are 96.26%

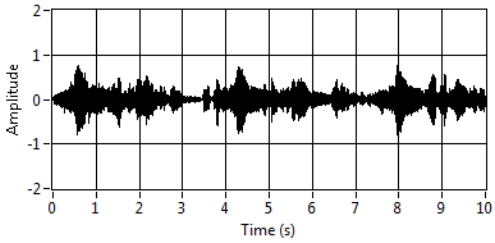
and 89.95%, respectively. BSLS can also successfully reconstruct the original signals pairing with their locations.



(a) Mixed signals at Channel #1 (woman's voice and harp music sound)



(b) Separated woman's voice using PSS



(c) Separated harp music sound using PSS

Figure 5.7. Separation of woman's voice and harp music sound with an orthogonal array of micro-phones. (a) Mixed signals of woman's voice and harp sound measured at Mic. #1. (b) Separated woman's voice by using PSS; (c) Separated harp sound by using PSS; (d) Separated woman's voice by using ICA; and (e) Separated harp sound by using ICA.

Summarizing the correlation coefficients of the results in the upper three cases, and comparing them with those obtained in Chapter 4, as Table 5.1 shows, it can be seen that BSLS is not as effective as the PSS. This is reasonable because PSS uses the input of the benchmark locations of the sound sources, but the blind source localization and separation program attempts to find out the locations of sound sources, and the locations calculated has some errors which may has a negative impact on the separation effect. However, BSLS has the advantages that it requires less input information than PSS: the locations of the sound sources are not required in its

program. This makes BSLS applicable, because the source locations are not available in many practical cases.

What is more, if the results by BSLS are compared to those by ICA, it is obvious that it has better separation results than ICA in general. Moreover, the results by BSLS have not only the separated signals, but also their locations, which cannot be gained by ICA. The only difference of the input for these two methods is that BSLS needs the positions of microphones in advance, which is very easy to get; however, BSLS achieves more information of the sound sources, including their time domain signals and positions, and its separation effectiveness is also better than ICA.

The processing time of BSLS is difficult to be predicted though, because the time costs in the loop in Figure 5.1 varies in every case, and it is the most time consuming part in the flowchart.

Table 5.1. Comparisons of the correlation coefficients under different algorithms

	$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **PSS**		$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **BSLS**		$\gamma_{s,s_0}(\boldsymbol{x}_n)$ for **ICA**	
	1st Source	2nd Source	1st Source	2nd Source	1st Source	2nd Source
Woman's voice vs. Man's voice	99.00%	98.29%	96.27%	87.12%	91.80%	82.24%
Woman's voice vs. Chopper sound	98.97%	99.25%	96.26%	94.59%	93.20%	81.58%
Woman's voice vs. Music sound	99.01%	98.74%	96.26%	89.95%	92.73%	69.83%

5.3 Impact of source localization error on PSS

As shown in Table 5.1, the separation effectiveness is influenced by the sound source locations determined in the localization procedure. Therefore it is necessary to test the separation at various localization errors to get a better understanding of the BSLS method. In this section, the error of the localization was systematically generated, and they are introduced to the PSS procedures, and the correlation coefficients of the results based on the locations with errors are calculated.

Suppose the location of the benchmark location is $\vec{r}_{benchmark}$, and with the definition of error in Equation (2.15), there can be infinite number of locations $\vec{r}$ which have the same error value, and there are distributed on a spherical surface of radius *d* centered at $\vec{r}_{benchmark}$. Some of the locations $\vec{r}$ may have stronger impact on the PSS procedure than the others. However, it is obvious not possible to test them one by one. In this dissertation, six $\vec{r}$ were selected, where the radius *d* parallel to the Cartesian coordinates are added to $\vec{r}_{benchmark}$. Thus they can expressed as $\vec{r}_i = \vec{r}_{benchmark} + \vec{d}_i$, where i=1, 2, …, 6, $\vec{d}_1 = (d, 0, 0)$, $\vec{d}_2 = (-d, 0, 0)$, $\vec{d}_3 = (0, d, 0)$, $\vec{d}_4 = (0, -d, 0)$, $\vec{d}_5 = (0, 0, d)$, $\vec{d}_6 = (0, 0, -d)$. The worst separated results by the six locations were used in the survey.

Table 5.2 demonstrates the survey results of the correlation coefficients, and Figure 5.8 uses the same data and plots them in a three-dimensional scale. The two sound source signals used in this test are woman's and man's voice, respectively. $\mathbf{E_1}$ and $\mathbf{E_2}$ indicate the error generated for the localization of these two sources, and the percentage values listed in the table are the correlation coefficients of the separated woman's voice. Therefore Table 5.2 and Figure 5.8 give us a general idea how the localization of the 1st and 2nd source influence the

reconstruction of the 1st signal. Values above 90% in the table are highlighted with yellow color, which means the reconstruction is completed and very clear.

Results show that the reconstruction effectiveness of the 1st signal is more dependent on the accuracy of the 1st source location than the 2nd one. If the location of the 1st signal is correctly located, the PSS program can lead to a satisfactory separated signal for the 1st source even when the error of the 2nd source location is 100%, which means the position of 2nd source is incorrect. One the other hand, when the error of 1st source location increases, the correlation coefficient of the 1st source result decreases rapidly. In conclusion, for the BSLS algorithm, the successful reconstruction of one of the sources is majorly dependent on the accuracy of localization of this particular source, while the localization accuracies of other sources have limited influence on it.

What is more, sound source separation doesn't have high constraint on the accuracy of the source locations. It can be seen from Table 5.2 that even the errors of both source locations are as high as 35%, the correlation coefficient can still reach 90.10%, which means the separation works good.

Table 5.2. Correlation coefficients in separated woman's voice as shown in Figure 4. $\mathbf{E_1}$ and $\mathbf{E_2}$ are errors in locating woman's and man's voices, respectively. The highlighted area implies the region where PSS can produce satisfactory sources separation with 90% and above correlation.

γ_{s,s_0} E_1 \ E_2	0%	17%	35%	52%	69%	87%	100%
0%	99.91%	99.40%	98.55%	98.20%	97.97%	98.30%	97.84%
17%	97.19%	96.84%	95.80%	94.88%	94.51%	95.02%	94.10%
35%	93.16%	91.76%	90.10%	88.63%	89.06%	90.03%	87.20%
52%	88.32%	86.46%	84.99%	82.82%	86.32%	82.87%	81.36%
69%	81.52%	78.99%	79.36%	77.25%	75.64%	74.55%	77.21%
87%	74.61%	70.80%	75.21%	70.44%	64.82%	67.79%	71.90%
100%	70.17%	65.44%	68.79%	64.96%	62.84%	67.25%	67.86%

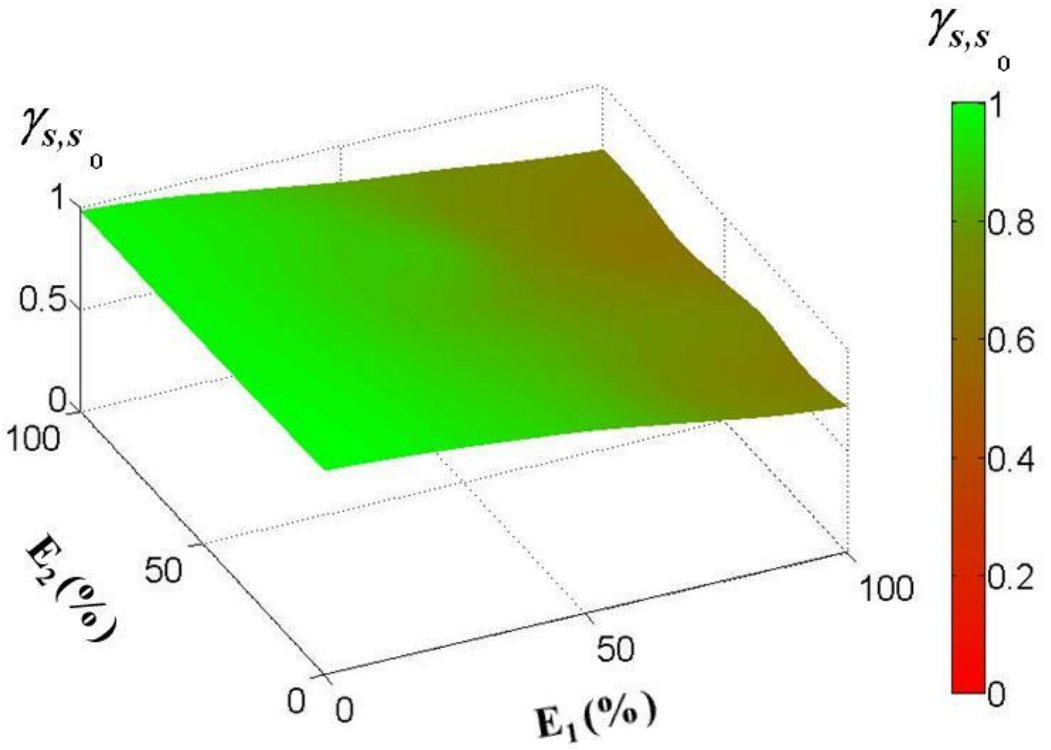


Figure 5.8. Surface contour plot of the correlation coefficients for separated woman's voice using PSS subject to errors in locating both sources for woman's and man's voices.

CHAPTER 6

NEURAL SIGNALS LOCALIZATION INSIDE BRAIN AUDITORY STRUCTURE

6.1 Neural signal de-noising

An important aspect in any signal processing is to sanitize the directly measured data, also known as de-noising signals. This is no exception to analyzing neural signals. De-noising algorithm is applied to the neural signals to eliminate the unwanted background such as the electrical noise, because the random fluctuating of the background may cause the error in the source localization algorithm. In this dissertation, averaging methodology is applied to the time domain raw data, and the random fluctuations in the background noise are suppressed during the averaging process. The averaging algorithm is based on the assumption that neural responses measured at a fixed point resulting from stimulus acting at any source position remain the same, if the stimulus remains the same. In other words, the neural responses retain similar shapes, and their changes are primarily caused by the interferences from random background noises. In the experimental validations, the same stimulus is repeated multiple times, and a group of raw data responding to the same stimulus is obtained. Meanwhile, random background noise may change each time. So by taking averages of these data, it is possible to suppress the random fluctuations due to background noise contained in the data set, resulting in a cleaner set of data.

Assume the measurement was repeated N times, and the raw data array is expressible as $x_n(i)$, where the subscript "n" is the number of the data array, $1 \le n \le N$, and the letter "i" indicate the index number of the array. The averaging algorithm consists of three steps:

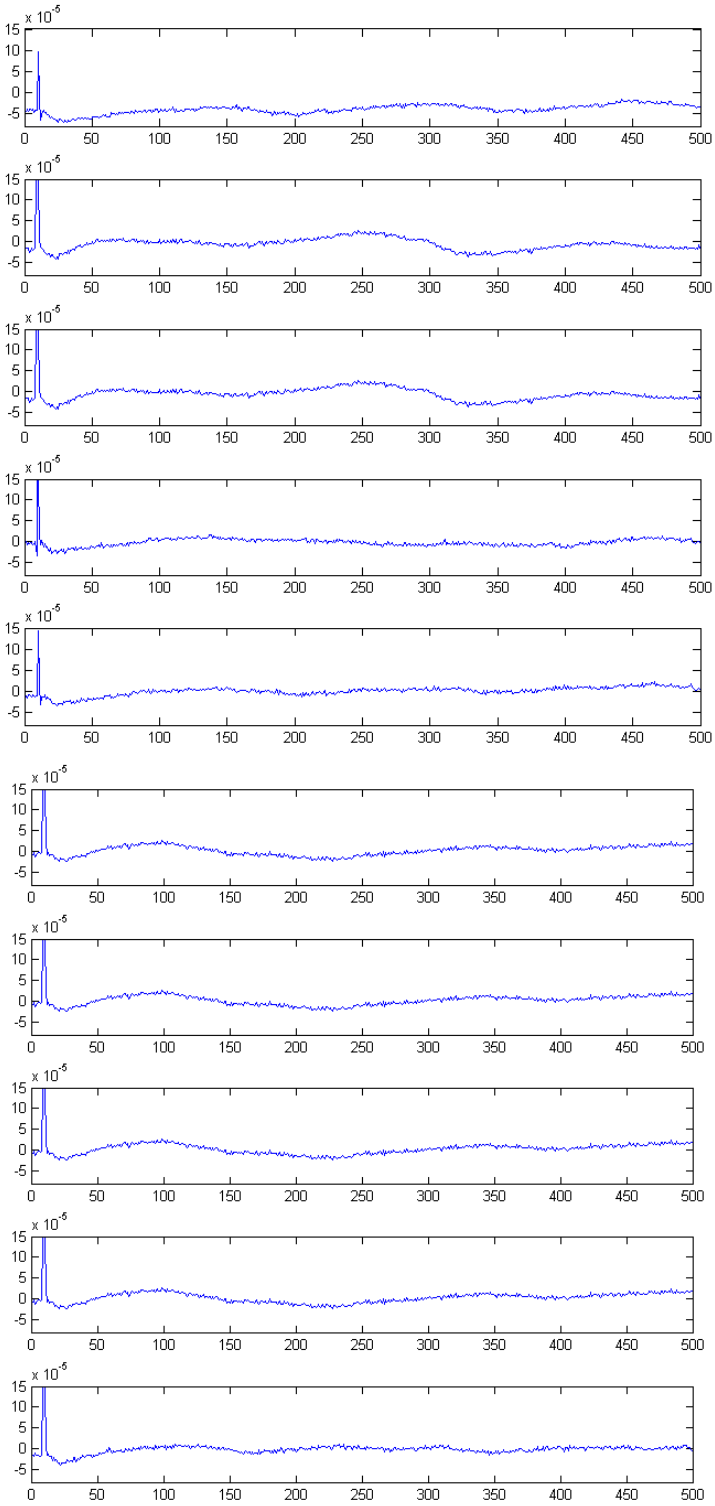
1) Synchronize the N set of data arrays using the position of the peaks. Find out the index of the peak position in each data array, and move the array back and

forward to make sure that the peak at the same index in every data array. The new data arrays can be expressed as $x_{syn,n}(i')$. Note that the index number i' is different from the index of the original data array i because the new array is moved along the time axis.

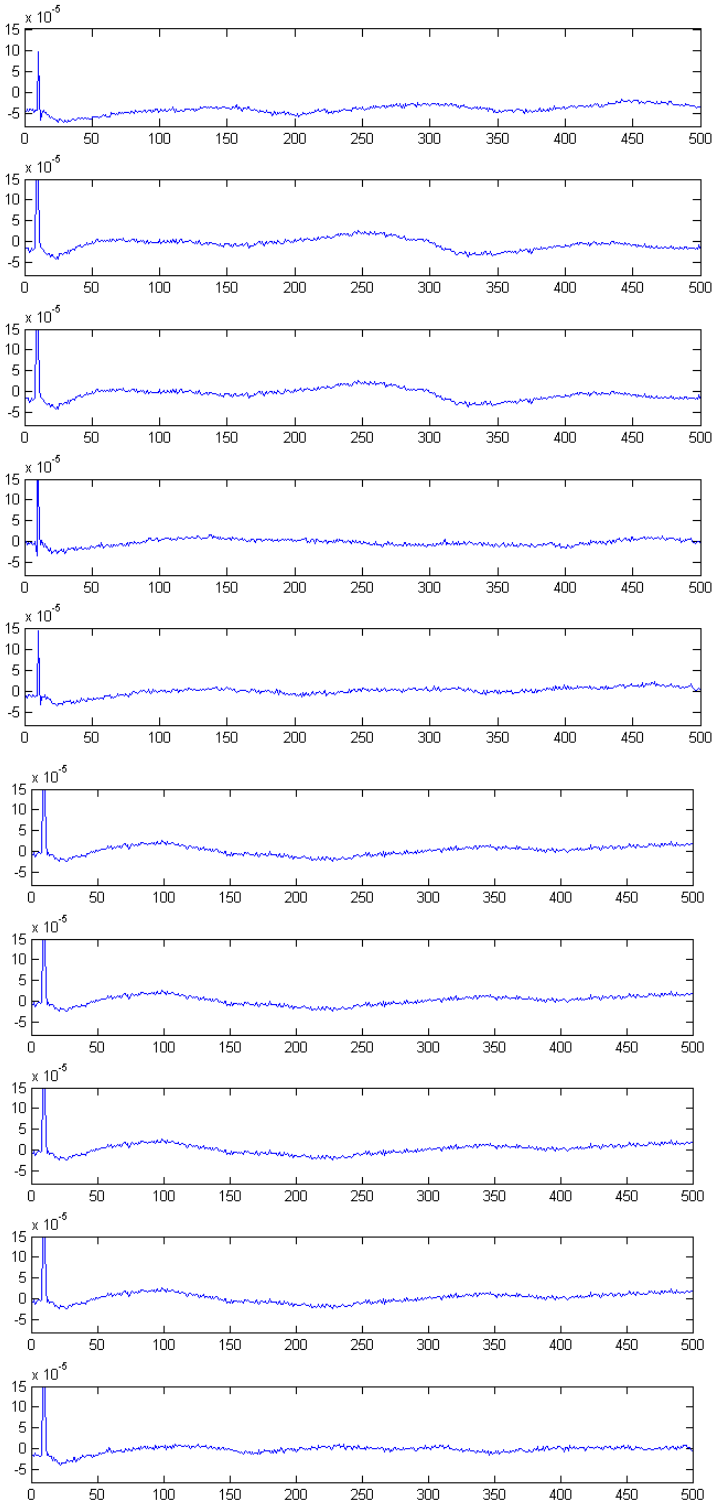
2) Calculate the mean of the synchronized arrays and get the averaged neuron response to the stimulus.

$$x_{\text{neuronresponse}} = \frac{1}{N}\sum_{n=1}^{N} x_{syn,n}(i') \tag{6.1}$$

For example, if ten measurements were conducted responding to the same stimulus, the time domain array are recorded as shown in Figure 6.1(a). Although the timing of the stimulus is controlled on purpose, the indices of the peaks still have some offsets, for example, the peak at the fourth array has some delay comparing to others. Thus, the index number of the peak at the fourth array is larger than the others. Synchronizing the data arrays leads to the results shown in Figure 6.1(b), where all the peaks at the tenth index in each array are aligned with each other. The shapes of these ten arrays are similar, but some random fluctuations in each array are observed. For example, a small peak occurs at around the 60^{th} index in the tenth array, while it doesn't happen in the other arrays. Taking the average of the ten arrays can significantly reduce the contribution of the background noise in the signal. The averaged signal (see Figure 6.2) shows that the fluctuation around the 60^{th} index is eliminated.



(a) Raw data before synchronized



(b) Raw data after synchronized

Figure 6.1 Raw data before and after synchronized

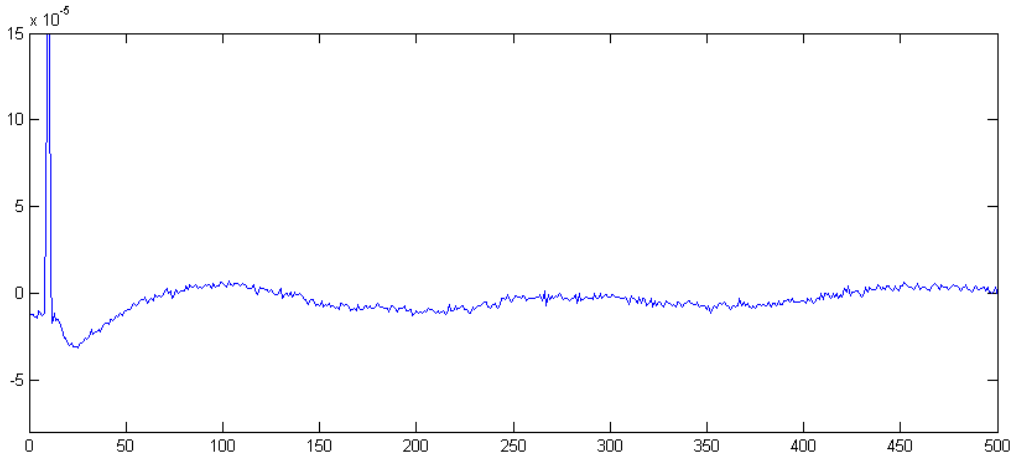


Figure 6.2 Averaged raw data

6.2 Neural signals localization using time reversal algorithm

Locating the tinnitus related hyper active neurons can help pinpoint the abnormalities in the brain auditory system, which is the first step in diagnosing tinnitus. Current technologies based on direct measurements of the neural activities can only observe neural activities at the tips electrodes, which makes it difficult to find out the precise location of the hyperactive neurons. One of the major problems of this approach is that the spatial resolution of the measured data is very low, which may lead to less effective treatment strategy for tinnitus. This difficulty is especially prevalent when the number of electrodes used in collecting neural activities is low.

Moreover, the signals traveling in the brain auditory system does not obey any decay rules, which means that the amplitude of the measured signals are not necessarily reflecting the source locations of the hyper activities. Therefore, the source locations based on observations of the amplitudes of the data measured at individual electrodes are unreliable.

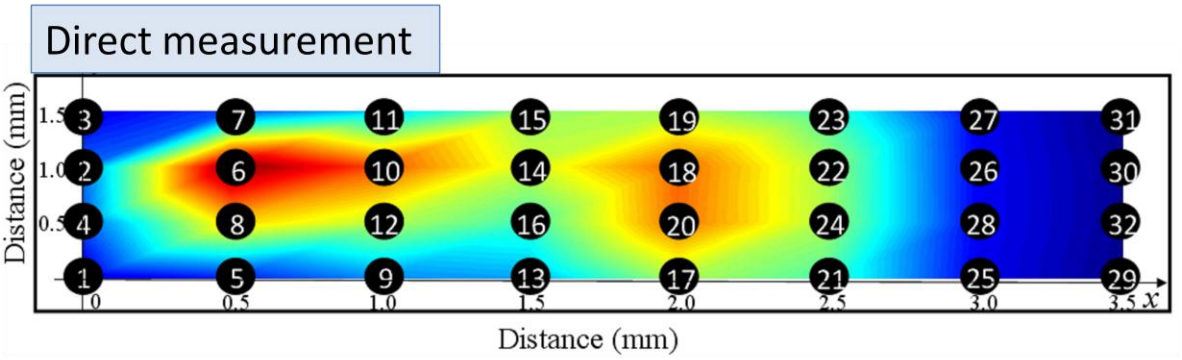
In this dissertation, TR is used to process the input raw data measured by electrodes. Basically, the measured data in each channel is inversed in the time history, each electrode is treated as an imaginary source, and the time-reversed signals are propagated from these

imaginary sources back to the real source. Finally, by summing these time-reversed and back propagating signals at every point in space, the actual source location is identified. This is because only at the true source location, will the amplitude of the signal peak out. Figure 6.3 shows the numerical validation of the TR algorithm applying to the brain auditory system to locate the sources. This simulation test mimics a 4 × 8 electrode array with 500μm between each sensor and each shank, and the locations of impulse sources are arbitrarily selected in space. The signal measured at each electrode was estimated and used as the input of the TR algorithm. Source localization was done through scanning at 100μm interval.

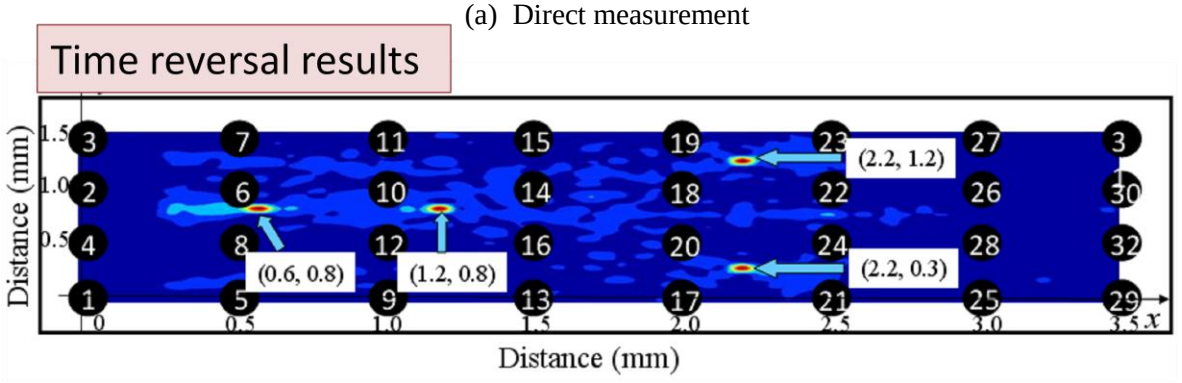
Figure 6.3(a) shows the contour of the amplitude measured at each electrode. Note that only the value exactly at the 32 sensors are reliable, and all the other color scale between the sensors are generated by interpolation, which may be reasonable, but not reliable enough to be used to determine the location of the source. Moreover, as discussed above, the amplitude is actually not necessarily reflecting the location of the source. Therefore even one can estimate the area of the source location by the direct measurement, it is still not reliable, and may not be the correct result of the source location.

Figure 6.3 (b) shows the result processed by TR. It is obvious that there are four sources in this numerical simulation, which are located at (0.6, 0.8), (1.2, 0.8), (2.2, 1.2), (2.2, 0.3) in millimeter in two-dimensional coordinate system. Unless the directly measured results in Figure 6.3 (b), the TR result can pinpoint the precise location of the sources of the hyper activities, and the spatial resolution of the results are determined by the scanning step size and the sampling frequency of the raw data, which make it possible to be much smaller than the sensor spacing. What is more, TR can not only locate the sources within the coverage of the electrode arrays, but

also estimate the position of the sources outside the sensor coverage, though there may be errors on the results outside the sensor coverage because of those results are diverged.



(a) Direct measurement



(b) TR results

Figure 6.3 Numerical simulation of TR applied in brain auditory system. (a) Signals measured by a 4 × 8 array of electrodes separated by 0.5 mm spacing. Results show that electrode positions #6, #10, #18, and #20 detected high activities. However, the spatial resolution is poor and interference of signals and noise contamination can distort the actual source locations. (b) Images produced by TR enable one to pinpoint the exact source locations, even when they fall between the electrodes. The TR results match nearly perfectly with the benchmark source locations.

CHAPTER 7

EXPERIMENTAL VALIDATION FOR NEURAL SIGNALS LOCALIZATION

7.1 Experimental setup

The experimental validation consists of two parts. The first part was the benchmark test, which was used to validate the feasibility of the TR method on source localization in the brain auditory system. Normal rats were used in this test, and ACES (define ACES) was applied at pre-determined locations to examine its effects on suppressing hyper active neural activities. The stimulus locations obtained by using the TR algorithm were compared to the benchmark stimulus locations, which were pre-selected. Also, the locations corresponding to spontaneous neural activities were identified by using the TR algorithm. The benchmark test was conducted by applying stimulus in the AC (define AC) of normal rats, and a 4 × 4 electrode array with 400μm sensor spacing was used to collect the raw data at 3,051 Hz sampling frequency.

The second part of the experimental validation was conducted on rats with and without tinnitus. The goals of this test are to: 1) compare the spontaneous activities of the tinnitus positive rats to those of normal rats, and 2) find out the exact positions of the hyper active neurons. Moreover, ACES treatments were performed on both normal and tinnitus positive rats. Spontaneous activities were collected again after ACES to see if ACES has any impact on suppressing the neural activities, especially for tinnitus positive rats. The second part of the experimental validation was also performed in the AC. A 2 × 8 electrode array with 400μm sensor spacing was used to record the raw data at 3,051 Hz sampling frequency.

7.2 Benchmark test in AC

Benchmark test was conducted in the AC of normal rats to validate the time reversal algorithm in the brain auditory structure. During the benchmark test, ACES was applied to pre-determined locations in the AC and was used to estimate the hyper active neural activities. The stimulus were generated at every 0.5 second from 50 – 500μA with a step of 5μA, 10 impulses per level. The raw data were collected by a 4 × 4 electrode array with 400μm between each sensor and each shank (See Figure 7.1), and the sampling frequency was as 3,051 Hz.

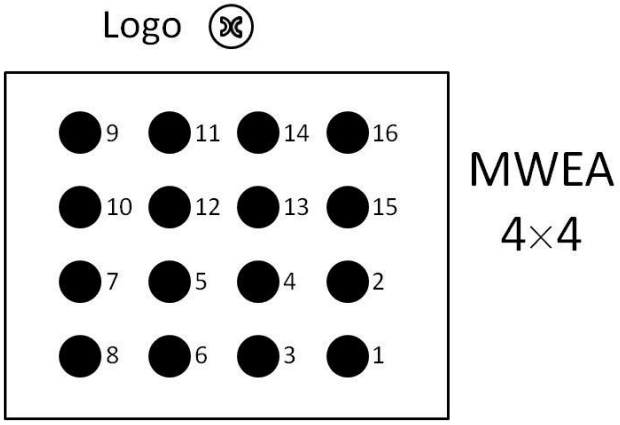


Figure 7.1 Electrode arrays used in the benchmark test. A 4×4 electrode arrays were used in the benchmark test in AC.

Two cases were measured in the benchmark test: the first was the spontaneous neural activities of normal rats without any stimulus; and the second was the neural activities during stimulus. After the neural activities were collected in these two cases, the resultant raw data were taken as the input to the TR algorithm. There were several blocks in each case, and a five-minute long data were collected in each block to be processed. TR algorithm was conducted 60 times on each block, in which each calculation represented a five second length raw data. The source localization was done through scanning at 40μm interval, which was 1/10 of the sensor spacing.

Results of the two cases were compared, and those results for the cases during ACES were compared to the benchmark stimulus location.

Note that 3,051 Hz is the highest sampling frequency possible for the data acquisition system to collect multi-channel signals over the required lengths of time histories in these tests. The travel speed of signals in AC is estimated to be 0.08 m/s, based on the existing data collected in APRL. Therefore in the scanning process, the highest spatial resolution achievable is equal to $0.08 / 3051 = 26 \times 10^{-6}$ m or 26 μm. The actual scanning step size may be equal to or larger than 26 μm, otherwise the results are not reliable. In this case the scanning step size is set at 40μm. If a higher sampling frequency is used in this case, a higher spatial resolution can be achieved because the scanning step size can be smaller.

The procedure of the data gathering in the benchmark test can be seen in Figure 7.2. Firstly, spontaneous activities without any stimulus were measured in the AC and the raw data were saved as Block 3. Secondly, the neuron activates during ACES were tested twice and the raw data were saved as Block 5 and 6, respectively. Thirdly, the spontaneous activities without stimulus were measured again as Block 7. Fourthly, the neural activities during ACES were measured as Block 9 and 10. Finally, the spontaneous activities were measured again as Block 11. As shown in Figure 7.2, there are total three blocks of the spontaneous activities, and four blocks of data collected during the ACES.

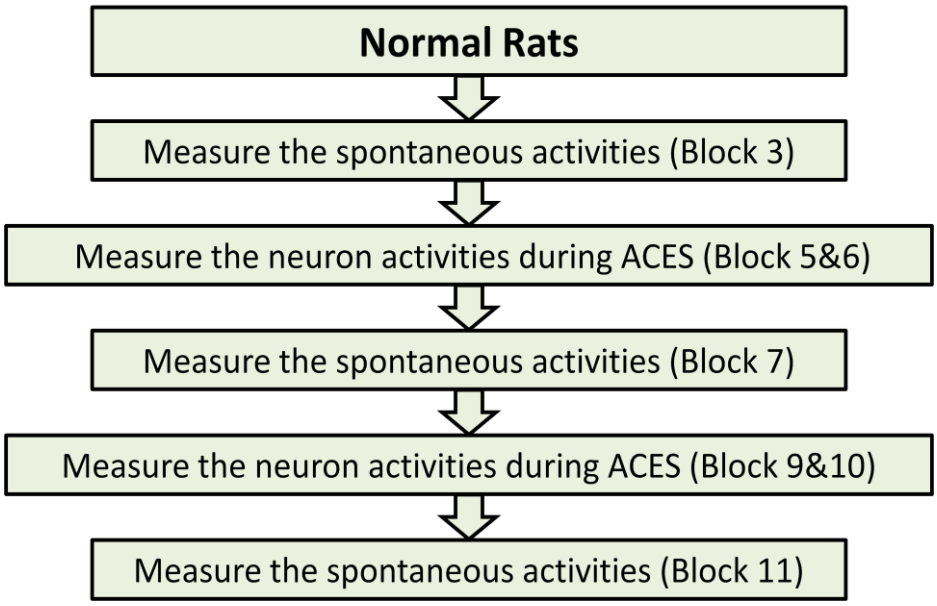
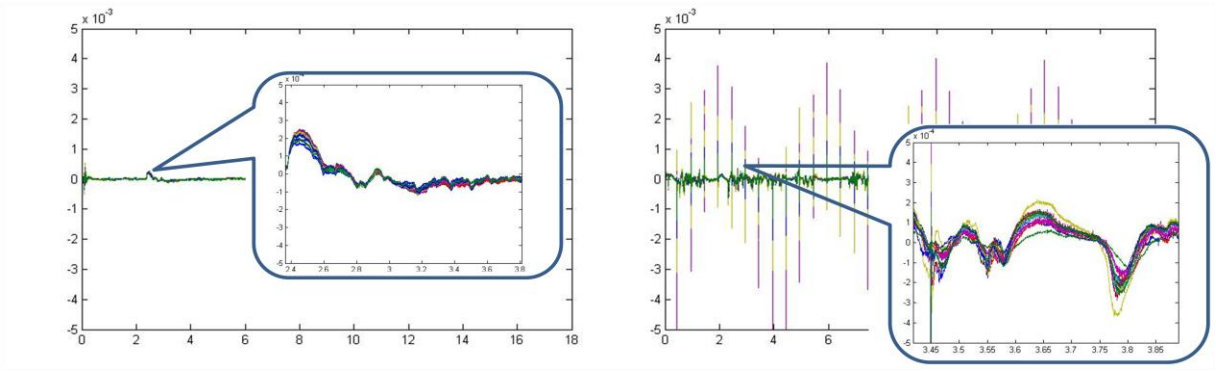


Figure 7.2 Benchmark test flowchart

The raw data collected in the two cases are illustrated in Figure 7.3. Figure 7.3 (a) shows the spontaneous activities in time domain, and the amplitude of the signals are very small. In the source localization procedure, the TR algorithm is applied to the signals at the spontaneous activities every five second and attempt to find a reasonable source based on that period of raw data. After 60 times of calculation, accumulative neural activities along the five minutes are distributed in space, which may denote the representation of the synaptic and discharging activities in the brain auditory structure.

Figure 7.3 (b) shows the activities during the stimulus, and the significant peaks represent the artifacts of the stimulus. As the stimulus at each level repeated ten times, the averaging algorithm is added before the TR procedure to eliminate the background noises. The averaging was conducted on every level at each electrode, which includes ten responses of the ACES, 0.5 second in each between. Moreover, the artifacts of the ACES at each point happen simultaneously with extremely high amplitude, but it is not the response of the neural activities

and should not be counted during TR. Therefore the artifacts were zeroed before the TR procedure. Similar as the TR procedure for the spontaneous activities, TR was also applied 60 times to the processed time domain signals in the case during the stimulus, and a source distribution in space was obtained which represents the neural activities within five minutes.



(a) Spontaneous activities signals (b) Neural activities during ACES

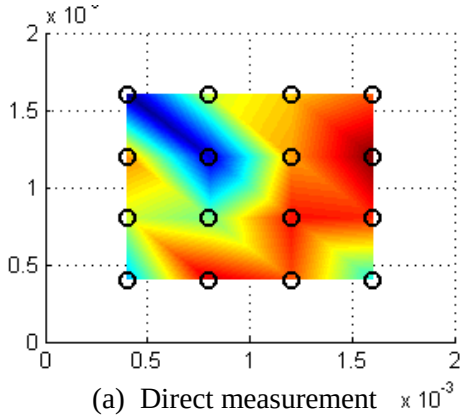
Figure 7.3 Time domain signals measure at spontaneous activities and during ACES. (a) shows the raw data of the spontaneous activities in time domain. (b) shows the raw data of the neural activities during ACES. The high peaks in the figure indicate the ACES.

7.2.1 Spontaneous activities in AC

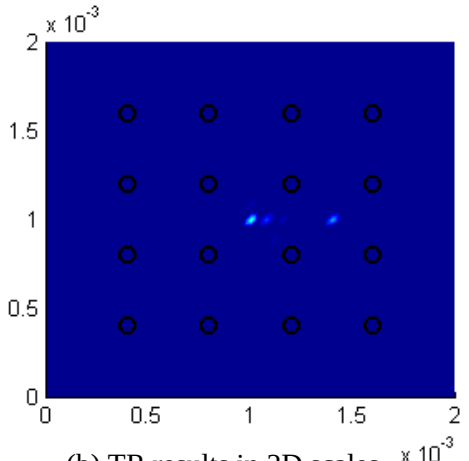
Figure 7.4 to 7.6 shows the measurements and the TR results of spontaneous activities in AC. Figure 7.4 (a) demonstrates the contour color map of the amplitude in the directly measured raw data in Block 3. The spatial resolution of the measurement is no better than the sensor spacing, and it cannot provide any information outside the coverage of the sensors. Figure 7.4 (b) and (c) show the TR results of Block 3 in two and three dimensional scale respectively. The scanning step size is 40μm and the scanning coverage extended the sensor coverage 400μm at each edge. Note that the scale range are fixed and the same in the spontaneous activities and the

neural activities during stimulus for easy comparison. It can be seen from Figure 7.4 that the neural activities of Block 3 are random and very low.

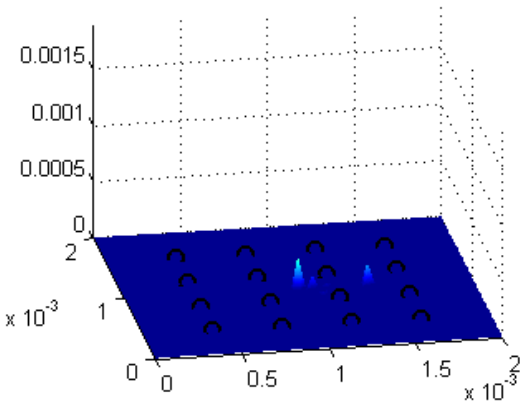
Figure 7.5 and 7.6 also show the measurement and TR results of spontaneous activities, which are obtained by the data of Block 7 and Block 11. Block 7 is the spontaneous activities measured between two set of ACES, and Block 11 is the post ACES measurement. It can be seen from Figure 7.4 to 7.6 that the spontaneous activities of normal rats are randomly distributed and relatively low.



(a) Direct measurement

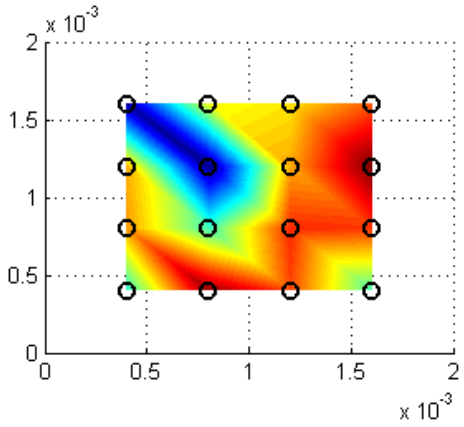


(b) TR results in 2D scales

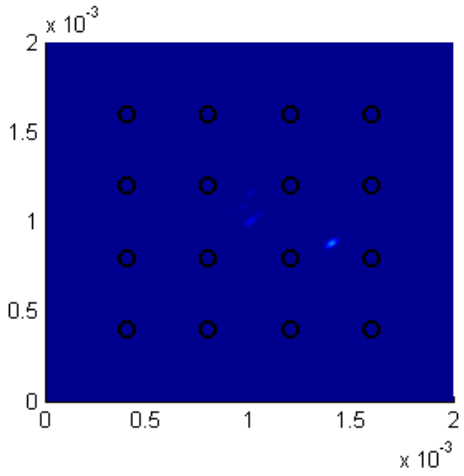


(3) TR results in 3D scales

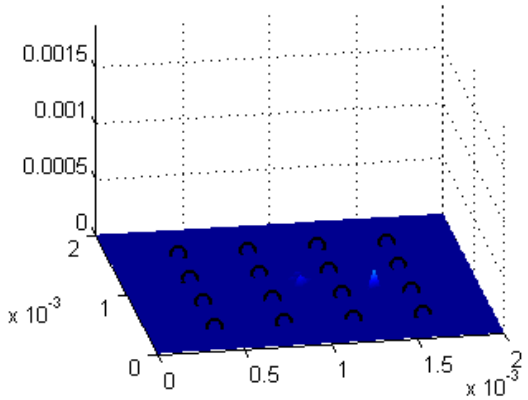
Figure 7.4 Direct measurement in Block 3 and its TR results for spontaneous activities in AC. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are relatively low.



(b) Direct measurement

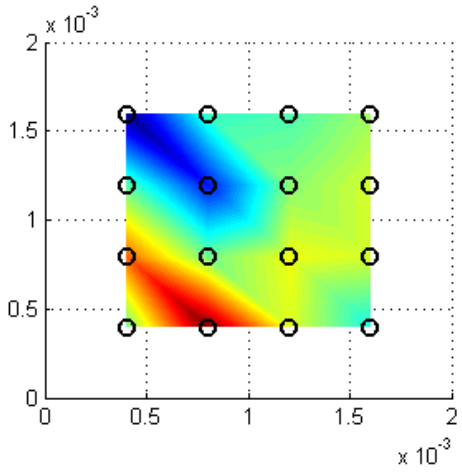


(b) TR results in 2D scales

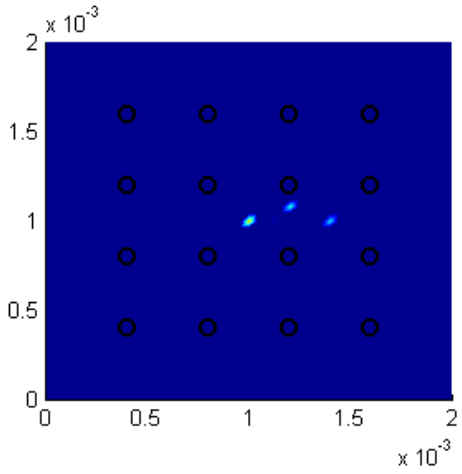


(3) TR results in 3D scales

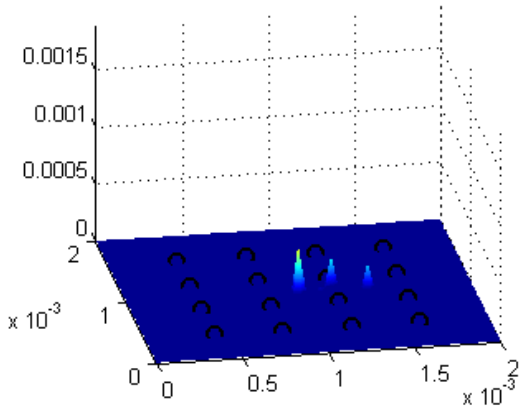
Figure 7.5 Direct measurement in Block 7 and its TR results for spontaneous activities in AC. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are relatively low.



(c) Direct measurement



(b) TR results in 2D scales



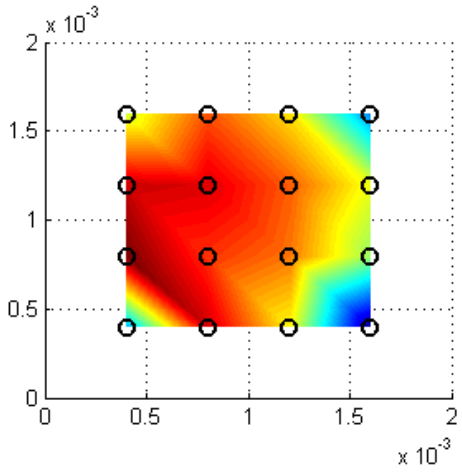
(3) TR results in 3D scales

Figure 7.6 Direct measurement in Block 11 and its TR results for spontaneous activities in AC. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are relatively low.

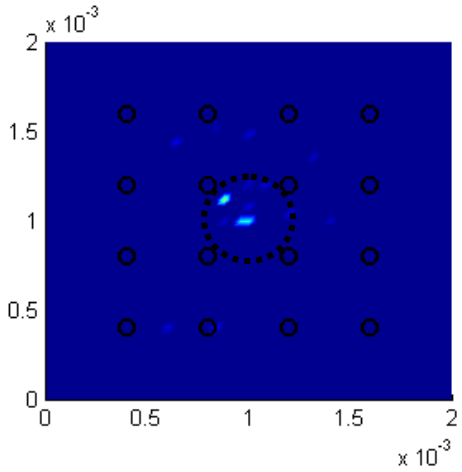
7.2.2 Neural activities in AC during electronic stimulus

Figure 7.7 to 7.10 demonstrate the measurement and source localization results of the blocks during ACES. It can be seen that the plotting of the directly measurements at each electrode cannot tell the correct source position. However, the TR results show that the amplitude of the activities around the center area is much higher than those in spontaneous activities. Note that the stimulus is accomplish at the center of the electrode arrays, and the black dashed circle indicates the position of the stimulus. The dimension of the ACES is about the same size as the sensor spacing, therefore any peak in the TR results that falls inside the black circle is correctly indicating the source signals.

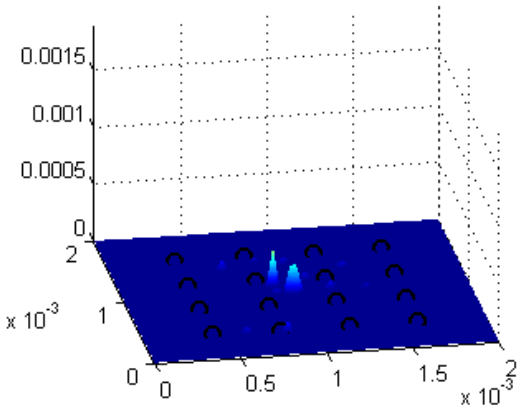
The results from Figure 7.4 to 7.10 indicate that the neuron responses to the ACES are much higher than spontaneous activities, and TR can successfully locate the position of hyper active neurons.



(d) Direct measurement

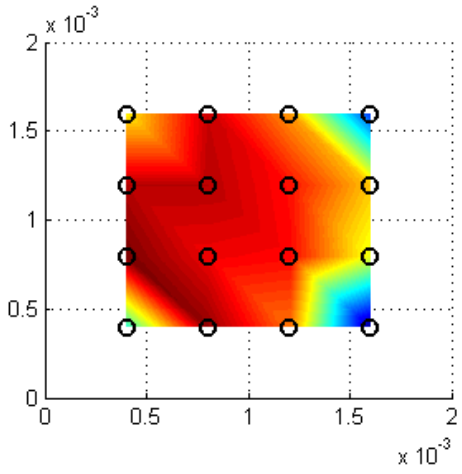


(b) TR results in 2D scales

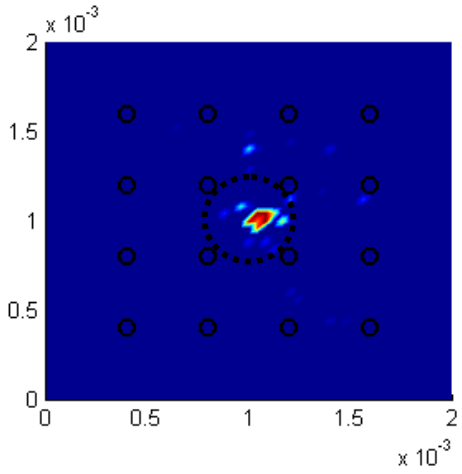


(3) TR results in 3D scales

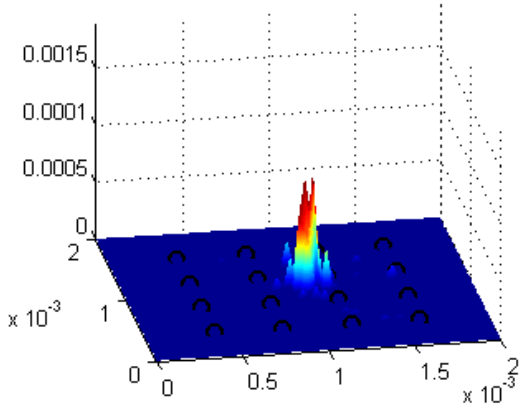
Figure 7.7 Direct measurement in Block 5 and its TR results for neural activities in AC during ACES. Black dots indicate electrode locations, and dashed circle shows the position of the stimulus. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC concentrates at the area of stimulus and the amplitude are high.



(e) Direct measurement

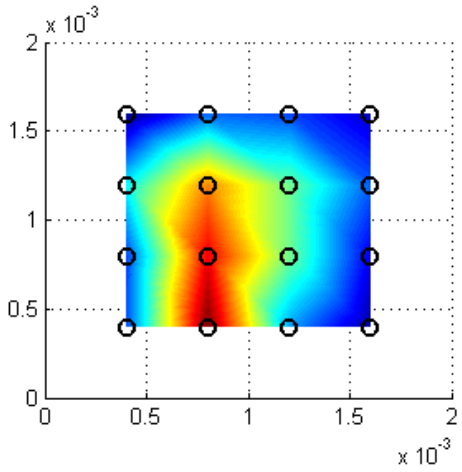


(b) TR results in 2D scales

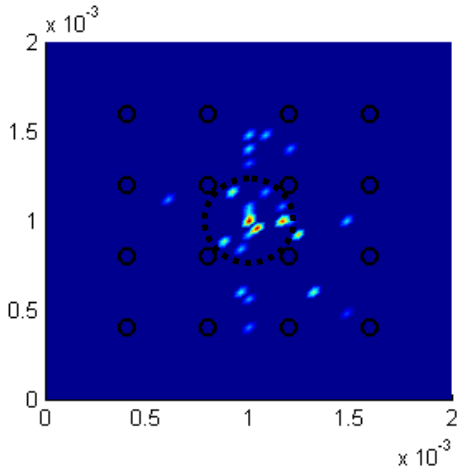


(3) TR results in 3D scales

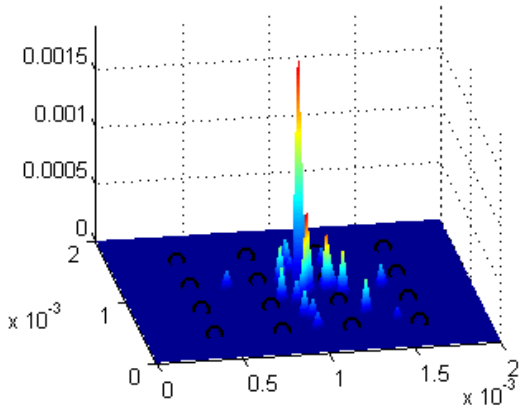
Figure 7.8 Direct measurement in Block 6 and its TR results for neural activities in AC during ACES. Black dots indicate electrode locations, and dashed circle shows the position of the stimulus. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC concentrates at the area of stimulus and the amplitude are significantly high.



(f) Direct measurement

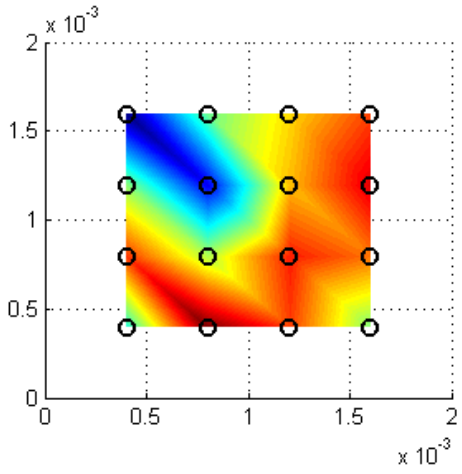


(b) TR results in 2D scales

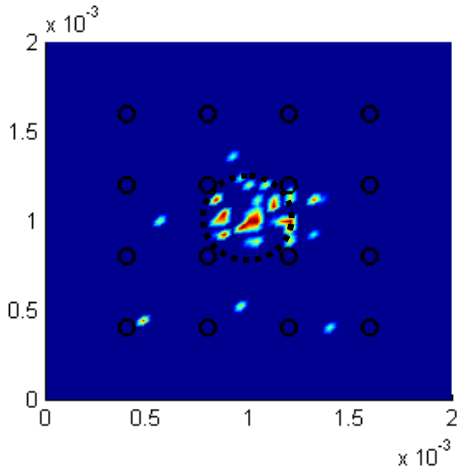


(3) TR results in 3D scales

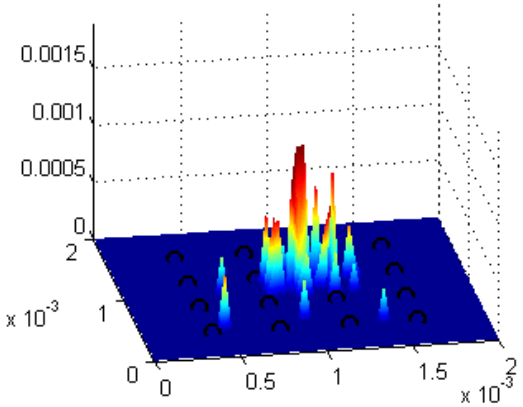
Figure 7.9 Direct measurement in Block 9 and its TR results for neural activities in AC during ACES. Black dots indicate electrode locations, and dashed circle shows the position of the stimulus. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC concentrates at the area of stimulus and the amplitude are significantly high.



(g) Direct measurement



(b) TR results in 2D scales



(3) TR results in 3D scales

Figure 7.10 Direct measurement in Block 10 and its TR results for neural activities in AC during ACES. Black dots indicate electrode locations, and dashed circle shows the position of the stimulus. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC locate at the area of stimulus and the amplitude are significantly high.

7.2.3 Comparison of results

As introduced above, Figure 7.4 to 7.6 indicate the spontaneous activities of a normal rat without any external stimulus, while the results in Figure 7.7 to 7.10 represent the neural activities responding to ACES. The peaks representing the spontaneous activities have a limited number and are randomly distributed in space. In addition, the amplitudes of these peaks are relatively small and even the highest one is under 0.0005.

On the other hand, the peaks in the TR results during the ACES are much more than those during spontaneous activities. The amplitudes of the peaks representing the neural activities during ACES are large and can be up to 0.0015. In other words, the neurons are much more active during the ACES than spontaneous activities. What is more, the peaks are mostly concentrated at the center of coverage area of the electrode arrays, which coincide with the position of ACES.

In conclusion, the results from Figure 7.4 to 7.10 indicate that the neuron responses to the ACES are much higher than spontaneous activities, and TR can successfully locate the position of hyper active neurons.

7.3 Impact of ACES on suppressing tinnitus related neural activities in AC

After the benchmark test, the TR algorithm is used to analyzing the raw data measured on the rats with and without tinnitus, before and after the ACES to locate the source signals and exam the effect of the ACES treatment on reducing the hyper neural activities in AC.

A 2 × 8 electrode arrays (See Figure 7.11) with 400μm sensor spacing were plugged in the AC area of the rats, as shown in Figure 7.12. The raw data of spontaneous activities were

measured at sampling frequency of 3,051 Hz, and then used as the input of TR to find the source location of the hyper activities.

In the localization procedure, one minute raw data was selected from the time domain measurement, and TR was applied to each second length. Therefore totally 60 times TR algorithm were conducted on each set of data. The source localization was done through scanning at 80μm interval, which is 1/5 of the sensor spacing, and the scanning coverage extended to 400 μm outside the edge of electrode arrays. Note that the scanning step size can be as small as 26 μm, which can lead to a high spatial resolution but at the cost of much longer computation time. As a compromise, a larger scanning step size at 80 μm is selected.

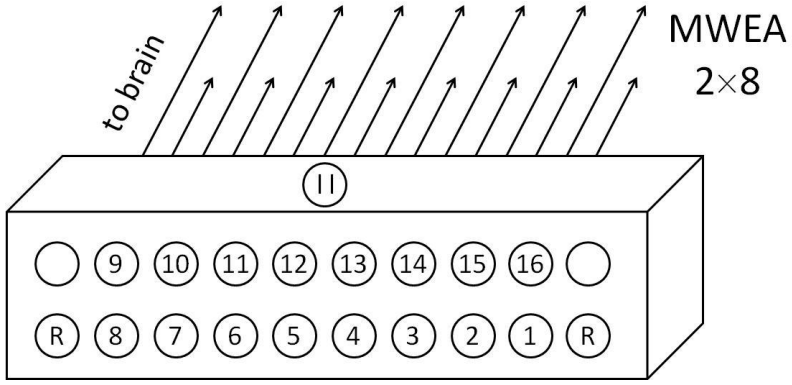


Figure 7.11 Electrode array used in experimental validation. A 2×8 electrode arrays were used in the experimental validation on the rats with and without tinnitus before and after ACES.

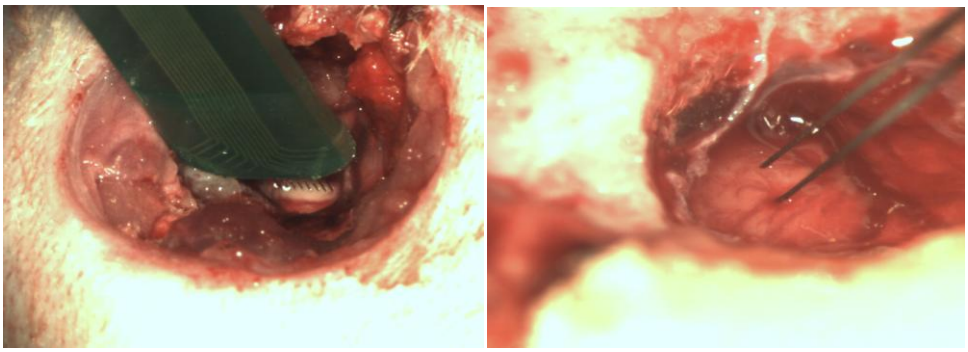


Figure 7.12 Experiment in AC area

Figure 7.13 shows the flowchart of the experimental validation procedure. Two groups of rats were employed in the test, one group is tinnitus positive, and the other group is tinnitus negative. At the beginning of the test, the spontaneous activities in time domain were collected in both groups, and TR was applied to the raw data and located the signal sources. Next, ACES were conducted on both groups of rats. Biphasic pulses at 75 μA intensity level of ACES were used in this test. The width of the pulses was 100 μs, ten pulses were generated per second, the train duration is 500 ms, and the ACES lasted for one minute. After the ACES procedure, the spontaneous activities of two groups were measured again. TR was conducted again on the measured raw data and located the sources. The TR results of two groups were compared to each other thus the differences between tinnitus positive and negative rats can be obtained. Additionally, the TR results before and after ACES in each group were also compared thus the effect of ACES can be examined.

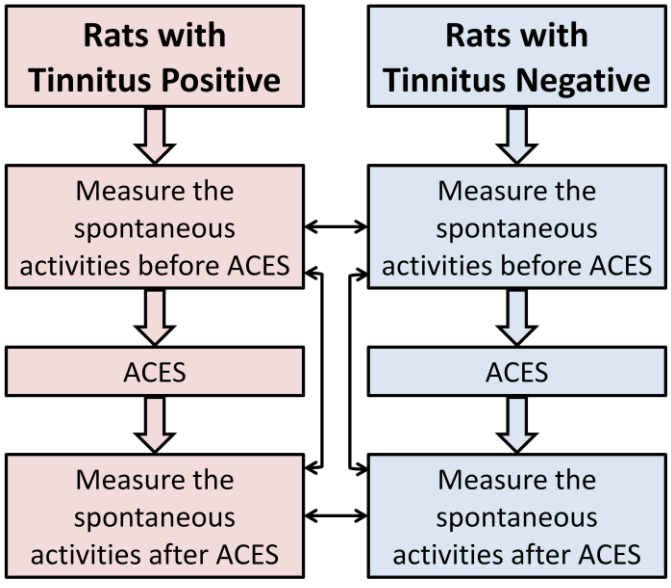
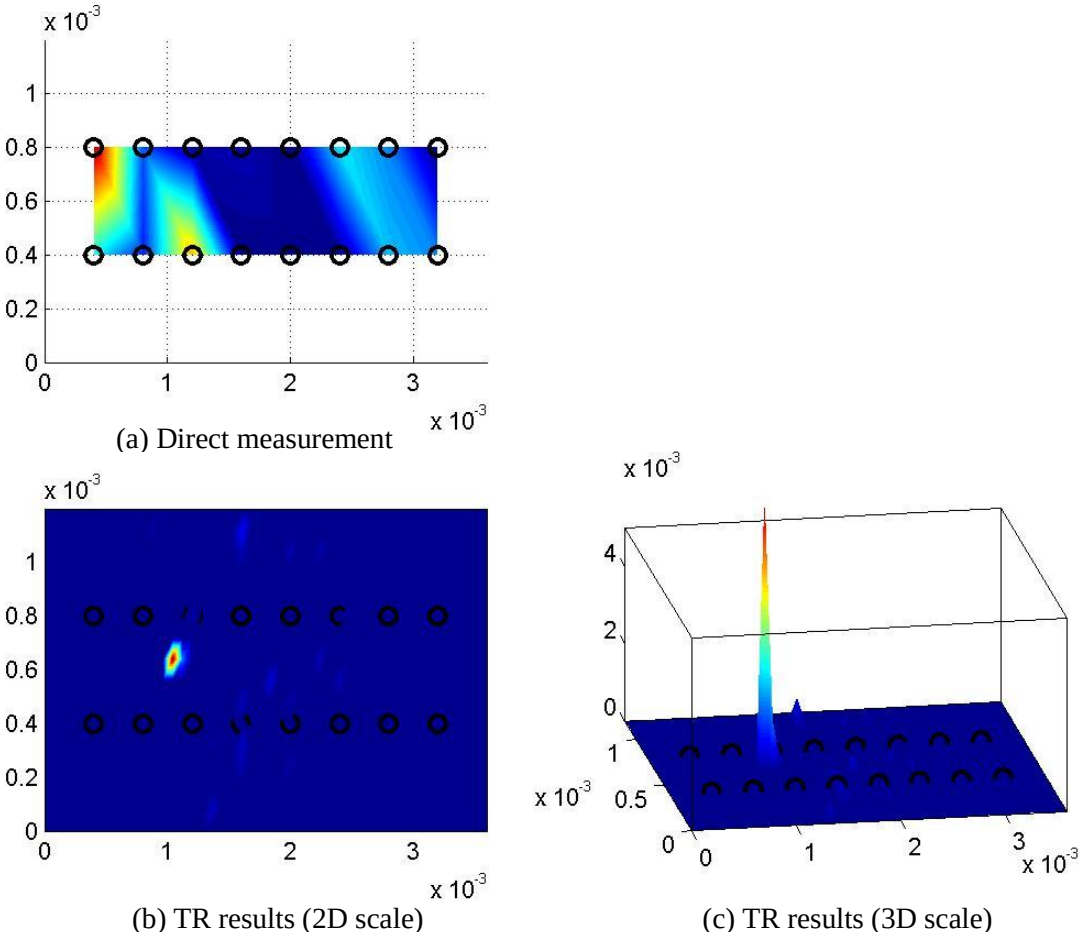


Figure 7.13 Flowchart of experimental validation on rats with and without tinnitus before and after ACES.

Data of two rats in each group are demonstrated in this dissertation. For the tinnitus positive group, the two rats are #ImpIE01 and #ImpIE11; and the rats in the tinnitus negative group are called #ImpIE06 and #ImpIE14.

7.3.1 Case 1: Spontaneous neural activities of tinnitus positive rats before ACES

Figure 7.14 and 7.15 illustrate the spontaneous neural activities of the rat #ImpIE01 and #ImpIE11 respectively, which are from the tinnitus positive group. The direct measurement in Figure 7.14 (a) and 7.15 (a) look similar, and they cannot tell the position of the hyper active signals. And as the TR result showed in Figure 7.14 (b) and (c), the source signals for the rat #ImpIE01 are mostly from a point between the electrodes #6, 7, 10, and 11, and the peak is significant which may represent the tinnitus related hyper activities. The source signals of rat #ImpIE11is also located between the electrodes #6, 7, 10, and 11, as Figure 7.15 shows, but the precise location is different from the one obtained by rat #ImpIE01. This is reasonable as the test is conducted on two rats, and difference between individuals exists.

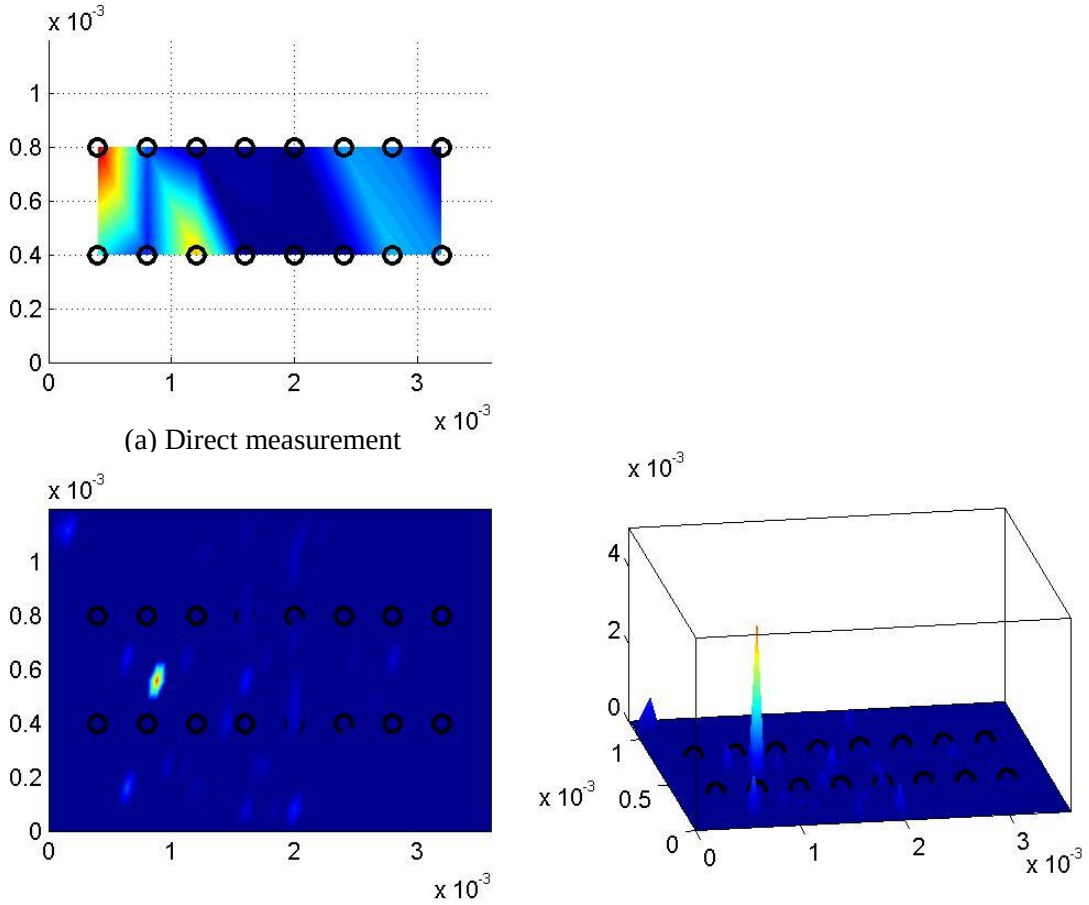


(a) Direct measurement

(b) TR results (2D scale)

(c) TR results (3D scale)

Figure 7.14 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE01. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are mainly at a particular position and the amplitude are significantly high high for the rats with tinnitus before ACES.



(a) Direct measurement

(b) TR results (2D scale)

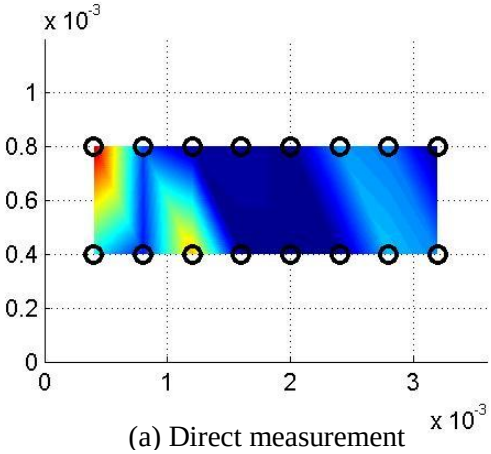
(c) TR results (3D scale)

Figure 7.15 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE11. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are mainly at a particular position and the amplitude are significantly high for the rats with tinnitus before ACES.

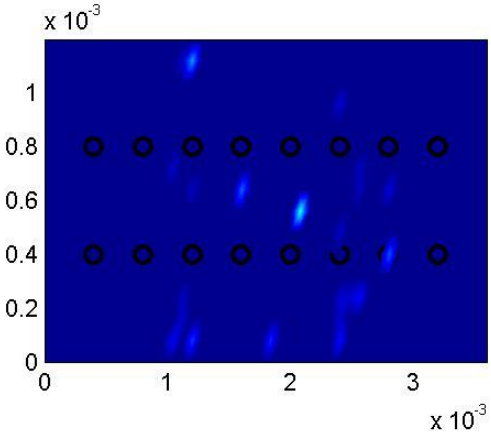
7.3.2 Case 2: Spontaneous neural activities of tinnitus negative rats before ACES

Figure 7.16 and 7.17 indicate the direct measurement and TR results of spontaneous activities in the tinnitus negative group, and the rats are named #ImpIE06 and #ImpIE14 It can be seen that the direct measurement color maps in this group are still similar to those in the tinnitus positive group, and one cannot find a reliable solution of the hyper active source positions. The TR results of tinnitus negative group demonstrated in Figure 7.16 (b), (c), 7.17 (b),

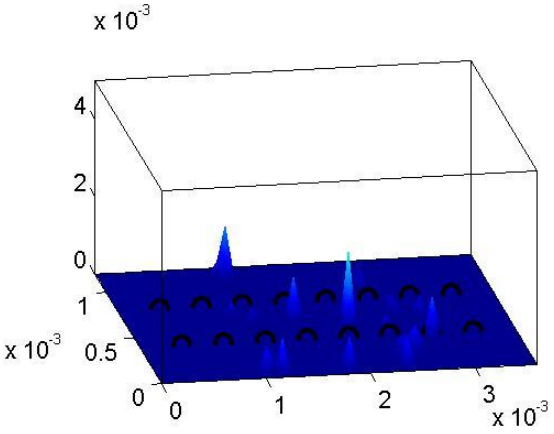
and (c) has much lower peaks than those in the tinnitus positive group, and the position of the peaks are arbitrarily located.



(a) Direct measurement

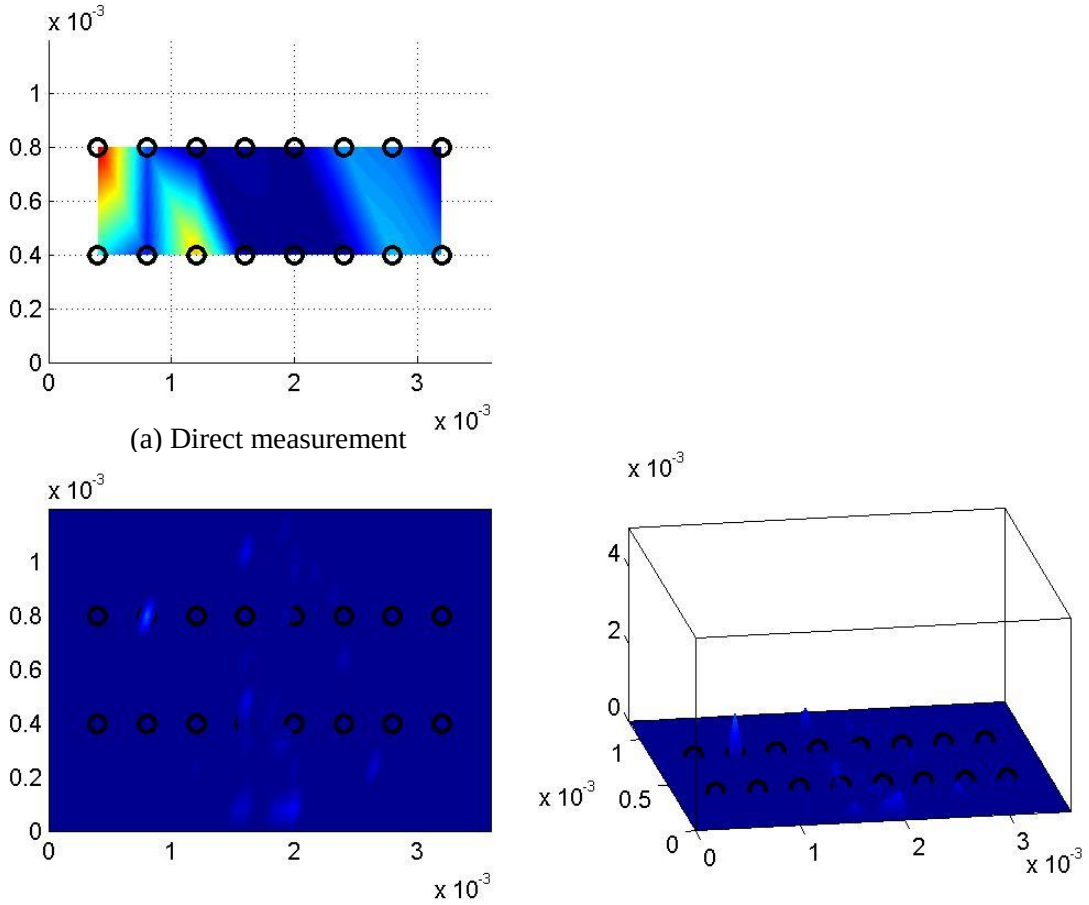


(b) TR results (2D scale)



(c) TR results (3D scale)

Figure 7.16 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE06. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are low for the rats without tinnitus before ACES.



(a) Direct measurement

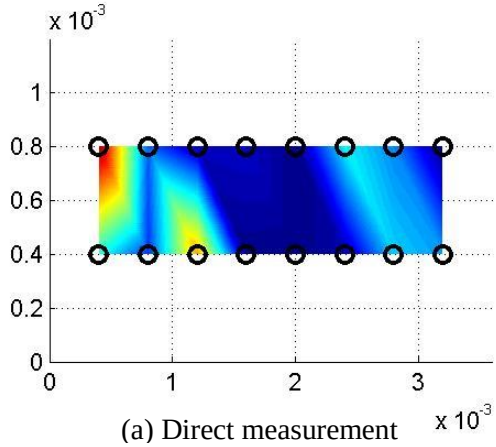
(b) TR results (2D scale)

(c) TR results (3D scale)

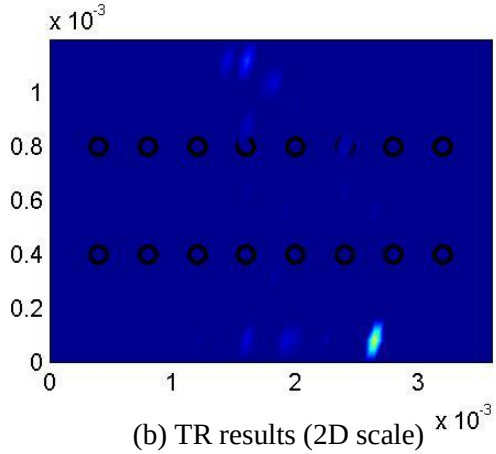
Figure 7.17 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE14. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are low for the rats without tinnitus before ACES.

7.3.3 Case 3: Spontaneous neural activities of tinnitus positive rats after ACES

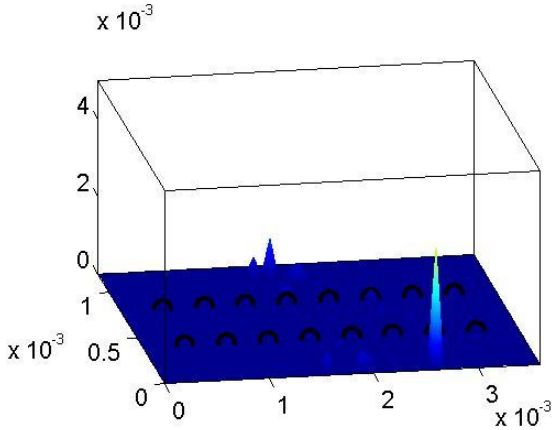
Figure 7.18 and 7.19 show the direct measurement and TR results of the tinnitus positive group after ACES. Note that ACES was applied on Channel #2 and 15 for both of the rats. The direct measurement cannot tell the position of the source signals. The amplitudes of the peaks in the TR results (See Figure 7.18 (b), (c), 7.19 (b), and (c)) are much smaller than those before ACES, which means that the neural activities of tinnitus positive rats were reduced after ACES.



(a) Direct measurement

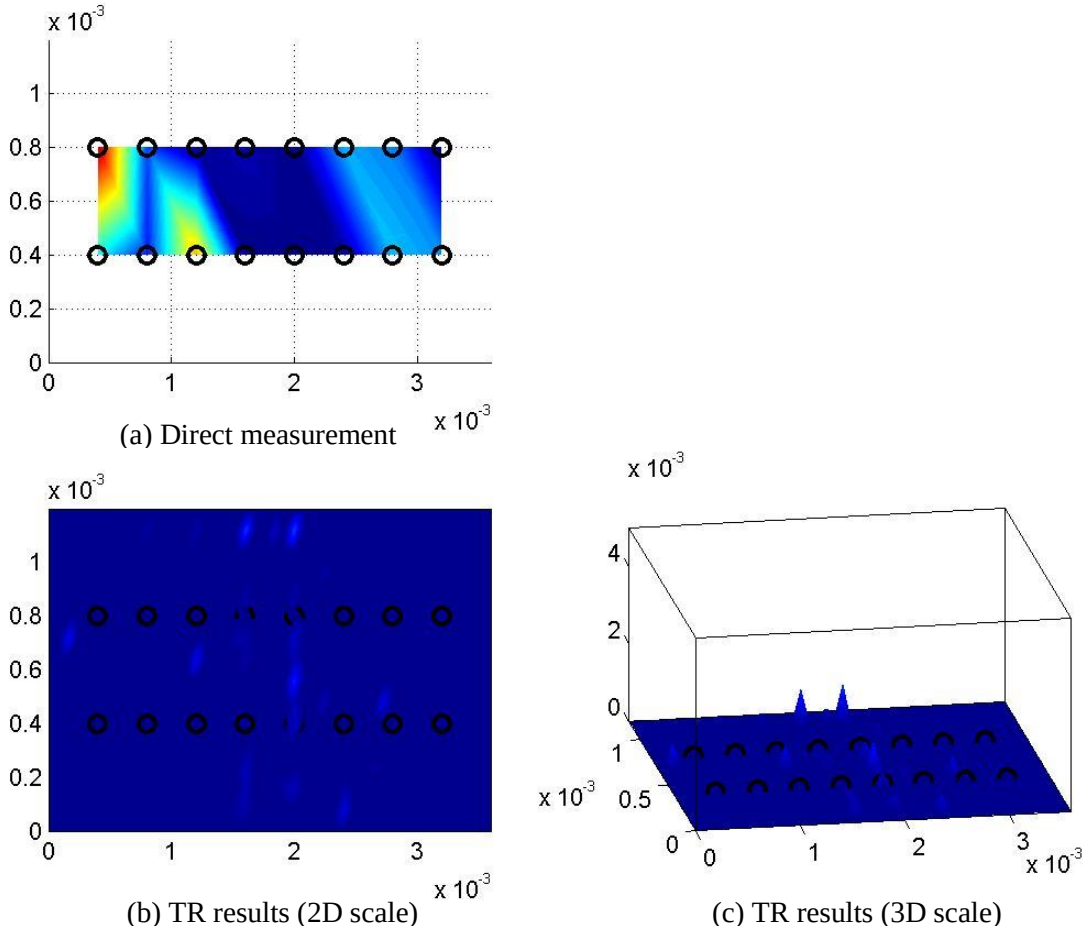


(b) TR results (2D scale)



(c) TR results (3D scale)

Figure 7.18 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE01. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are reduced for the rats with tinnitus before ACES.



(a) Direct measurement

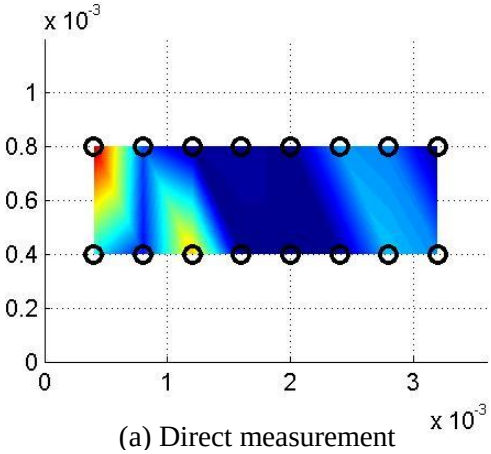
(b) TR results (2D scale)

(c) TR results (3D scale)

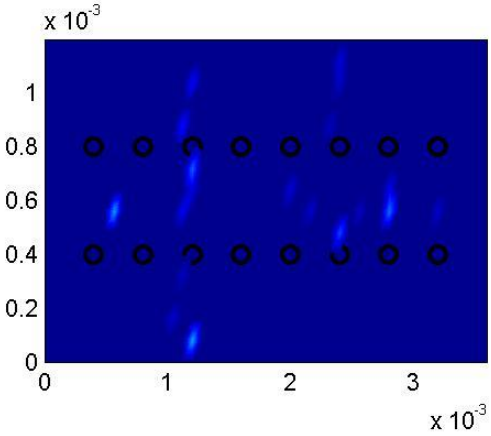
Figure 7.19 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE11. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are reduced for the rats with tinnitus before ACES.

7.3.4 Case 4: Spontaneous neural activities of tinnitus negative rats after ACES

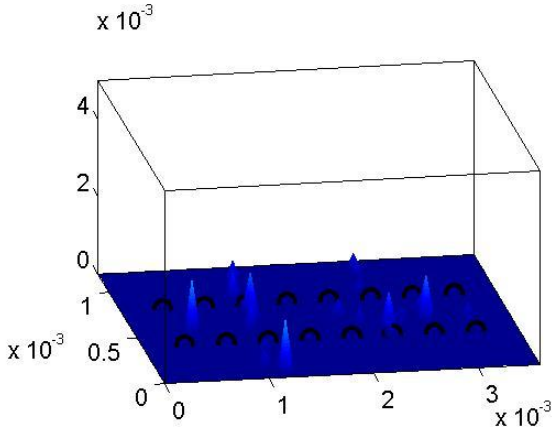
Figure 7.20 and 7.21 show the direct measurement and TR results of the tinnitus negative group after ACES. Note that ACES was applied on Channel #7 and 10 for rat #ImpIE06 and on Channel #2 and 15 for rat #ImpIE14. The direct measurement cannot tell the position of the source signals, and the neural activities after ACES are still small and random for the tinnitus negative rats.



(a) Direct measurement

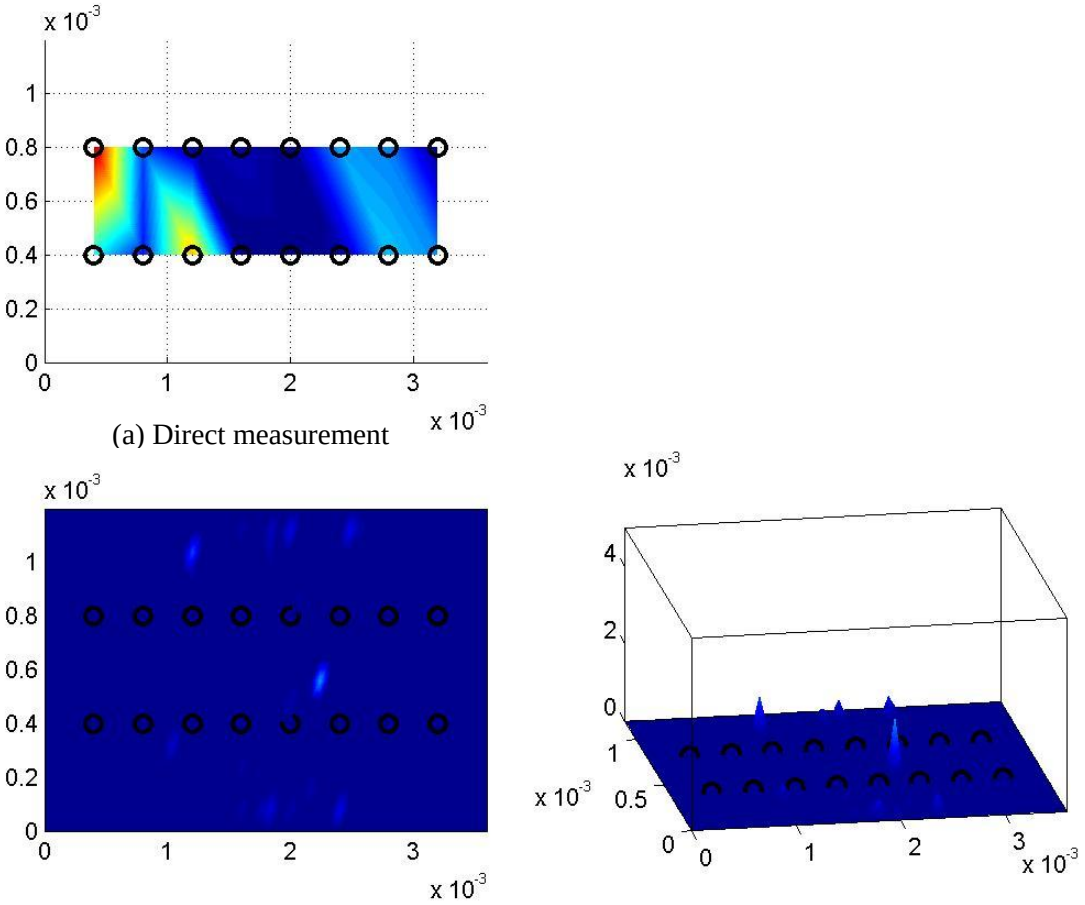


(b) TR results (2D scale)



(c) TR results (3D scale)

Figure 7.20 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE06. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are low for the rats without tinnitus after ACES.



(a) Direct measurement

(b) TR results (2D scale)

(c) TR results (3D scale)

Figure 7.21 Direct measurement in and its TR results for neural activities in AC of rat #ImpIE14. Black dots indicate electrode locations. (a) shows the direct measurement contour, (b) and (c) demonstrate the TR results in 2D and 3D scales, respectively. The TR results show that the spontaneous activities in AC are random and the amplitude are low for the rats without tinnitus after ACES.

Comparing the direct measurement and the TR results of the four cases demonstrated above, it can be seen that the tinnitus positive rats have strong activities in the AC, and the position of the hyper active neurons can be pinpointed by the TR algorithm. The spontaneous activities in the AC of the tinnitus negative rats are randomly distributed and the amplitudes are very small. ACES seems to be able to reduce the hyper neural activities in the AC for rats with tinnitus, while it doesn't have very obvious effect on those rats without tinnitus.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

This dissertation presents the innovative acoustic modeling based method for locating arbitrary and incoherent sound sources in 3D space in real time by using a minimal number of microphones, and the Point Source Separation (PSS) method for extracting target signals from directly measured mixed signals. Combining these two approaches leads to a novel technology known as Blind Sources Localization and Separation (BSLS) that enables one to locate multiple incoherent sound signals in 3D space and separate original individual sources simultaneously, based on the directly measured mixed signals. These technologies have been validated through numerical simulations and experiments conducted in various non-ideal environments where there are non-negligible, unspecified sound reflections and reverberation as well as interferences from random background noise. The advantages and limitations of the model based localization, PSS, and BSLS are summarized in Table 8.1.

Another innovation presented in this dissertation is concerned with applications of the TR algorithm to pinpoint the exact locations of hyper-active neurons in the brain auditory structure that are directly correlated to the tinnitus perception. Benchmark tests conducted on normal rats have confirmed the localization results provided by the TR algorithm. Results demonstrate that the spatial resolution of this source localization can be as high as the micrometer level. This high precision localization may lead to a paradigm shift in conventional tinnitus diagnosis, which may in turn produce a more cost-effective treatment for tinnitus than any of the existing ones.

Table 8.1 Advantages and limitations of the model based localization, PSS, and BSLS

	Advantages	Limitations
Model based localization	1. It can provide the precise locations of multiple sound sources in 3D space in real time. 2. It can handle a variety of sound signals including continuous, random, impulsive, narrow- and broadband signals. 3. It can cover a localization range up to seven times the microphone spacing.	1. It is effective in a free field. 2. Inside a confined space, its effectiveness will be affected by the presence of sound reflection and reverberation. 3. It is not suitable for locating sound sources that emit tonal sounds.
PSS	It can exactly reconstruct the source signals in terms of amplitudes and phases.	1. It needs source locations. 2. It is suitable for a free field.
BSLS	1. It can provide the precise locations of multiple sound sources in 3D space in real time. 2. It can handle a variety of sound signals including continuous, random, impulsive, narrow- and broadband signals. 3. It can cover a localization range up to seven times the microphone spacing. 4. It can exactly reconstruct the source signals in terms of amplitudes and phases	1. It is effective in a free field. 2. Inside a confined space, its effectiveness will be affected by the presence of sound reflection and reverberation. 3. It is not suitable for locating sound sources that emit tonal sounds.

8.2 Future work

The research performed in this dissertation on source localization and separation has produced satisfactory results. However, there is much room for further improvements both theoretically and experimentally.

First of all, both of the localization and separation methods introduced in this dissertation assume that the environment is a free field. Such an idealized situation does not exist in reality. There are always sound reflections and reverberation, and interferences from background noise sources. Moreover, these effects are unknown a priori, and may vary from one environment to another. Therefore, it is necessary to account for these effects in the modeling and calculations so that the results may be more robust, repeatable and accurate.

Second, experimental validations should be conducted with different types of sounds in various environments for PSS and BSLS algorithms. Moreover, more numerical simulations are needed to get a better understanding of the underlying characteristics of these methodologies.

Third, more benchmark tests should be done on using the TR algorithm to locate hyper active neurons in the brain auditory structure. In particular, tests should be performed in the AC, inferior colliculus (IC) and dorsal cochlear nucleus (DCN) to ensure the feasibility of this new approach to pinpoint the precise locations of active neurons inside the brain auditory structure.

Fourth, TR algorithm should be extended to locate neuron activities in 3D space. This can be accomplished by using implanting multi-channel electrodes in the 3D brain auditory structure, which will lead to more accurate diagnosis of the mechanisms underlying tinnitus pathology.

APPENDIX A – REAL TIME SOUND SOURCE LOCALIZATION PROGRAM

Master/Slave Design Pattern
C:\Users\AVNC\Desktop\programs\sound sources localization Folder\sound sources localization.vi
Last modified on 11/17/2011 at 14:24
Printed on 11/17/2011 at 14:25

Connector Pane

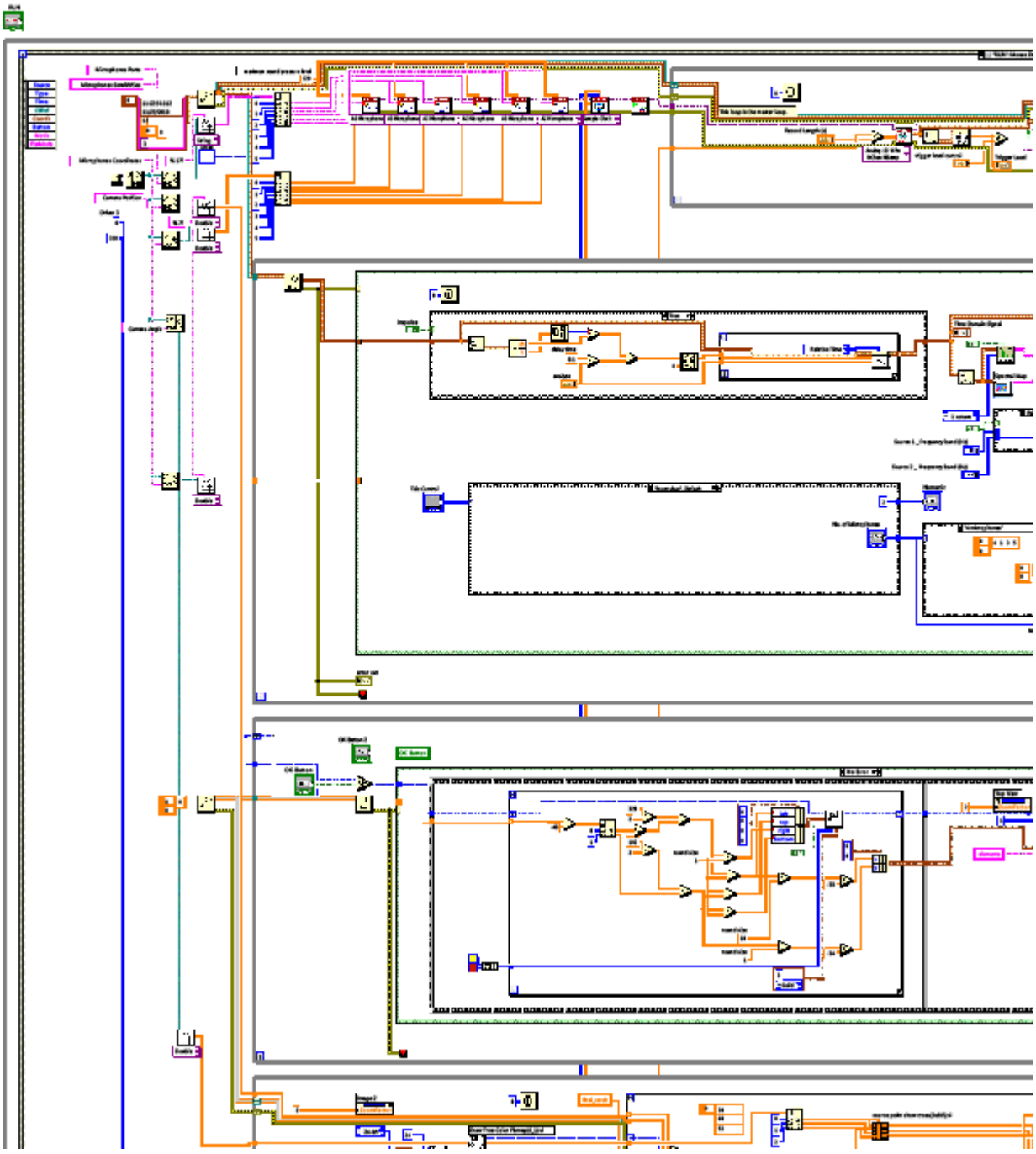
Master/Slave Design Pattern

Use this template to build a master/slave design pattern. The master loop always executes. It notifies one or more slave loops to execute their code. The slave loop(s) continue executing until they complete, then wait for another notification. Contrast this with the Producer/Consumer pattern in which the consumer loops execute only when they have data in their queue.

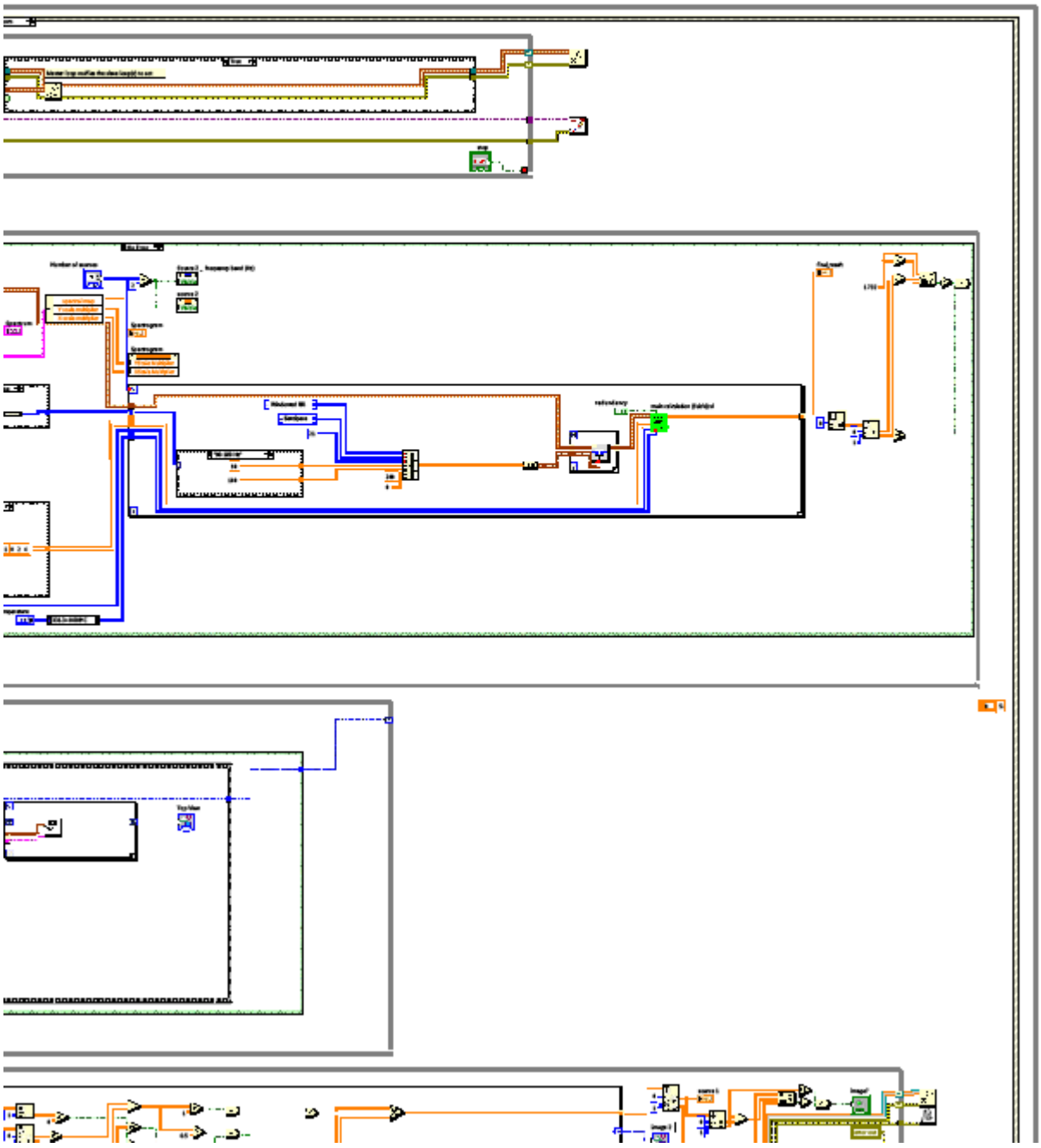
Front Panel

Master/Slave Design Pattern
C:\Users\AVNC\Desktop\programs\sound sources localization Folder\sound sources localization.vi
Last modified on 11/17/2011 at 14:24
Printed on 11/17/2011 at 14:25

Block Diagram

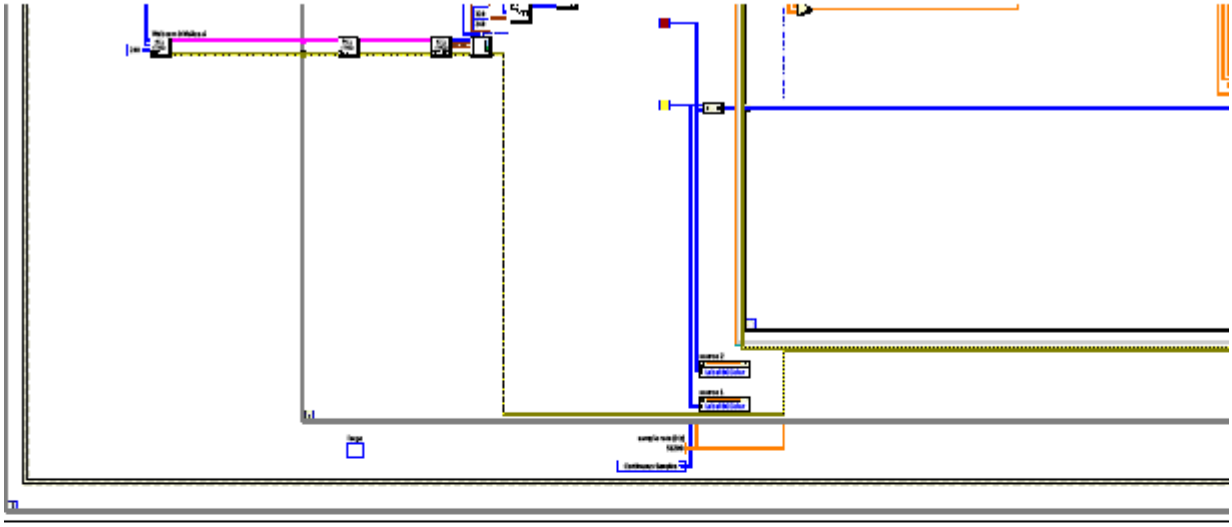


Master/Slave Design Pattern
C:\Users\AVNC\Desktop\programs\sound sources localization Folder\sound sources localization.vi
Last modified on 11/17/2011 at 14:24
Printed on 11/17/2011 at 14:25



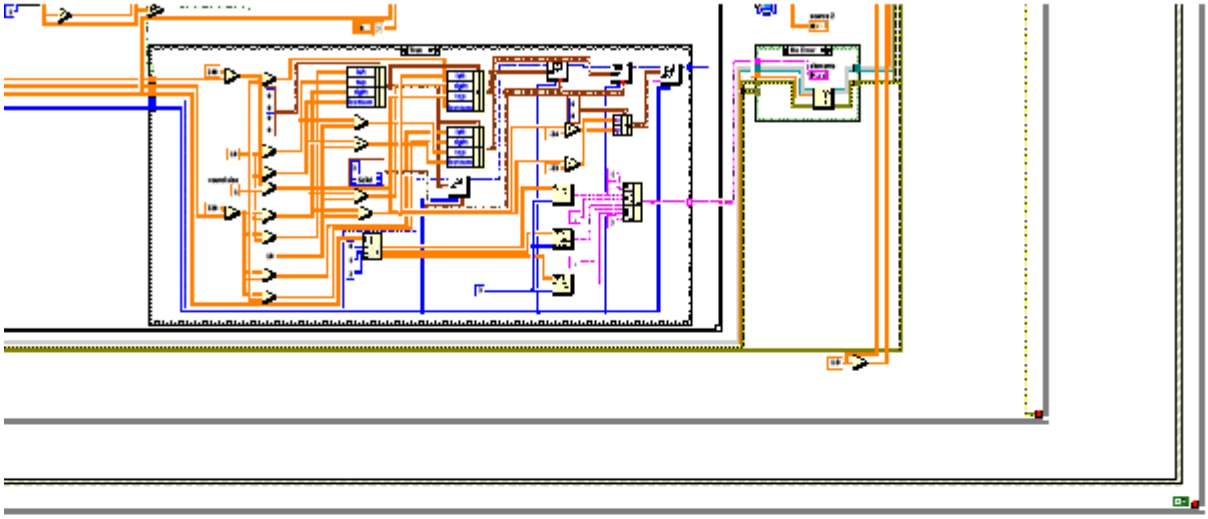
Page 4

Master/Slave Design Pattern
C:\Users\AVNC\Desktop\programs\sound sources localization Folder\sound sources localization.vi
Last modified on 11/17/2011 at 14:24
Printed on 11/17/2011 at 14:25



Express VI Configuration Information

Master/Slave Design Pattern
C:\Users\AVNC\Desktop\programs\sound sources localization Folder\sound sources localization.vi
Last modified on 11/17/2011 at 14:24
Printed on 11/17/2011 at 14:25



Master/Slave Design Pattern
C:\Users\AVNC\Desktop\programs\sound sources localization Folder\sound sources localization.vi
Last modified on 11/17/2011 at 14:24
Printed on 11/17/2011 at 14:25

Spectral Map
Spectral Map
Computes the magnitude or power spectral map of an input signal as a function of time or rotational speed. This Express VI can return the spectral map as a colormap or a waterfall graph.

This Express VI is configured as follows:

Export colormap: True
Export waterfall graph: True
Plot type: Frequency-Time
- Limit range: False
- Time interval [s]: 1m

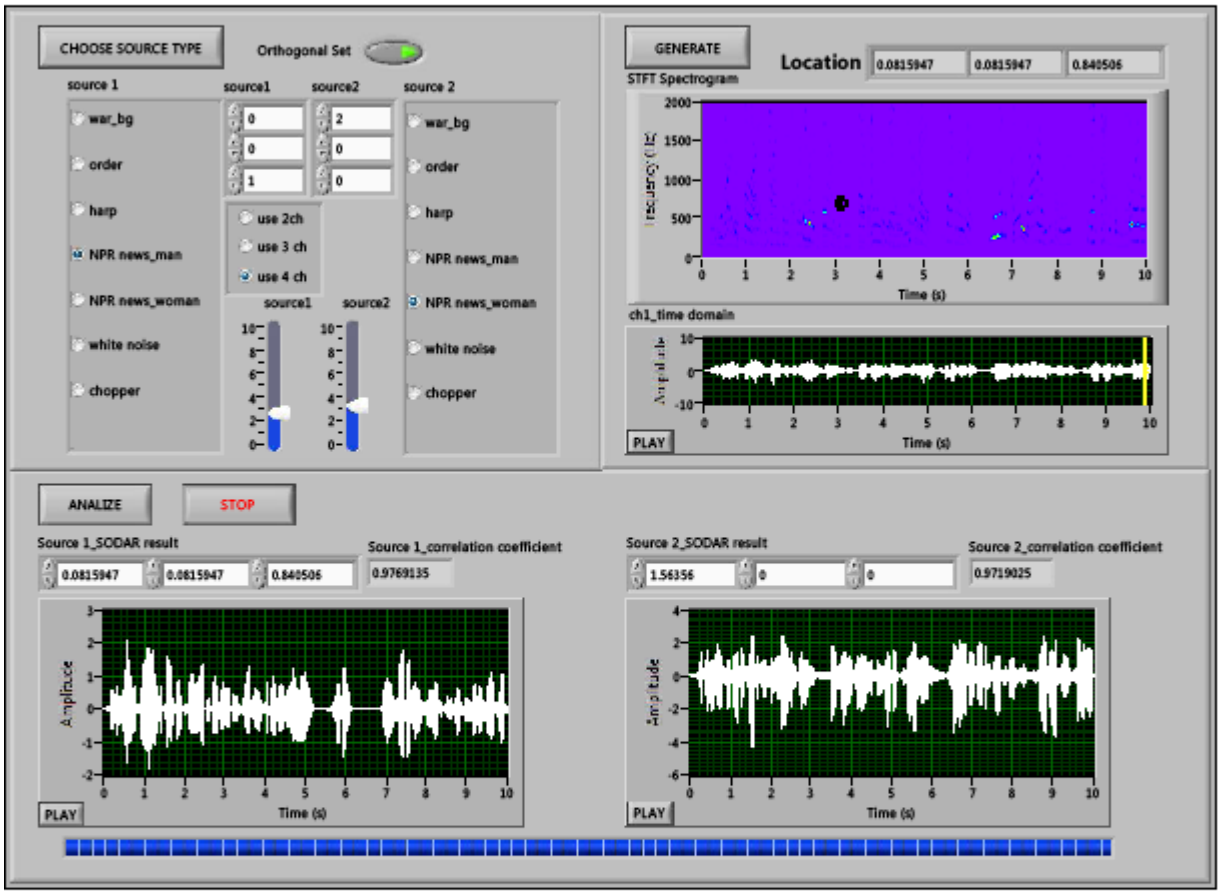
Window: None
Window length [samples]: 2048
Colormap scaling:
- Spectrum type: Power
- Peak conversion: RMS
- Colormap scale: dB

Waterfall graph scaling:
- Spectrum type: Power
- Peak conversion: RMS
- Waterfall scale: Linear

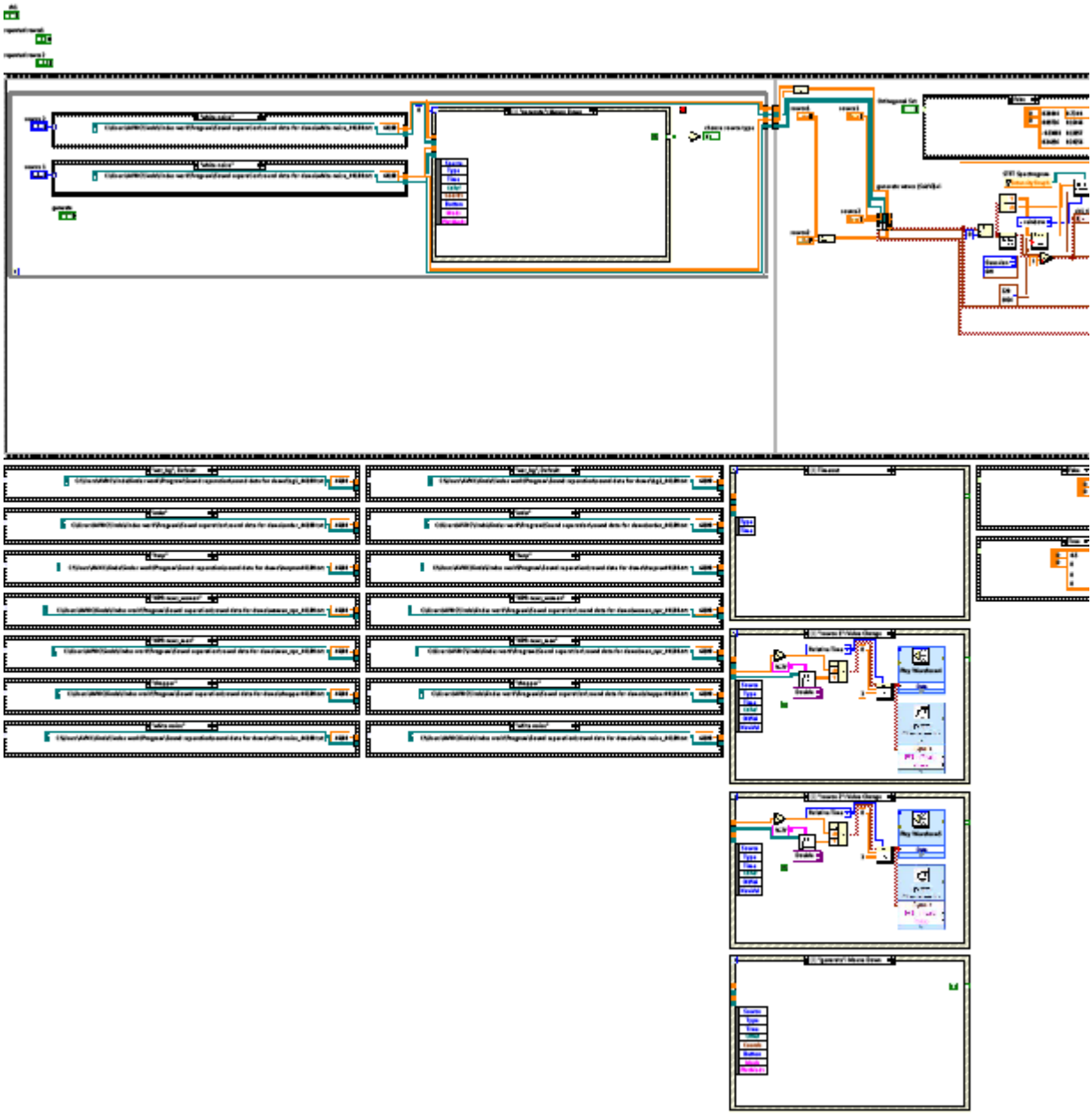
APPENDIX B – BLIND SOURCES LOCALIZATION AND SEPARATION

BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:58

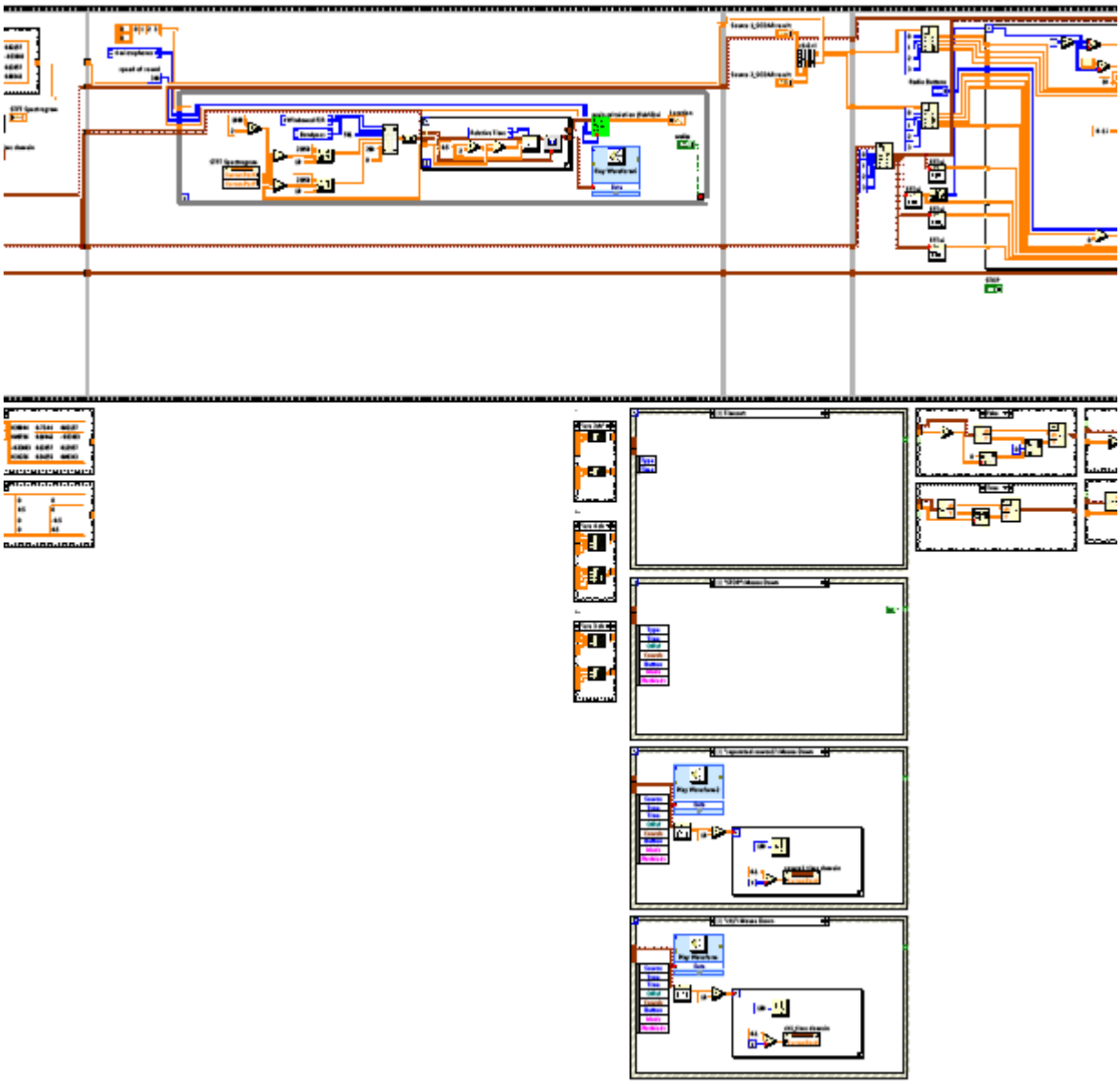
BSLS.vi



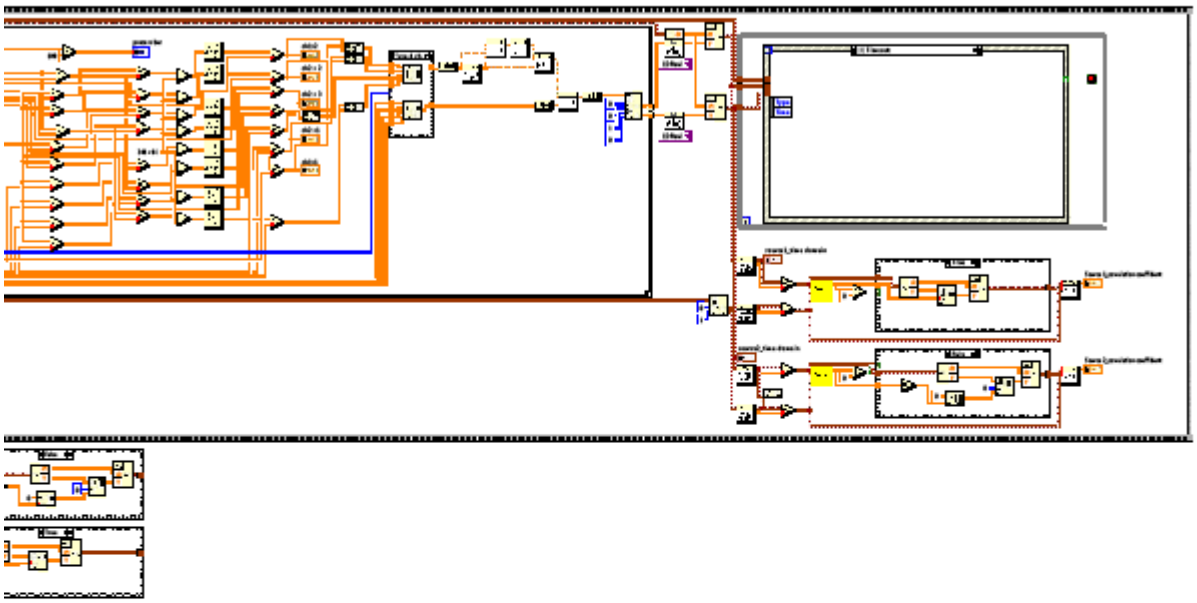
BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:58



BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:58



BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:58



BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:59

Play Waveform3
Play Waveform
Plays data from the sound output device using finite sampling. This Express VI automatically configures an output task and clears the task after the output completes.

This Express VI is configured as follows:

Device: 0

Play Waveform
Play Waveform
Plays data from the sound output device using finite sampling. This Express VI automatically configures an output task and clears the task after the output completes.

This Express VI is configured as follows:

Device: 0

Play Waveform2
Play Waveform
Plays data from the sound output device using finite sampling. This Express VI automatically configures an output task and clears the task after the output completes.

This Express VI is configured as follows:

Device: 0

WDT Waveform Duration DBL.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\WDT Waveform Duration DBL.vi

Waveform Duration.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\Waveform Duration.vi

NI_Matrix.lvlib:Matrix To Array - CM.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Matrix To Array - CM.vi

BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:59

BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:59

NI_Matrix.lvlib:Matrix To Array.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Matrix To Array.vi

NI_Matrix.lvlib:Array To Matrix - CA.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Array To Matrix - CA.vi

NI_AALBase.lvlib:Complex Inverse Matrix.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\baseanly.llb\Complex Inverse Matrix.vi

NI_AALBase.lvlib:Inverse Matrix.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\baseanly.llb\Inverse Matrix.vi

NI_AALBase.lvlib:Complex A x B.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\baseanly.llb\Complex A x B.vi

NI_AALBase.lvlib:A x B.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\baseanly.llb\A x B.vi

NI_AALPro.lvlib:Complex Conjugate Transpose Matrix.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\7linalg.llb\Complex Conjugate Transpose Matrix.vi

NI_AALPro.lvlib:Transpose Matrix.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\7linalg.llb\Transpose Matrix.vi

NI_Matrix.lvlib:Array To Matrix - CA2.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Array To Matrix - CA2.vi

NI_Matrix.lvlib:Array To Matrix.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Array To Matrix.vi

NI_Gmath.lvlib:Correlation Coefficient.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\gmath\statdist.llb\Correlation Coefficient.vi

cross_index(SubVI).vi
C:\Users\AVNC\Desktop\BSLS\cross_index(SubVI).vi

NI_AALPro.lvlib:Inverse Real FFT.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\2dsp.llb\Inverse Real FFT.vi

NI_AALPro.lvlib:Inverse FFT.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\2dsp.llb\Inverse FFT.vi

NI_AALPro.lvlib:FFT.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\2dsp.llb\FFT.vi

c1s2.vi
C:\Users\AVNC\Desktop\BSLS\c1s2.vi

NI_MAPro.lvlib:FIR Filter for 1 Chan.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\measure\macond.llb\FIR Filter for 1 Chan.vi

BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:59

NI_MAPro.lvlib:Digital FIR Filter.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\measure\macond.llb\Digital FIR Filter.vi

main calculation (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\main calculation (SubVI).vi

Play Waveform6
Play Waveform
Plays data from the sound output device using finite sampling. This Express VI automatically configures an output task and clears the task after the output completes.

This Express VI is configured as follows:

Device: 0

STFT Spectrogram Display.vi
C:\Users\AVNC\Desktop\BSLS\sig_proc.llb\STFT Spectrogram Display.vi

NI_Gmath.lvlib:STFT Spectrograms.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\gmath\trans.llb\STFT Spectrograms.vi

Waveform Min Max.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\Waveform Min Max.vi

generate waves (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\generate waves (SubVI).vi

WDT Get Waveform Subset DBL.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\WDT Get Waveform Subset DBL.vi

Get Waveform Subset.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\Get Waveform Subset.vi

Read From Spreadsheet File (DBL).vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Utility\file.llb\Read From Spreadsheet File (DBL).vi

Read From Spreadsheet File.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Utility\file.llb\Read From Spreadsheet File.vi

BSLS.vi
C:\Users\AVNC\Desktop\BSLS\BSLS.vi
Last modified on 11/17/2011 at 11:49
Printed on 11/17/2011 at 11:59

Play Waveform5
Play Waveform
Plays data from the sound output device using finite sampling. This Express VI automatically configures an output task and clears the task after the output completes.

This Express VI is configured as follows:

Device: 0

exnSpectralBlock spectral measurements peak WFM.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\express\express analysis\Spectral\Block\exnSpectralBlock spectral measurements peak WFM.vi

exnSpectralBlock spectral measurements peak poly.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\express\express analysis\Spectral\Block\exnSpectralBlock spectral measurements peak poly.vi

Play Waveform4
Play Waveform
Plays data from the sound output device using finite sampling. This Express VI automatically configures an output task and clears the task after the output completes.

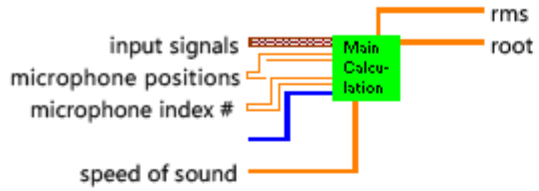
This Express VI is configured as follows:

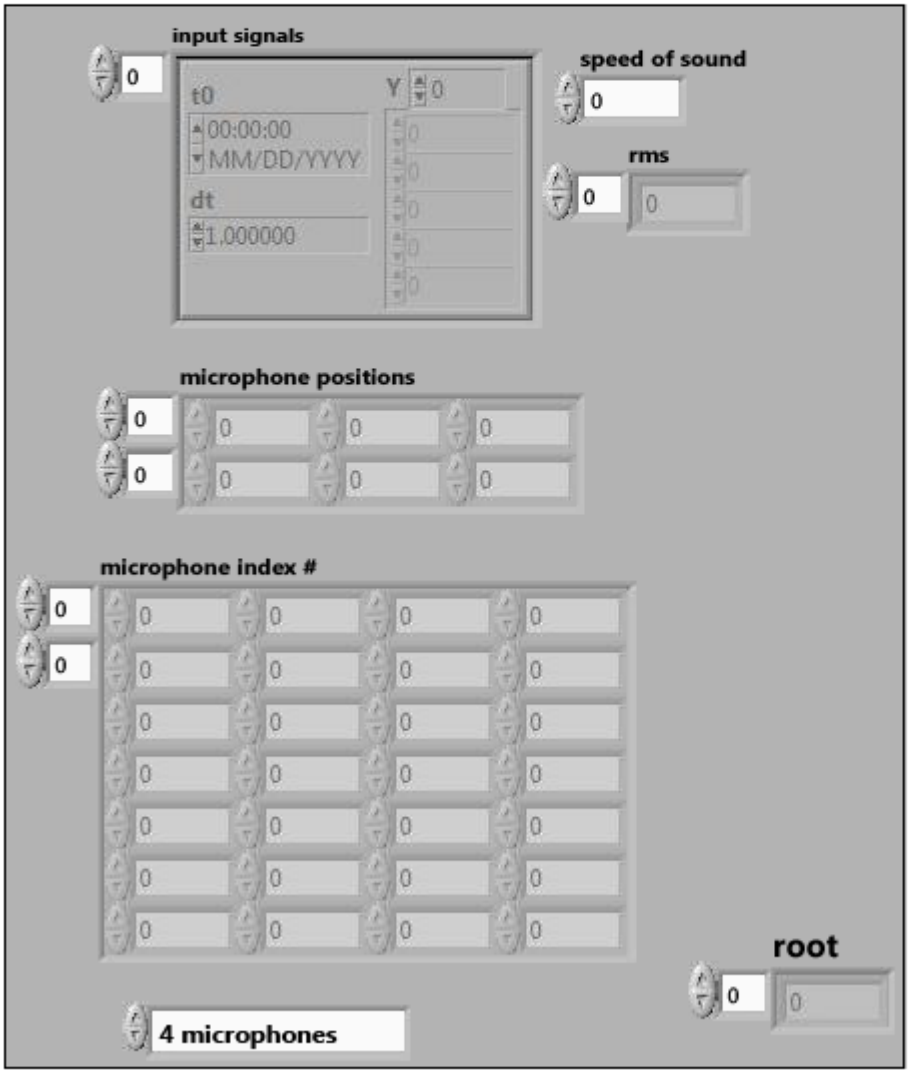
Device: 0

APPENDIX C – LOCALIZATION MAIN PROGRAM (SUB VI)

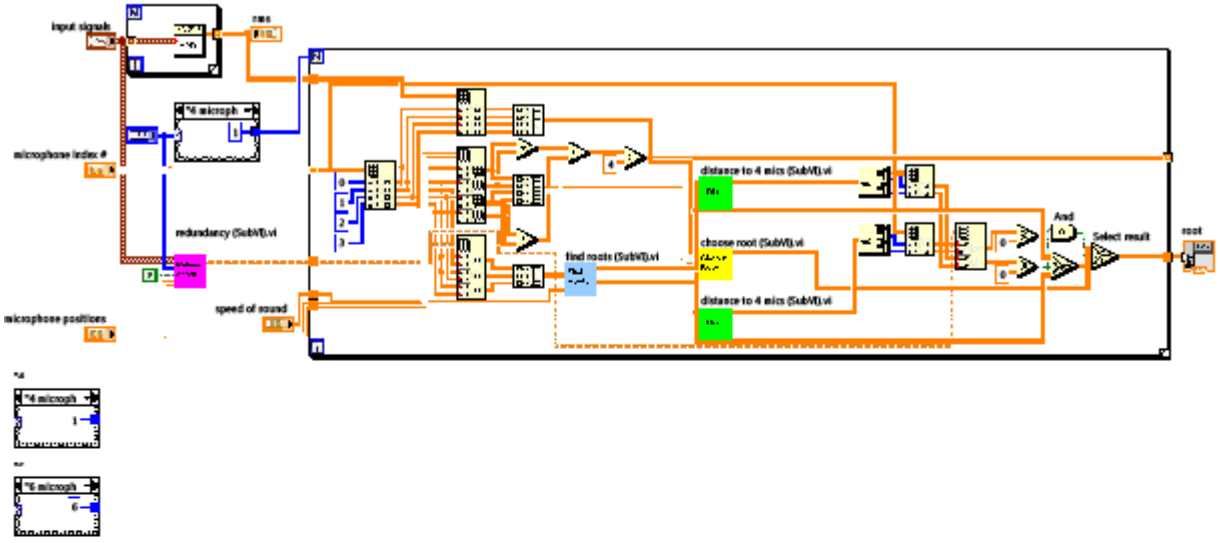
main calculation (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\main calculation (SubVI).vi
Last modified on 11/17/2011 at 11:28
Printed on 11/17/2011 at 13:41

main calculation (SubVI).vi





main calculation (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\main calculation (SubVI).vi
Last modified on 11/17/2011 at 11:28
Printed on 11/17/2011 at 13:41



Choose Root **choose root (SubVI).vi**
C:\Users\AVNC\Desktop\BSLS\choose root (SubVI).vi

Dis **distance to 4 mics (SubVI).vi**
C:\Users\AVNC\Desktop\BSLS\distance to 4 mics (SubVI).vi

Find Roots **find roots (SubVI).vi**
C:\Users\AVNC\Desktop\BSLS\find roots (SubVI).vi

RMS **NI_AALPro.lvlib:RMS (DBL).vi**
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\5stat.llb\RMS (DBL).vi

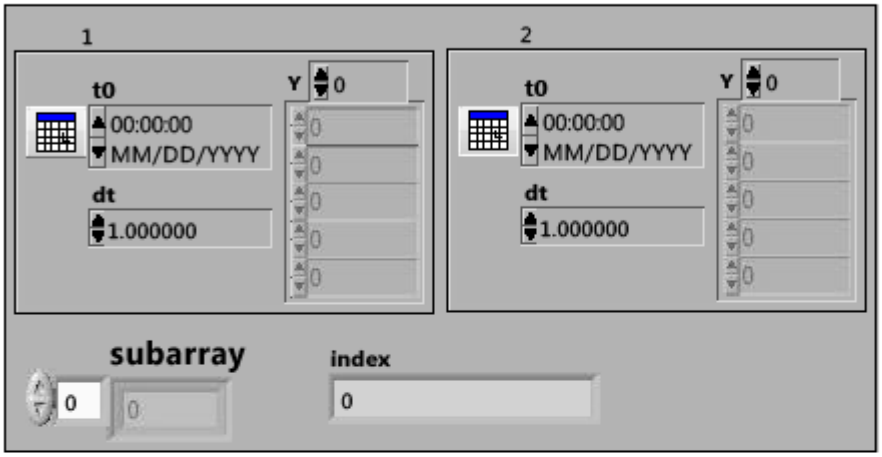
RMS **NI_AALPro.lvlib:RMS.vi**
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\5stat.llb\RMS.vi

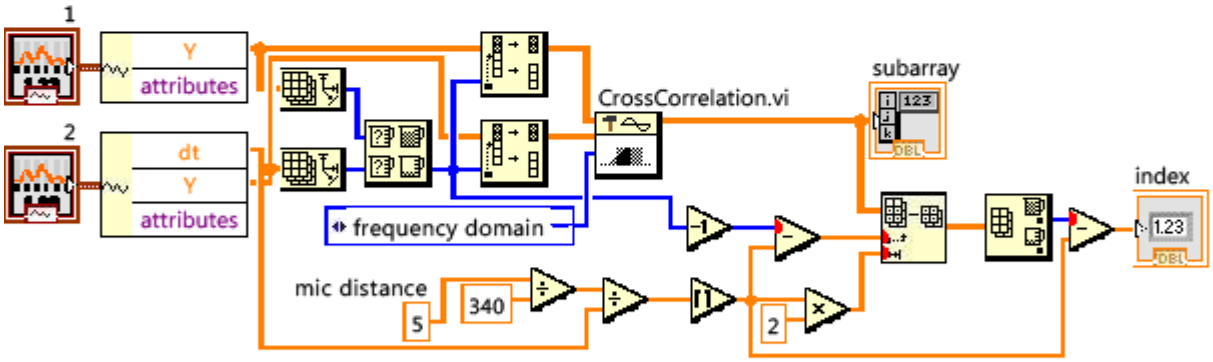
Redun-dancy **redundancy (SubVI).vi**
C:\Users\AVNC\Desktop\BSLS\redundancy (SubVI).vi

APPENDIX D – TDOA ESTIMATION (SUB VI)

cross_index(SubVI).vi
C:\Users\AVNC\Desktop\BSLS\cross_index(SubVI).vi
Last modified on 11/17/2011 at 11:26
Printed on 11/17/2011 at 13:42

cross_index(SubVI).vi





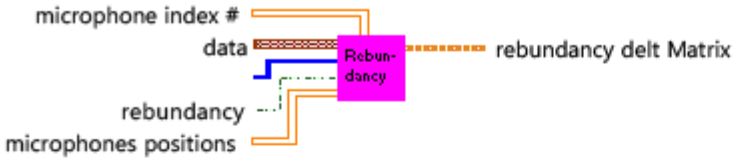
NI_AALPro.lvlib:1D Cross Correlation (DBL).vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\2dsp.llb\1D Cross Correlation (DBL).vi

NI_AALPro.lvlib:CrossCorrelation.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\2dsp.llb\CrossCorrelation.vi

APPENDIX E – REDUNDANCY CHECK ON TDOAS (SUB VI)

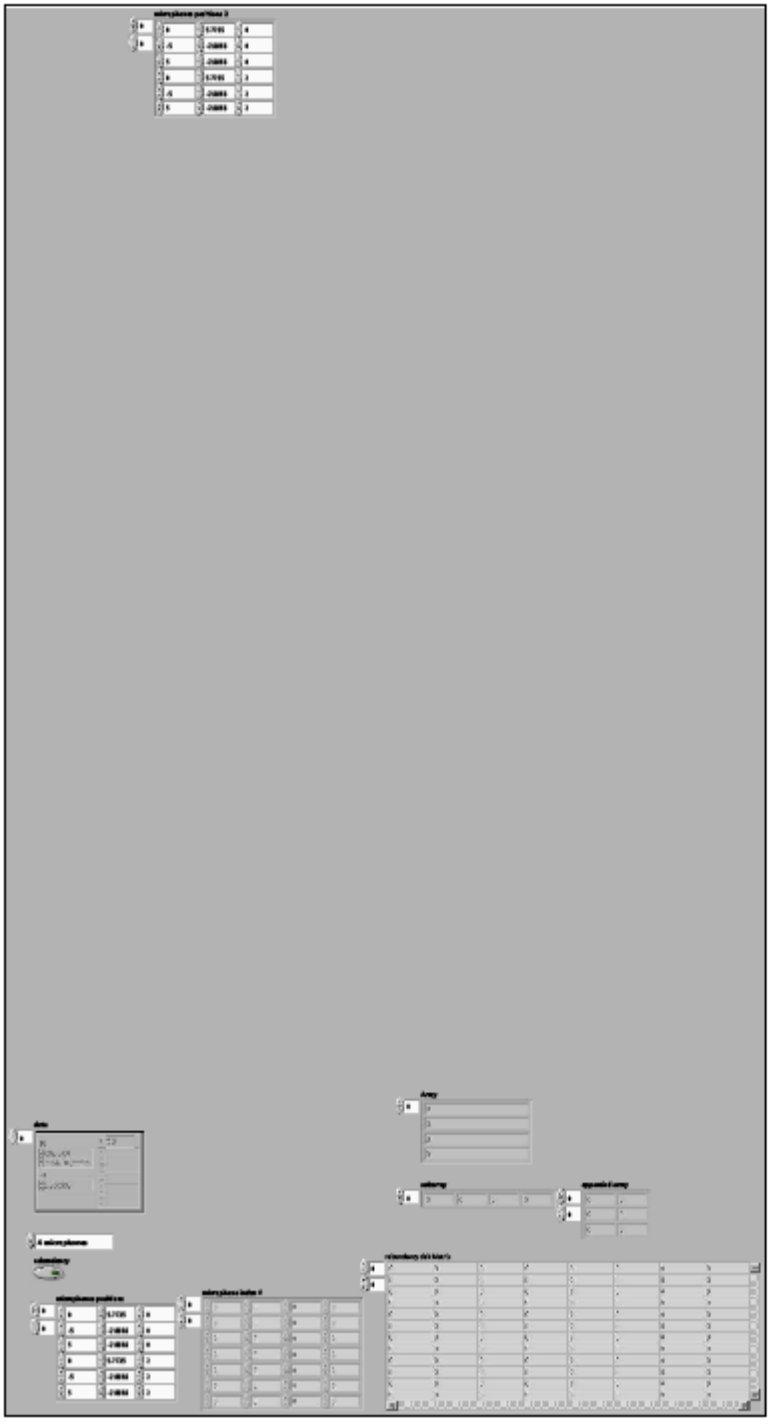
redundancy (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\redundancy (SubVI).vi
Last modified on 11/17/2011 at 11:29
Printed on 11/17/2011 at 13:42

redundancy (SubVI).vi

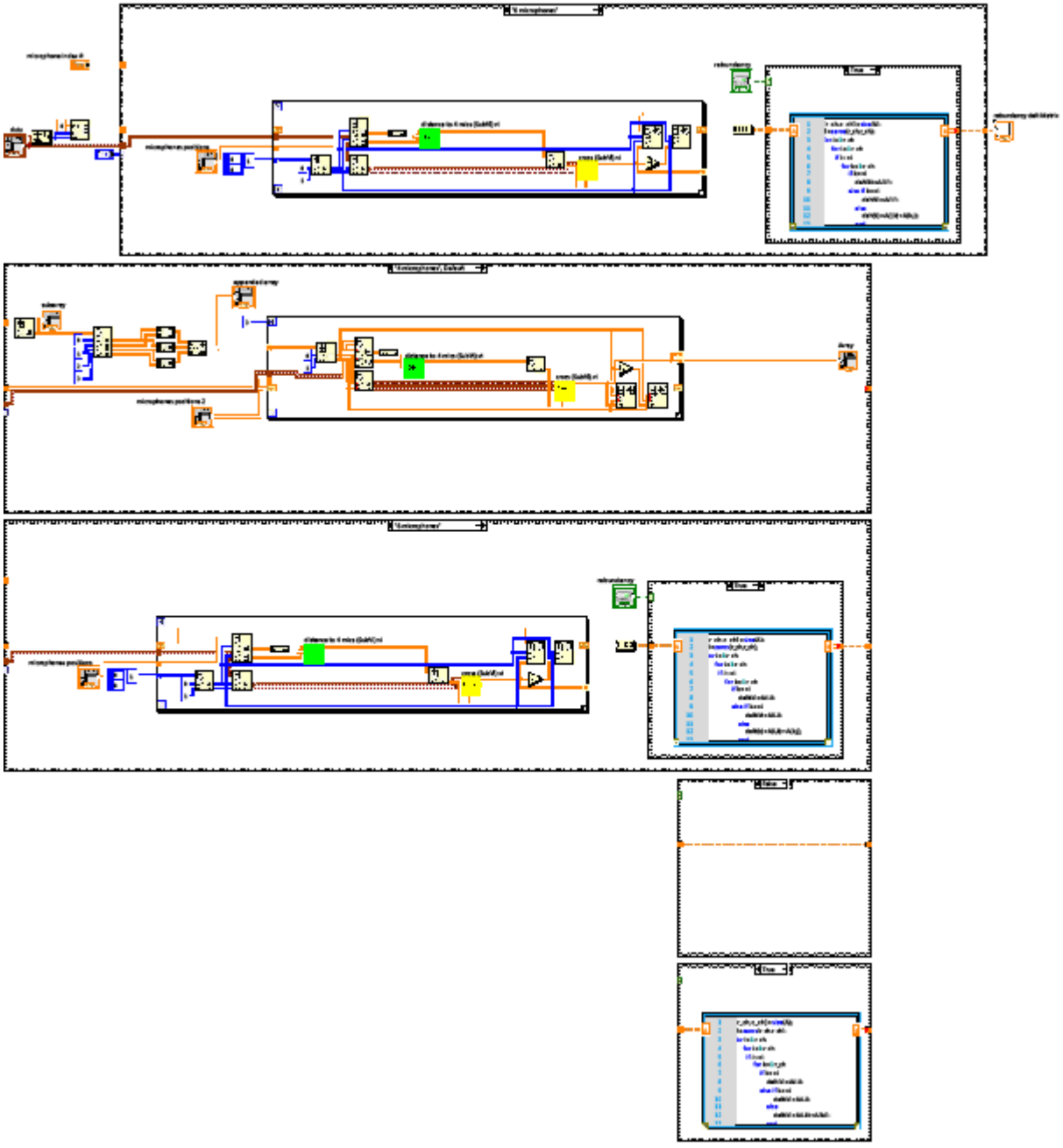


Redun-
dancy

redundancy (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\redundancy (SubVI).vi
Last modified on 11/17/2011 at 11:29
Printed on 11/17/2011 at 13:42



redundancy (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\redundancy (SubVI).vi
Last modified on 11/17/2011 at 11:29
Printed on 11/17/2011 at 13:42



NI_Matrix.lvlib:RealMatrix.ctl
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Datatypes\RealMatrix.ctl

redundancy (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\redundancy (SubVI).vi
Last modified on 11/17/2011 at 11:29
Printed on 11/17/2011 at 13:43

distance to 4 mics (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\distance to 4 mics (SubVI).vi

NI_Matrix.lvlib:Array To Matrix.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Array To Matrix.vi

NI_Matrix.lvlib:Array To Matrix - RA2.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Array To Matrix - RA2.vi

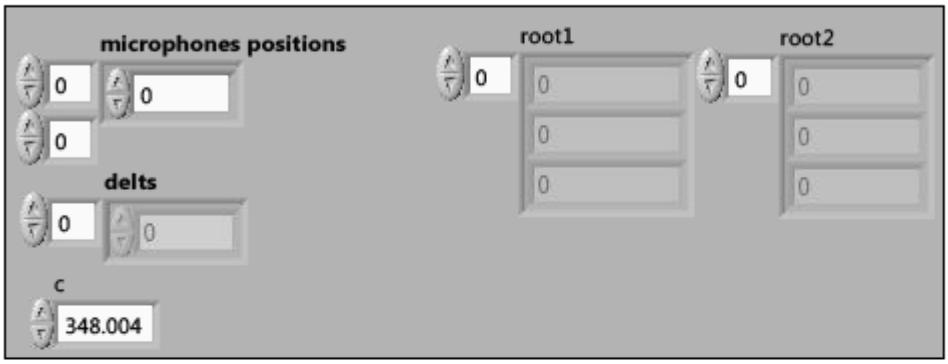
NI_Matrix.lvlib:Matrix To Array - RM.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Analysis\Matrix\Numeric\Conversions\Matrix To Array - RM.vi

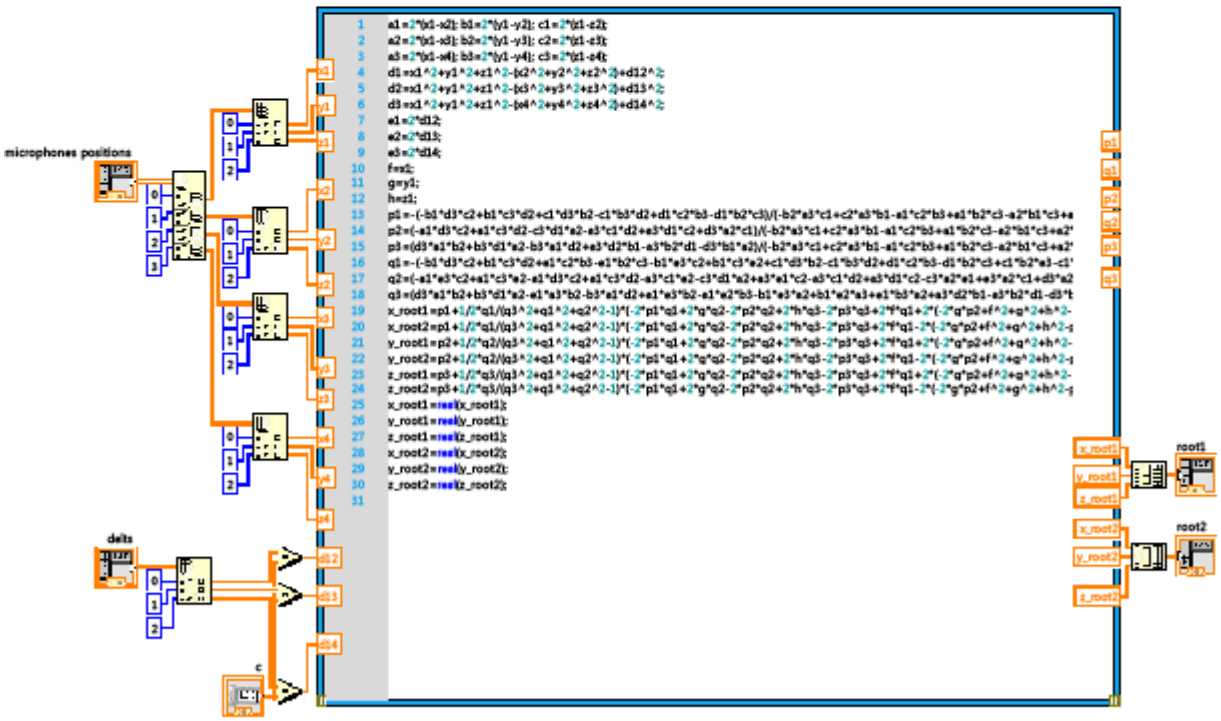
cross (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\cross (SubVI).vi

APPENDIX F – SOLVING EQUATION SET IN LOCALIZATION ALGORITHM (SUB VI)

find roots (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\find roots (SubVI).vi
Last modified on 11/17/2011 at 11:18
Printed on 11/17/2011 at 13:43

find roots (SubVI).vi



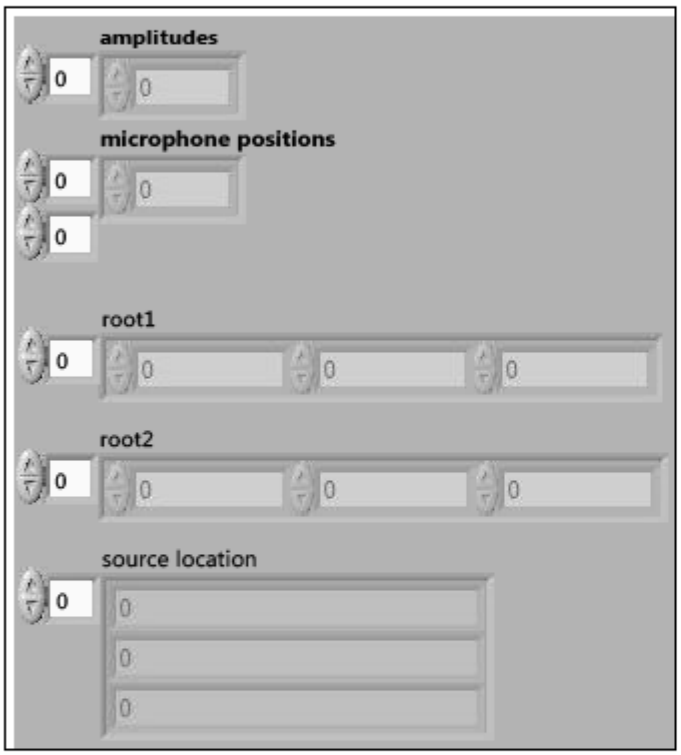


APPENDIX G – SELECTION OF THE LOCATION FROM TWO ROOTS (SUB VI)

Choose Root

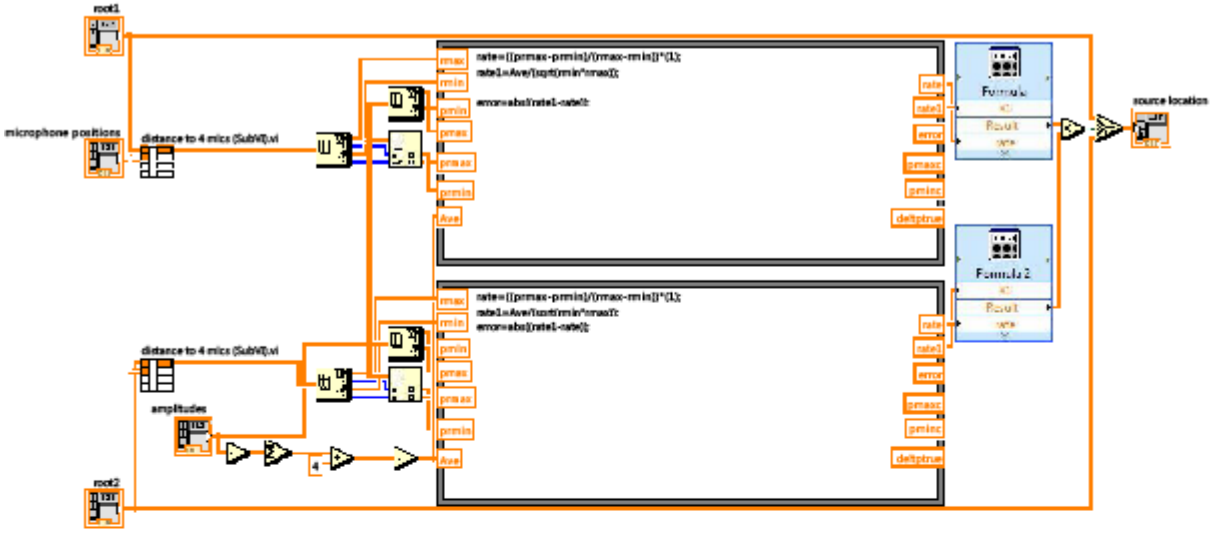
choose root (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\choose root (SubVI).vi
Last modified on 11/17/2011 at 11:24
Printed on 11/17/2011 at 13:43

choose root (SubVI).vi



Page 2 Choose Root

choose root (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\choose root (SubVI).vi
Last modified on 11/17/2011 at 11:24
Printed on 11/17/2011 at 13:43

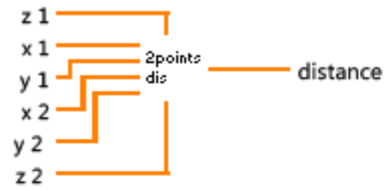


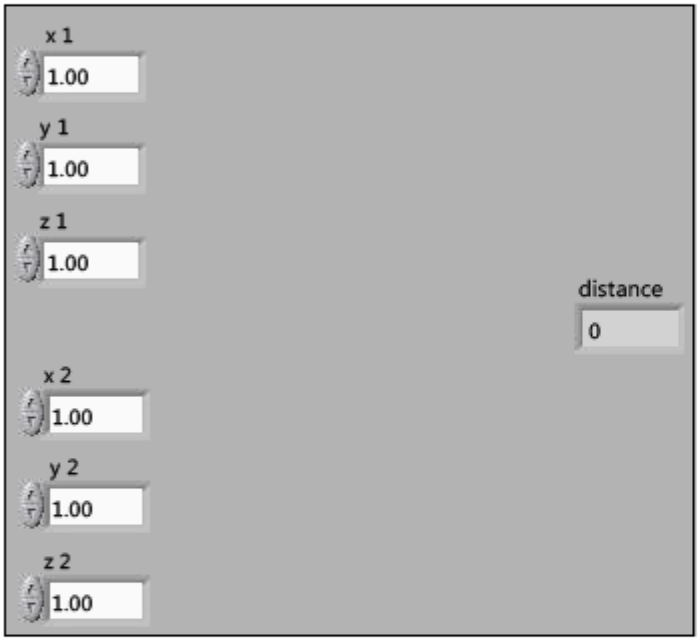
Dis **distance to 4 mics (SubVI).vi**
C:\Users\AVNC\Desktop\BSLS\distance to 4 mics (SubVI).vi

APPENDIX H – DISTANCE OF TWO POINTS IN SPACE (SUB VI)

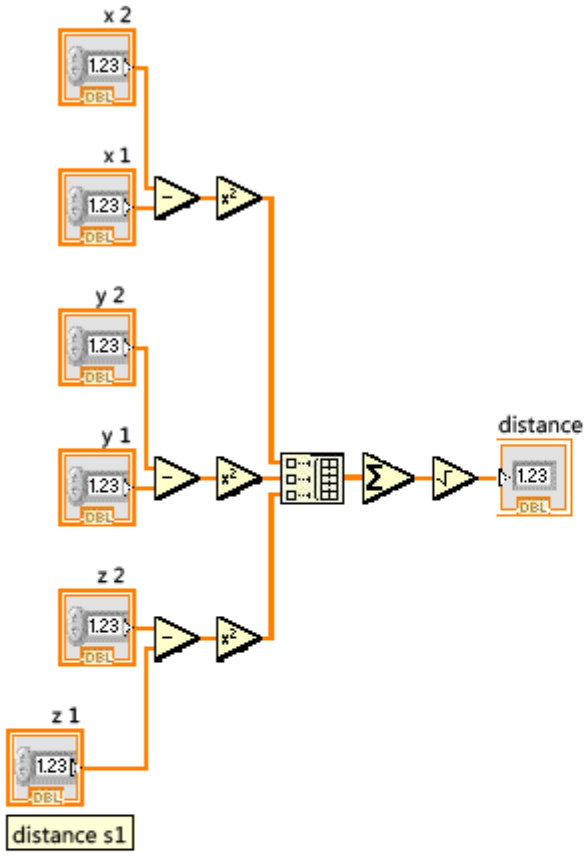
distance between 2 points (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\distance between 2 points (SubVI).vi
Last modified on 11/17/2011 at 11:26
Printed on 11/17/2011 at 13:42

distance between 2 points (SubVI).vi

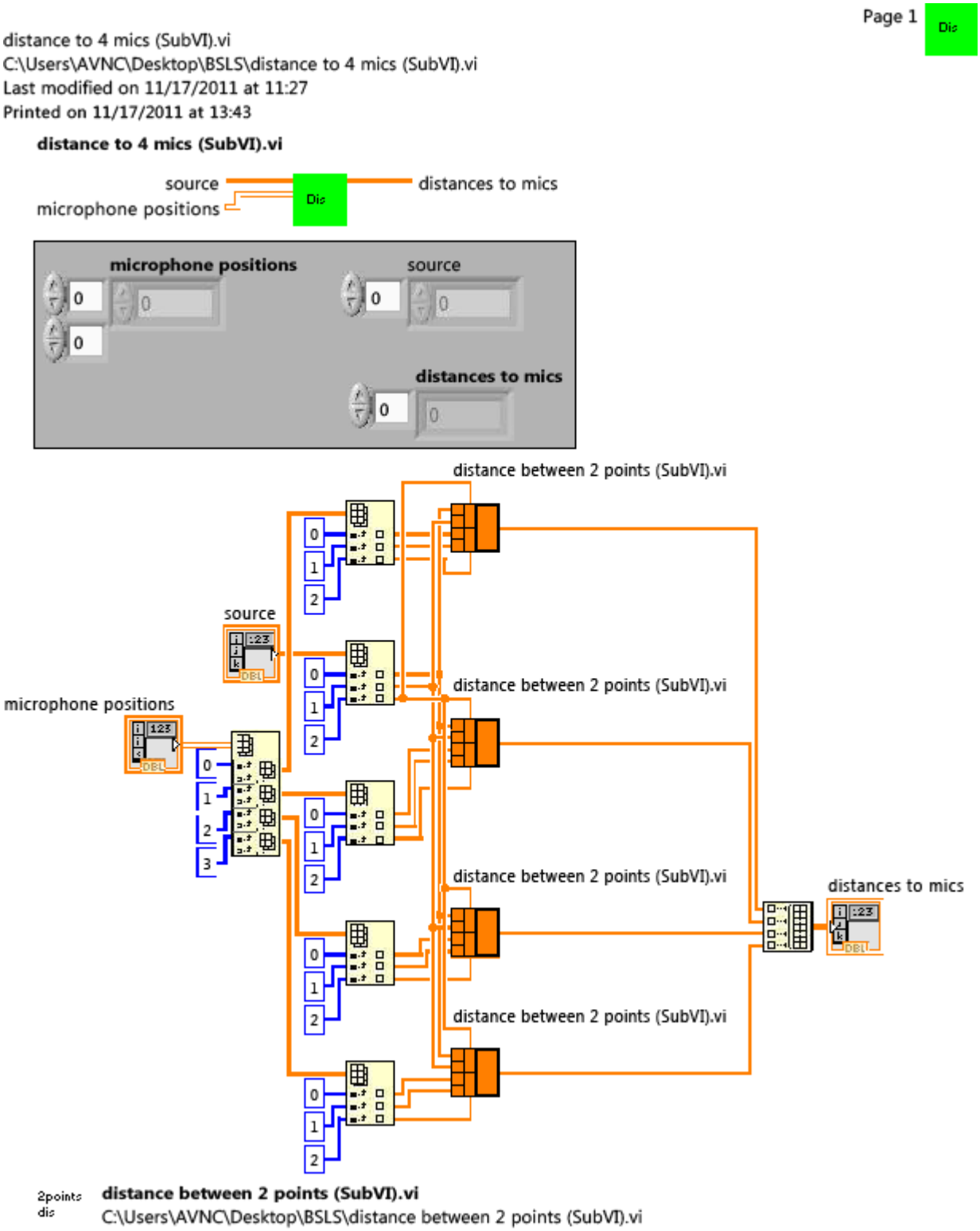




distance between 2 points (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\distance between 2 points (SubVI).vi
Last modified on 11/17/2011 at 11:26
Printed on 11/17/2011 at 13:42



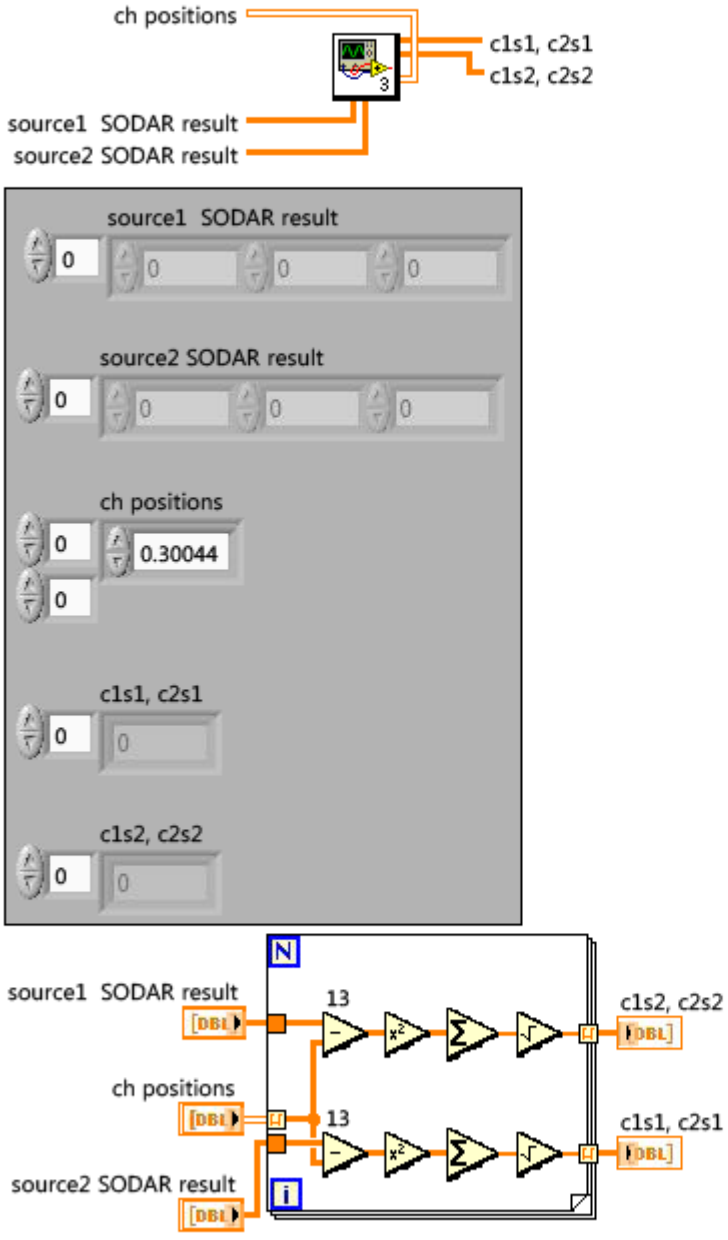
APPENDIX I – SOURCE RANGE CALCULATION (SUB VI)



APPENDIX J – SOURCE RANGES FOR MULTIPLE SOURCES (SUB VI)

c1s2.vi
C:\Users\AVNC\Desktop\BSLS\c1s2.vi
Last modified on 11/17/2011 at 11:24
Printed on 11/17/2011 at 13:44

c1s2.vi

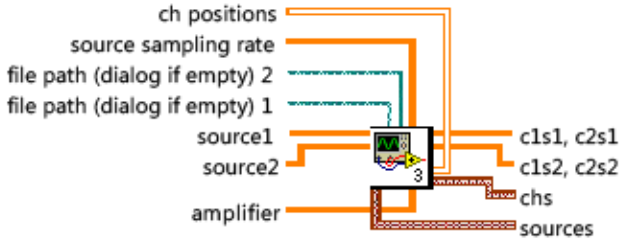


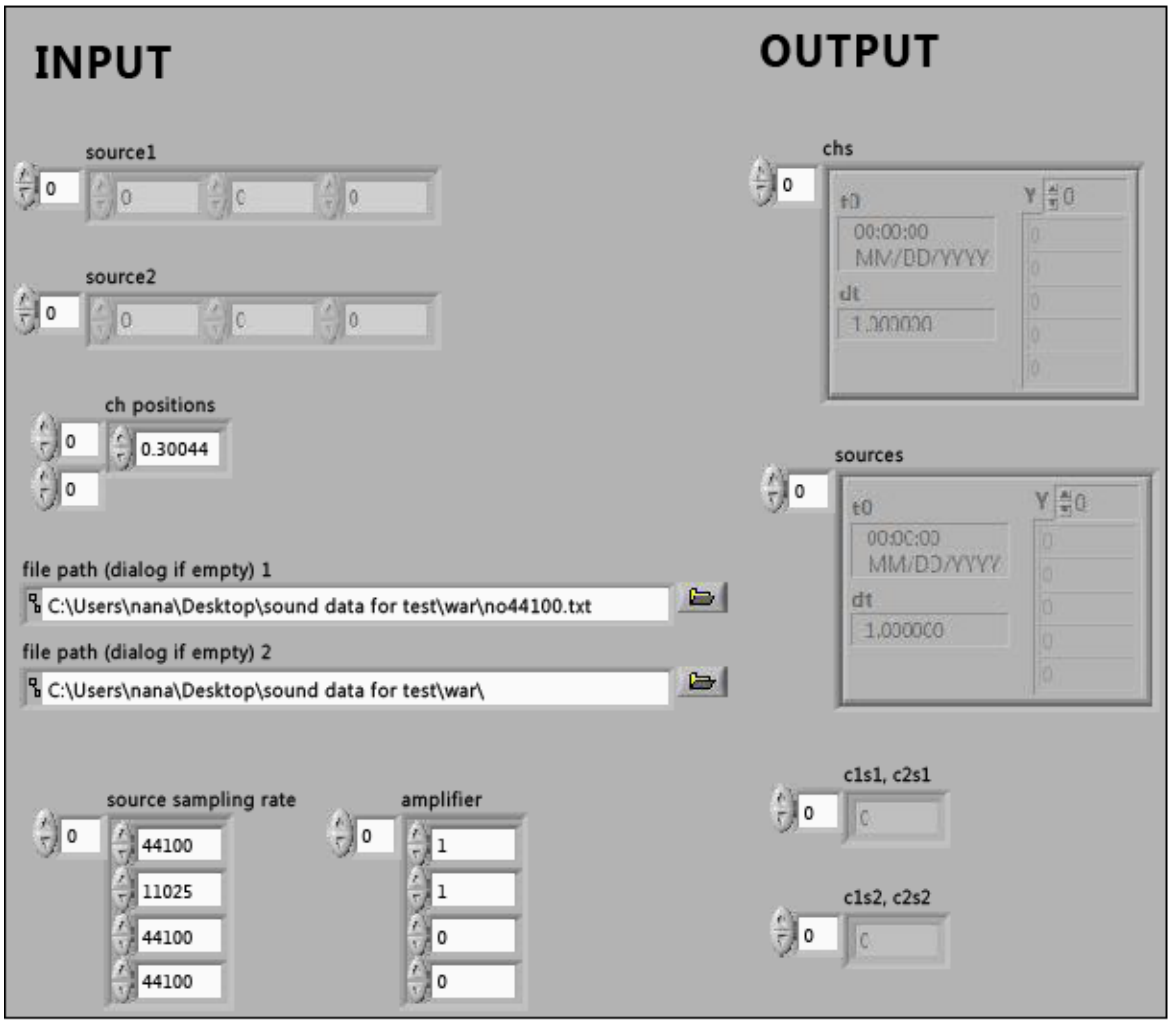
APPENDIX K – GENERATION OF MIXED SIGNALS IN NUMERICAL SIMULATION (SUB VI)

Page 1

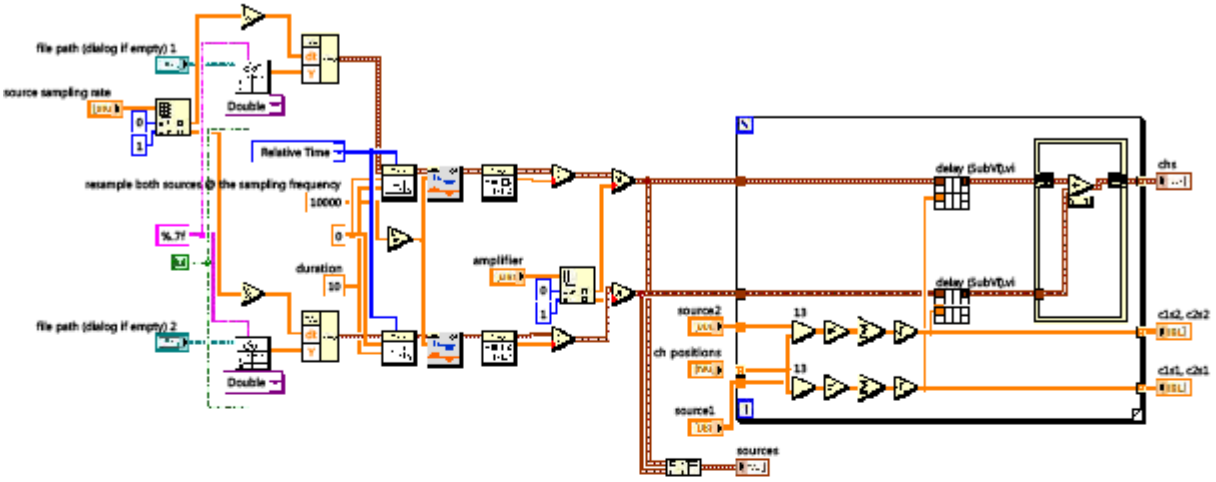
generate waves (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\generate waves (SubVI).vi
Last modified on 11/17/2011 at 11:28
Printed on 11/17/2011 at 13:44

generate waves (SubVI).vi





generate waves (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\generate waves (SubVI).vi
Last modified on 11/17/2011 at 11:28
Printed on 11/17/2011 at 13:44



Waveform Min Max.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\Waveform Min Max.vi

delay (SubVI).vi
C:\Users\AVNC\Desktop\BSLS\delay (SubVI).vi

WDT Get Waveform Subset DBL.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\WDT Get Waveform Subset DBL.vi

Get Waveform Subset.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Waveform\WDTOps.llb\Get Waveform Subset.vi

NI_MAPro.lvlib:Resample Waveform (continuous).vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\measure\maresamp.llb\Resample Waveform (continuous).vi

NI_MAPro.lvlib:Resample Waveforms (continuous).vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\measure\maresamp.llb\Resample Waveforms (continuous).vi

Read From Spreadsheet File (DBL).vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Utility\file.llb\Read From Spreadsheet File (DBL).vi

Read From Spreadsheet File.vi
C:\Program Files\National Instruments\LabVIEW 2011\vi.lib\Utility\file.llb\Read From Spreadsheet File.vi

APPENDIX L – NEURAL ACTIVITIES LOCALIZATION BY TR

```
%%%%%%%%%%%%%%%%define parameters and input data%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear
load block3
SF=3051.757813;speed=0.08;stimbreak=0.5;
pro_length=200;sensor_spacing=0.4*10^(-3);
%assume the source is within the sensor coverage area, then the min delay
%is zero, while the max delay is 65 index. using the pro_length=200 should
%be long enough to cover the peak. Also the delayed signal's length can be
%up to two times of pro_length.
ch_position=[4 1;4 2;3 1;3 2;2 2;2 1;1 2;1 1;1 4;1 3;2 4;2 3;3 3;3 4;4 3;4
4]*sensor_spacing; %sensor
x=0:sensor_spacing/10:5*sensor_spacing;y=x;[x_scan y_scan]=meshgrid(x,y); %
scan
[R,L]=size(x_scan);x_scan= reshape(x_scan,R*L,1);y_scan=
reshape(y_scan,R*L,1);
%%
peak_contour=zeros(R*L,1);count=0;
for k=1:60

%%%%%%E-stim%%%%%%%%%%%%%%%%overlapping averaging begin%%%%%%%%%%%%%%%%%%%%%%
%     for n=1:16
%         temp=zeros(pro_length,1);
%         for l=1:10
%             start_index=fix(SF*stimbreak*10*(k-1)+(l-
1)*SF*stimbreak)+1;end_index=start_index+fix(l*SF*stimbreak)-1;
%             [C,I]=max(data(start_index:end_index,n));
%             temp=temp+data(I+start_index:I+start_index+pro_length-
1,n); %Estim
%         end
%         ch(:,n)=temp;clear temp;
%         ch(:,n)=ch(:,n)/10;
%         ch(:,n)=cleartrend(ch(:,n)); %clear trendency caused by artifact
%         ch(1:15,n)=ch(1:15,n)*0;%only used in E-stim
%     end
%%%%%%%regular%%%%%%%%%%%%%%%%no overlapping averaging
end%%%%%%%%%%%%%%%%%%%%%%%%
%%
ch=zeros(pro_length,16);
start_index=fix(SF*stimbreak*10*(k-1))+1;
[C,I]=max(data(start_index:start_index+SF*10*stimbreak,:));
[hist_n,xout]=hist(I);
if (max(C<0.1e-3))%&&(max(hist_n)<8)
I=max(I(1),pro_length/4);
for n=1:16
ch(:,n)=data(start_index+I-pro_length/4:start_index+I+3*pro_length/4-1,n);
ch(:,n)=cleartrend(ch(:,n)); %clear trendency caused by artifact
end
%%%%%%%%%%%%%%%time reveral program%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
ch_reverse=flipud(ch);
peak=zeros(R*L,1);
for m=1:R*L
scan_position=[x_scan(m,1) y_scan(m,1)];
clear sum
```

```
sum=zeros(pro_length*2,1);
for n=1:16;
clear temp dis t
dis=sqrt((scan_position(1)-ch_position(n,1))^2+(scan_position(2)-
ch_position(n,2))^2);
t=dis/speed;zero_number=round(t*SF);
temp=[zeros(zero_number,1)
ch_reverse(:,n)]; %subplot(2,1,1); plot(temp);axis([0 400 -20e-3 20e-
3])%temp=temp(1:pro_length);
sum(1:size(temp,1),1)=sum(1:size(temp,1),1)+temp(1:size(temp,1),1);%subplot(2
,1,2); plot(sum);axis([0 400 -50e-3 50e-3]); pause(1)
end
peak(m)=max(sum);
end
%%
[C_max,I_max]=max(peak);
if C_max>0
peak=fix(peak/C_max)*C_max;
else
peak=peak*0;
end
x_scan_single = reshape(x_scan,R,L);y_scan_single =
reshape(y_scan,R,L);peak_single= reshape(peak,R,L);
clf;
%subplot(2,1,1);
peak_single(1:4,:)=peak_single(1:4,:)*0;
peak_single(18:21,:)=peak_single(18:21,:)*0;
peak_single(:,18:21)=peak_single(:,18:21)*0;
peak_single(:,1:4)=peak_single(:,1:4)*0;
surf(x_scan_single,y_scan_single,peak_single);caxis([0 6e-4]);axis([0 0.002 0
0.002 0 1e-2]);
hold
on;plot3(ch_position(:,1),ch_position(:,2),ch_position(:,2)*0,'MarkerSize',8,
'Marker','o','LineWidth',2,'LineStyle','none',...
'Color',[0 0 0]);hold on;
refreshdata(peak_single,'caller');drawnow;
%subplot(2,1,2); plot(hist_n); pause(1)
%refreshdata(ch,'caller');drawnow;
count=count+1;
peak_contour=peak_contour+peak;

disp(['finished=',num2str(k),'%']);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
end  %only for regular one
end
%%
%%%%%%%%%%%%%%final plot%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%time reversal results%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
x_scan = reshape(x_scan,R,L);
y_scan = reshape(y_scan,R,L);
peak_contour = reshape(peak_contour,R,L);
% peak_contour(1:10,:)=peak_contour(1:10,:)*0;
% peak_contour(41:51,:)=peak_contour(41:51,:)*0;
% peak_contour(:,41:51)=peak_contour(:,41:51)*0;
% peak_contour(:,1:10)=peak_contour(:,1:10)*0;
subplot(2,2,4);
```

```
plot(ch_position(:,1),ch_position(:,2),'MarkerSize',8,'Marker','o','LineWidth
',2,'LineStyle','none',...
'Color',[0 0 0]);
surf(x_scan,y_scan,peak_contour);shading interp;view([-10 30]);hold on;
subplot(2,2,3);plot3(ch_position(:,1),ch_position(:,2),ch_position(:,2)+100,'
MarkerSize',8,'Marker','o','LineWidth',2,'LineStyle','none',...
'Color',[0 0 0]);hold on;
surf(x_scan,y_scan,peak_contour);shading interp;view([0 90]);
hold on;
plot3(ch_position(:,1),ch_position(:,2),ch_position(:,2)*0,'MarkerSize',8,'Ma
rker','o','LineWidth',2,'LineStyle','none',...
'Color',[0 0 0]);   axis([0 0.002 0 0.002]);
%%%%%%%%%directed measured data%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
loudness=max(data);loudness_x=(ch_position(:,1))';loudness_y=(ch_position(:,2
))';
temp=loudness;surf_loudness=[temp(9) temp(11) temp(14) temp(16);temp(10)
temp(12) temp(13) temp(15);temp(7) temp(5) temp(4) temp(2);temp(8) temp(6)
temp(3) temp(1)];clear temp
temp=loudness_x;surf_loudness_x=[temp(9) temp(11) temp(14) temp(16);temp(10)
temp(12) temp(13) temp(15);temp(7) temp(5) temp(4) temp(2);temp(8) temp(6)
temp(3) temp(1)];clear temp
temp=loudness_y;surf_loudness_y=[temp(9) temp(11) temp(14) temp(16);temp(10)
temp(12) temp(13) temp(15);temp(7) temp(5) temp(4) temp(2);temp(8) temp(6)
temp(3) temp(1)];clear temp
subplot(2,2,1);surf(surf_loudness_x,surf_loudness_y,surf_loudness);shading
interp;view([0 90]);
hold on;
plot3(ch_position(:,1),ch_position(:,2),ch_position(:,2)*100,'MarkerSize',8,'
Marker','o','LineWidth',2,'LineStyle','none',...
'Color',[0 0 0]);   axis([0 0.002 0 0.002]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function output=cleartrend(input)
input=input';
t=1:1:length(input);
p=polyfit(t,input,2);
ch_trend=polyval(p,t);
output=input-ch_trend;
```

REFERENCES

[1] J. S. Snyder and C. Alain, "Toward a neurophysiological theory of auditory stream segregation," *Psychol Bull,* vol. 133, pp. 780-99, Sep 2007.

[2] S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 94, pp. 10979-10984, Sep 30 1997.

[3] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology,* vol. 37, pp. 163-178, Mar 2000.

[4] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods,* vol. 134, pp. 9-21, Mar 15 2004.

[5] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping,* vol. 6, pp. 160-188, 1998.

[6] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *Neuroimage,* vol. 9, pp. 416-429, Apr 1999.

[7] C. F. Beckmann and S. A. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Transactions on Medical Imaging,* vol. 23, pp. 137-152, Feb 2004.

[8] R. O. Schmidt, "New approach to geometry of range difference location," *IEEE Transactions on Aerospace and Electronic Systems,* vol. Aes8, pp. 821-835, 1972.

[9] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Transactions on Information Theory,* vol. 19, pp. 608-614, 1973.

[10] A. Tobias, "Acoustic emission source location in two dimensions by an array of three sensors," *Non-Destructive Testing,* vol. 9, pp. 9-12, 1976.

[11] G. C. Carter, "Passive ranging errors due to receiving hydrophone position uncertainty," *Journal of the Acoustical Society of America,* vol. 65, pp. 528-530, 1979.

[12] E. J. Hilliard and R. F. Pinkos, "Analysis of triangulation ranging using beta-density angular errors," *Journal of the Acoustical Society of America,* vol. 65, pp. 1218-1228, 1979.

[13] S. S. Reddi, "Multiple source location a digital approach," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 15, pp. 95-105, 1979.

[14] C. W. Clark, "A teal-time direction finding device for determining the bearing to the underwater sounds of southern right whales, Eubalaena-Australis," *Journal of the Acoustical Society of America,* vol. 68, pp. 508-511, 1980.

[15] J. C. Hassab, B. W. Guimond, and S. C. Nardone, "Estimation of location and motion parameters of a moving source observed from a linear-array," *Journal of the Acoustical Society of America,* vol. 70, pp. 1054-1061, 1981.

[16] E. J. Hilliard and J. J. Perruzzi, "Time-delay triangulation ranging using beta-density measurement error statistics," *Journal of the Acoustical Society of America,* vol. 72, pp. 1831-1837, 1982.

[17] D. J. Torrieri, "Statistical-theory of passive location systems," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 20, pp. 183-198, 1984.

[18] R. O. Schmidt, "Multiple emitter location and signal parameter-estimation," *IEEE Transactions on Antennas and Propagation,* vol. 34, pp. 276-280, Mar 1986.

[19] J. M. Delosme, M. Morf, and B. Friedlander, "Source location from time differences of arrival: Identifiability and estimation," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. ASSP-35, pp. 818 – 824 1987.

[20] H. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 35, pp. 1223-1225, Aug 1987.

[21] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 35, pp. 1661-1669, Dec 1987.

[22] P. C. Ching and Y. T. Chan, "Adaptive time-delay estimation with constraints," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 36, pp. 599-601, Apr 1988.

[23] J. L. Spiesberger and K. M. Fristrup, "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," *American Naturalist,* vol. 135, pp. 107-153, Jan 1990.

[24] M. R. Gorman, "Plate wave acoustic-emission," *Journal of the Acoustical Society of America,* vol. 90, pp. 358-364, Jul 1991.

[25] S. M. Ziola and M. R. Gorman, "Source location in thin plates using cross-correlation," *Journal of the Acoustical Society of America,* vol. 90, pp. 2551-2556, Nov 1991.

[26] J. F. Figueroa and J. S. Lamancusa, "A method for accurate detection of time of arrival - analysis and design of an ultrasonic ranging system," *Journal of the Acoustical Society of America,* vol. 91, pp. 486-494, Jan 1992.

[27] F. Figueroa and A. Mahajan, "A robust method to determine the coordinates of a wave source for 3-D position sensing," *Journal of Dynamic Systems Measurement and Control-Transactions of the Asme,* vol. 116, pp. 505-511, Sep 1994.

[28] H. A. Canistraro and E. H. Jordan, "Projectile-impact-location determination: An acoustic triangulation method," *Measurement Science & Technology,* vol. 7, pp. 1755-1760, Dec 1996.

[29] P. K. McGregor, T. Dabelsteen, C. W. Clark, J. L. Bower, J. P. Tavares, and J. Holland, "Accuracy of a passive acoustic location system: empirical studies in terrestrial habitats," *Ethology Ecology & Evolution,* vol. 9, pp. 269-286, Sep 1997.

[30] B. G. Ferguson, "Variability in the passive ranging of acoustic sources in air using a wavefront curvature technique," *Journal of the Acoustical Society of America,* vol. 108, pp. 1535-1544, Oct 2000.

[31] B. G. Ferguson, L. G. Criswick, and K. W. Lo, "Locating far-field impulsive sound sources in air by triangulation," *Journal of the Acoustical Society of America,* vol. 111, pp. 104-16, Jan 2002.

[32] J. D. Chen, J. Benesty, and Y. A. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Transactions on Speech and Audio Processing,* vol. 11, pp. 549-557, Nov 2003.

[33] P. T. Coverley and W. J. Staszewski, "Impact damage location in composite structures using optimized sensor triangulation procedure," *Smart Materials & Structures,* vol. 12, pp. 795-803, Oct 2003.

[34] T. Kundu, S. Das, and K. V. Jata, "Point of impact prediction in isotropic and anisotropic plates from the acoustic emission data," *Journal of the Acoustical Society of America,* vol. 122, pp. 2057-66, Oct 2007.

[35] J. H. Wilson, "Signal-detection and localization using the Fourier-series method (Fsm) and cross-sensor data," *Journal of the Acoustical Society of America,* vol. 73, pp. 1648-1656, 1983.

[36] A. H. Nuttall and J. H. Wilson, "Estimation of the acoustic field directionality by use of planar and volumetric arrays via the Fourier-series method and the Fourier integral method," *Journal of the Acoustical Society of America,* vol. 90, pp. 2004-2019, Oct 1991.

[37] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *Journal of the Acoustical Society of America,* vol. 91, pp. 1662-76, Mar 1992.

[38] W. Soede, A. J. Berkhout, and F. A. Bilsen, "Development of a directional hearing instrument based on array technology," *Journal of the Acoustical Society of America,* vol. 94, pp. 785-98, Aug 1993.

[39] M. Kompis and N. Dillier, "Noise-reduction for hearing-aids - combining directional microphones with an adaptive beamformer," *Journal of the Acoustical Society of America,* vol. 96, pp. 1910-1913, Sep 1994.

[40] J. P. Fabre and J. H. Wilson, "Minimum detectable level evaluation of inverse beamforming using outpost SUNRISE data," *Journal of the Acoustical Society of America,* vol. 98, pp. 3262-3278, Dec 1995.

[41] R. J. van Hoesel and G. M. Clark, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *Journal of the Acoustical Society of America,* vol. 97, pp. 2498-503, Apr 1995.

[42] J. H. Wilson, "Applications of inverse beamforming theory," *Journal of the Acoustical Society of America,* vol. 98, pp. 3250-3261, Dec 1995.

[43] J. Vanden Berghe and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *Journal of the Acoustical Society of America,* vol. 103, pp. 3621-3626, Jun 1998.

[44] A. H. Nuttall and J. H. Wilson, "Adaptive beamforming at very low frequencies in spatially coherent, cluttered noise environments with low signal-to-noise ratio and finite-averaging times," *Journal of the Acoustical Society of America,* vol. 108, pp. 2256-65, Nov 2000.

[45] M. Kompis and N. Dillier, "Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. II. Experimental verification of the predictions," *Journal of the Acoustical Society of America,* vol. 109, pp. 1134-43, Mar 2001.

[46] M. Kompis and N. Dillier, "Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement," *Journal of the Acoustical Society of America,* vol. 109, pp. 1123-33, Mar 2001.

[47] J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects," *Journal of the Acoustical Society of America,* vol. 109, pp. 185-193, Jan 2001.

[48] A. Crewe, F. Perrin, V. Benoit, and K. Haddad, "Real-time pass-by noise source identification using a beam-forming approach," *SAE Paper 03NVC-25,* 2002.

[49] M. V. Greening and J. E. Perkins, "Adaptive beamforming for nonstationary arrays," *Journal of the Acoustical Society of America,* vol. 112, pp. 2872-2881, Dec 2002.

[50] K. Haddad and V. Benoit, "Capabilities of a beamforming technique for acoustic measurements inside a moving car," *The 2002 International Congress and Exposition on Noise Control Engineering,* 2002.

[51] K. Haddad and V. Benoit, "Localization of aeroacoustic sources all along a vehicle by means of the acoustical imaging system ANT64 in the semi-anechoic wind-tunnel Pininfarina," *Euronoise,* 2003.

[52] K. Haddad and V. Benoit, "An acoustical imaging processing for the localization of acoustic sources inside a vehicle: method and results for tests on road," *SAE International 03NVC-384,* 2003.

[53] G. L. D'Spain, E. Terrill, C. D. Chadwell, J. A. Smith, and S. D. Lynch, "Active control of passive acoustic fields: Passive synthetic aperture/Doppler beamforming with data from an autonomous vehicle," *Journal of the Acoustical Society of America,* vol. 120, pp. 3635-3654, Dec 2006.

[54] M. E. Lockwood and D. L. Jones, "Beamformer performance with acoustic vector sensors in air," *Journal of the Acoustical Society of America,* vol. 119, pp. 608-619, Jan 2006.

[55] L. C. Parra, "Steerable frequency-invariant beamforming for arbitrary arrays," *Journal of the Acoustical Society of America,* vol. 119, pp. 3839-3847, Jun 2006.

[56] A. Aubry, A. Derode, P. Roux, and A. Tourin, "Coherent backscattering and far-field beamforming in acoustics," *Journal of the Acoustical Society of America,* vol. 121, pp. 70-77, Jan 2007.

[57] K. S. Kim, J. Liu, and M. F. Insana, "Beamforming using spatial matched filtering with annular arrays," *Journal of the Acoustical Society of America,* vol. 121, pp. 1852-1855, Apr 2007.

[58] C. A. Shera and J. J. Guinan, "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive calibration of intracochlear distortion sources," *Journal of the Acoustical Society of America,* vol. 121, pp. 1003-1016, Feb 2007.

[59] S. F. Yan, C. H. Hou, X. C. Ma, and Y. L. Ma, "Convex optimization based time-domain broadband beamforming with sidelobe control (L)," *Journal of the Acoustical Society of America,* vol. 121, pp. 46-49, Jan 2007.

[60] A. Cigada, M. Lurati, F. Ripamonti, and M. Vanali, "Moving microphone arrays to reduce spatial aliasing in the beamforming technique: theoretical background and numerical investigation," *Journal of the Acoustical Society of America,* vol. 124, pp. 3648-3658, Dec 2008.

[61] D. Döbler, "Time-domain beamforming using zeropadding," *Berlin Beamforming Conference (BeBeC),* 2008.

[62] R. P. Dougherty, "What is beamforming?," *Berlin Beamforming Conference (BeBeC),* 2008.

[63] T. Folegot, G. Martinelli, P. Guerrini, and J. M. Stevenson, "An active acoustic tripwire for simultaneous detection and localization of multiple underwater intruders," *Journal of the Acoustical Society of America,* vol. 124, pp. 2852-2860, Nov 2008.

[64] S. Gu´erin and C. Weckm¨uller, "Frequency-domain reconstruction of the point-spread function for moving sources," *Berlin Beamforming Conference (BeBeC),* 2008.

[65] Y. Liu, A. R. Quayle, A. P. Dowling, and P. Sijtsma, "Beamforming correction for dipole measurement using two-dimensional microphone arrays," *Journal of the Acoustical Society of America,* vol. 124, pp. 182-191, Jul 2008.

[66] M. A. Poletti and U. P. Svensson, "Beamforming synthesis of binaural responses from computer simulations of acoustic spaces," *Journal of the Acoustical Society of America,* vol. 124, pp. 301-315, Jul 2008.

[67] W. Song, W. Ellermeier, and J. Hald, "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *Journal of the Acoustical Society of America,* vol. 123, pp. 910-924, Feb 2008.

[68] H. L. Tao and J. L. Krolik, "Waveguide invariant focusing for broadband beamforming in an oceanic waveguide," *Journal of the Acoustical Society of America,* vol. 123, pp. 1338-1346, Mar 2008.

[69] T. Yardibi, J. Li, P. Stoica, and L. N. Cattafesta, "Sparsity constrained deconvolution approaches for acoustic source mapping," *Journal of the Acoustical Society of America,* vol. 123, pp. 2631-2642, May 2008.

[70] D. Cassereau and M. Fink, "Time-reversal of ultrasonic fields .3. Theory of the closed time-reversal cavity," *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control,* vol. 39, pp. 579-592, Sep 1992.

[71] M. Fink, "Time-reversal of ultrasonic fields .1. Basic principles," *IEEE transactions on ultrasonics ferroelectrics and frequency control,* vol. 39, pp. 555-566, Sep 1992.

[72] F. Wu, J. L. Thomas, and M. Fink, "Time-reversal of ultrasonic fields .2. Experimental results," *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control,* vol. 39, pp. 567-578, Sep 1992.

[73] M. Fink, "Time-reversal mirrors," *Journal of Physics D-Applied Physics,* vol. 26, pp. 1333-1350, Sep 14 1993.

[74] M. Fink, "Time-reversal of acoustic-waves," *Recherche,* vol. 25, pp. 392-400, Apr 1994.

[75] M. Fink, "Time reversal in acoustics," *Contemporary Physics,* vol. 37, pp. 95-109, Mar-Apr 1996.

[76] C. Draeger, D. Cassereau, and M. Fink, "Theory of the time-reversal process in solids," *Journal of the Acoustical Society of America,* vol. 102, pp. 1289-1295, Sep 1997.

[77] M. Fink, "Time reversed acoustics," *Physics Today,* vol. 50, pp. 34-40, Mar 1997.

[78] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," *Journal of the Acoustical Society of America,* vol. 103, pp. 25-40, Jan 1998.

[79] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," *Journal of the Acoustical Society of America,* vol. 103, pp. 3234-3240, Jun 1998.

[80] J. H. Rose, M. Bilgen, P. Roux, and M. Fink, "Time-reversal mirrors and rough surfaces: Theory," *Journal of the Acoustical Society of America,* vol. 106, pp. 716-723, Aug 1999.

[81] P. Roux, J. De Rosny, M. Fink, and J. H. Rose, "Time-reversal mirrors and rough surfaces: Experiment," *Journal of the Acoustical Society of America,* vol. 106, pp. 724-732, Aug 1999.

[82] P. Roux, A. Derode, A. Peyre, A. Tourin, and M. Fink, "Acoustical imaging through a multiple scattering medium using a time-reversal mirror," *Journal of the Acoustical Society of America,* vol. 107, pp. L7-L12, Feb 2000.

[83] M. Tanter, J. L. Thomas, F. Coulouvrat, and M. Fink, "Breaking of time reversal invariance in nonlinear acoustics," *Physical Review E,* vol. 64, Jul 2001.

[84] M. Fink, "Acoustic time-reversal mirrors," *Imaging of Complex Media with Acoustic and Seismic Waves,* vol. 84, pp. 17-42, 2002.

[85] M. Fink, G. Montaldo, and M. Tanter, "Time-reversal acoustics in biomedical engineering," *Annual Review of Biomedical Engineering,* vol. 5, pp. 465-497, 2003.

[86] E. Kerbrat, C. Prada, D. Cassereau, and M. Fink, "Imaging in the presence of grain noise using the decomposition of the time reversal operator," *Journal of the Acoustical Society of America,* vol. 113, pp. 1230-1240, Mar 2003.

[87] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," *Journal of the Acoustical Society of America,* vol. 113, pp. 1533-1543, Mar 2003.

[88] G. Lerosey, J. de Rosny, A. Tourin, A. Derode, G. Montaldo, and M. Fink, "Time reversal of electromagnetic waves and telecommunication," *Radio Science,* vol. 40, Sep 17 2005.

[89] G. Ribay, J. de Rosny, and M. Fink, "Time reversal of noise sources in a reverberation room," *Journal of the Acoustical Society of America,* vol. 117, pp. 2866-2872, May 2005.

[90] M. Fink, "Time-reversal acoustics in complex environments," *Geophysics,* vol. 71, pp. Si151-Si164, Jul-Aug 2006.

[91] G. Lerosey, J. de Rosny, A. Tourin, A. Derode, and M. Fink, "Time reversal of wideband microwaves," *Applied Physics Letters,* vol. 88, Apr 10 2006.

[92] A. Tourin, F. Van Der Biest, and M. Fink, "Time reversal of ultrasound through a phononic crystal," *Physical Review Letters,* vol. 96, Mar 17 2006.

[93] G. Derveaux, G. Papanicolaou, and C. Tsogka, "Time reversal imaging for sensor networks with optimal compensation in time," *Journal of the Acoustical Society of America,* vol. 121, pp. 2071-2085, Apr 2007.

[94] G. Lerosey, J. De Rosny, A. Tourin, and M. Fink, "Focusing beyond the diffraction limit with far-field time reversal," *Science,* vol. 315, pp. 1120-1122, Feb 23 2007.

[95] D. H. Liu, S. Vasudevan, J. Krolik, G. Bal, and L. Carin, "Electromagnetic time-reversal source localization in changing media: Experiment and analysis," *Ieee Transactions on Antennas and Propagation,* vol. 55, pp. 344-354, Feb 2007.

[96] D. Francoeur and A. Berry, "Time reversal of flexural waves in a beam at audible frequency," *Journal of the Acoustical Society of America,* vol. 124, pp. 1006-1017, Aug 2008.

[97] F. Ciampa and M. Meo, "Acoustic emission localization in complex dissipative anisotropic structures using a one-channel reciprocal time reversal method," *Journal of the Acoustical Society of America,* vol. 130, pp. 168-175, Jul 2011.

[98] A. M. Thode, W. A. Kuperman, G. L. D'Spain, and W. S. Hodgkiss, "Localization using Bartlett matched-field processor sidelobes," *Journal of the Acoustical Society of America,* vol. 107, pp. 278-286, Jan 2000.

[99] M. Cooke, G. J. Brown, M. Crawford, and P. Green, "Computational auditory scene analysis - Listening to several things at once," *Endeavour,* vol. 17, pp. 186-190, 1993.

[100] M. P. Cooke and G. J. Brown, "Computational auditory scene analysis - Exploiting principles of perceived continuity," *Speech Communication,* vol. 13, pp. 391-399, Dec 1993.

[101] G. J. Brown and M. Cooke, "Computational auditory scene analysis," *Computer Speech and Language,* vol. 8, pp. 297-336, Oct 1994.

[102] D. Godsmark and G. J. Brown, "A blackboard architecture for computational auditory scene analysis," *Speech Communication,* vol. 27, pp. 351-366, Apr 1999.

[103] P. Li, Y. Guan, B. Xu, and W. J. Liu, "Monaural speech separation based on computational auditory scene analysis and objective quality assessment of speech," *IEEE Transactions on Audio Speech and Language Processing*, vol. 14, pp. 2014-2023, Nov 2006.

[104] Y. Shao, S. Srinivasan, Z. Z. Jin, and D. L. Wang, "A computational auditory scene analysis system for speech segregation and robust speech recognition," *Computer Speech and Language*, vol. 24, pp. 77-93, Jan 2010.

[105] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources .2. Problems statement," *Signal Processing*, vol. 24, pp. 11-20, Jul 1991.

[106] C. Jutten and J. Herault, "Blind separation of sources .1. An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1-10, Jul 1991.

[107] E. Sorouchyari, "Blind separation of sources .3. Stability analysis," *Signal Processing*, vol. 24, pp. 21-29, Jul 1991.

[108] A. J. Bell and T. J. Sejnowski, "An Information Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, Nov 1995.

[109] A. Belouchrani, K. AbedMeraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, pp. 434-444, Feb 1997.

[110] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the Ieee*, vol. 86, pp. 2009-2025, Oct 1998.

[111] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, pp. 87-90, Apr 1999.

[112] A. Mansour, A. K. Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications," *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. E83a, pp. 1498-1512, Aug 2000.

[113] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari. (2009). *Nonnegative matrix and tensor factorizations applications to exploratory multi-way data analysis and blind source separation*.

[114] M. A. Bee and C. Micheyl, "The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?," *Journal of Comparative Psychology,* vol. 122, pp. 235-251, Aug 2008.

[115] J. H. McDermott, "The cocktail party problem," *Current Biology,* vol. 19, pp. R1024-R1027, Dec 1 2009.

[116] P. Comon, "Independent component analysis, a new concept," *Signal Processing,* vol. 36, pp. 287-314, Apr 1994.

[117] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation,* vol. 9, pp. 1483-1492, Oct 1 1997.

[118] A. de Cheveigne and H. Kawahara, "Multiple period estimation and pitch perception model," *Speech Communication,* vol. 27, pp. 175-185, Apr 1999.

[119] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature,* vol. 401, pp. 788-791, Oct 21 1999.

[120] D. L. L. Wang and G. J. Brown, "Separation of speech from interfering sounds based on oscillatory correlation," *IEEE Transactions on Neural Networks,* vol. 10, pp. 684-697, May 1999.

[121] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks,* vol. 13, pp. 411-430, May-Jun 2000.

[122] L. Potamitis, N. Fakotakis, and G. Kokkinakis, "Independent component analysis applied to feature extraction for robust automatic speech recognition," *Electronics Letters,* vol. 36, pp. 1977-1978, Nov 9 2000.

[123] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent component analysis*. New York: J. Wiley, 2001.

[124] A. Cichocki, S. Amari, R. Zdunek, R. Kompass, G. Hori, and Z. H. He, "Extended SMART algorithms for non-negative matrix aactorization," *Artificial Intelligence and Soft Computing - Icaisc 2006, Proceedings*, vol. 4029, pp. 548-562, 2006.

[125] A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's divergences for non-negative matrix factorization: Family of new algorithms," *Independent Component Analysis and Blind Signal Separation, Proceedings*, vol. 3889, pp. 32-39, 2006.

[126] P. von Bunau, F. C. Meinecke, F. C. Kiraly, and K. R. Muller, "Finding stationary subspaces in multivariate time series," *Physical Review Letters*, vol. 103, Nov 20 2009.

[127] D. Kolossa, R. F. Astudillo, E. Hoffmann, and R. Orglmeister, "Independent component analysis and time-frequency masking for speech recognition in multitalker conditions," *Eurasip Journal on Audio Speech and Music Processing*, 2010.

[128] A. Ozerov and C. Fevotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, pp. 550-563, Mar 2010.

[129] M. Kuroda, Y. Mori, M. Iizuka, and M. Sakakihara, "Acceleration of the alternating least squares algorithm for principal components analysis," *Computational Statistics & Data Analysis*, vol. 55, pp. 143-153, Jan 1 2011.

[130] A. H. Quazi, "An overview on the time-delay estimate in active and passive systems for target localization," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 29, pp. 527-533, 1981.

[131] C. H. Knapp and G. C. Carter, "Generalized correlation method for estimation of time-delay," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 24, pp. 320-327, 1976.

[132] J. E. Ehrenberg, T. E. Ewart, and R. D. Morris, "Signal-processing techniques for resolving individual pulses in a multipath signal," *Journal of the Acoustical Society of America*, vol. 63, pp. 1861-1865, 1978.

[133] G. C. Carter, "Time-delay estimation for passive sonar signal-processing," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 29, pp. 463-470, 1981.

[134] D. M. Etter and S. D. Stearns, "Adaptive estimation of time delays in sampled data systems," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 29, pp. 582-587, 1981.

[135] J. C. Hassab and R. E. Boucher, "Performance of the generalized cross correlator in the presence of a strong spectral peak in the signal," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 29, pp. 549-555, 1981.

[136] L. E. Miller and J. S. Lee, "Error analysis of time-delay estimation using a finite integration time correlator," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 29, pp. 490-496, 1981.

[137] N. L. Owsley and G. R. Swope, "Time-delay estimation in a sensor array," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 29, pp. 519-523, 1981.

[138] F. A. Reed, P. L. Feintuch, and N. J. Bershad, "Time-delay estimation using the Lms adaptive filter - static behavior," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 29, pp. 561-571, 1981.

[139] J. P. Ianniello, "Time-delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 30, pp. 998-1003, 1982.

[140] D. H. Youn, N. Ahmed, and G. C. Carter, "On using the Lms algorithm for time-delay estimation," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 30, pp. 798-801, 1982.

[141] M. Azaria and D. Hertz, "Time-delay estimation by generalized cross-correlation methods," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 32, pp. 280-285, 1984.

[142] D. H. Youn and V. J. Mathews, "Adaptive realizations of the maximum-likelihood processor for time-delay estimation," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 32, pp. 938-940, 1984.

[143] T. J. Abatzoglou, "A fast local maximum-likelihood estimator for time-delay estimation," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 34, pp. 375-378, Apr 1986.

[144] J. P. Ianniello, "Large and small error performance limits for multipath time-delay estimation," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 34, pp. 245-251, Apr 1986.

[145] W. Lindemann, "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *Journal of the Acoustical Society of America,* vol. 80, pp. 1608-22, Dec 1986.

[146] J. A. Stuller, "Maximum-likelihood-estimation of time-varying delay .1.," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 35, pp. 300-313, Mar 1987.

[147] R. J. Tremblay, G. C. Carter, and D. W. Lytle, "A practical approach to the estimation of amplitude and time-delay parameters of a composite signal," *IEEE Journal of Oceanic Engineering,* vol. 12, pp. 273-278, Jan 1987.

[148] Y. Barshalom, F. Palmieri, A. Kumar, and H. M. Shertukde, "Analysis of wide-band cross-correlation for time-delay estimation," *IEEE Transactions on Signal Processing,* vol. 41, pp. 385-387, Jan 1993.

[149] J. K. Tugnait, "Time-delay estimation with unknown spatially correlated Gaussian-noise," *IEEE Transactions on Signal Processing,* vol. 41, pp. 549-558, Feb 1993.

[150] T. G. Manickam, R. J. Vaccaro, and D. W. Tufts, "A least-squares algorithm for multipath time-delay estimation," *IEEE Transactions on Signal Processing,* vol. 42, pp. 3229-3233, Nov 1994.

[151] P. M. Schultheiss, H. Messer, and G. Shor, "Maximum likelihood time delay estimation in non-Gaussian noise," *IEEE Transactions on Signal Processing,* vol. 45, pp. 2571-2575, Oct 1997.

[152] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *Journal of the Acoustical Society of America,* vol. 107, pp. 384-391, Jan 2000.

[153] J. Benesty, J. D. Chen, and Y. T. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Transactions on Speech and Audio Processing,* vol. 12, pp. 509-519, Sep 2004.

[154] J. D. Chen, J. Benesty, and Y. T. Huang, "Time delay estimation in room acoustic environments: An overview," *Eurasip Journal on Applied Signal Processing*, 2006.

[155] C. I. Chang and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, pp. 2387-2396, Sep 1999.

[156] P. Lehmann, "Fast time-frequency analysis based on the short time Fourier transform," *Technisches Messen*, vol. 64, pp. 247-258, Jun 1997.

[157] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *IEEE Transactions on Signal Processing*, vol. 48, pp. 2964-2972, Oct 2000.

[158] R. M. Rao and A. S. Bopardikar, *Wavelet transforms : introduction to theory and applications*. Reading, MA: Addison-Wesley, 1998.

[159] L. Debnath, *Wavelet transforms and time-frequency signal analysis*. Boston: Birkhäuser, 2001.

[160] L. Debnath, *Wavelet transforms and their applications*. Boston: Birkhäuser, 2002.

[161] J. B. Snow, *Epidemiology of tinnitus in Tinnitus : theory and management*. Hamilton, Ont.Lewiston, NY: B.C. Decker ;Sales and distribution, U.S., B.C. Decker, 2004.

[162] A. Man and L. Naggan, "Characteristics of tinnitus in acoustic trauma," *Audiology*, vol. 20, pp. 72-78, 1981.

[163] W. Muhlnickel, T. Elbert, E. Taub, and H. Flor, "Reorganization of auditory cortex in tinnitus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 10340-10343, Aug 18 1998.

[164] H. Komiya and J. J. Eggermont, "Spontaneous firing activity of cortical neurons in adult cats with reorganized tonotopic map following pure-tone trauma," *Acta Oto-Laryngologica*, vol. 120, pp. 750-756, Sep 2000.

[165] A. J. Norena and J. J. Eggermont, "Changes in spontaneous neural activity immediately after an acoustic trauma: implications for neural correlates of tinnitus," *Hearing Research*, vol. 183, pp. 137-153, Sep 2003.

[166] W. Arnold, P. Bartenstein, E. Oestreicher, W. Romer, and M. Schwaiger, "Focal metabolic activation in the predominant left auditory cortex in patients suffering from tinnitus: A PET study with [F-18]deoxyglucose," *Orl-Journal for Oto-Rhino-Laryngology and Its Related Specialties,* vol. 58, pp. 195-199, Jul-Aug 1996.

[167] J. J. Eggermont, "Correlated neural activity as the driving force for functional changes in auditory cortex," *Hearing Research,* vol. 229, pp. 69-80, Jul 2007.

[168] C. A. Bauer, J. G. Turner, D. M. Caspary, K. S. Myers, and T. J. Brozoski, "Tinnitus and inferior colliculus activity in chinchillas related to three distinct patterns of cochlear trauma," *Journal of Neuroscience Research,* vol. 86, pp. 2564-2578, Aug 15 2008.

[169] A. D. Pierce, *Acoustics : an introduction to its physical principles and applications*, 1989 ed. Woodbury, N.Y.: Acoustical Society of America, 1989.

[170] A. D. Pierce, *Acoustics : an introduction to its physical principles and applications*, 1989 ed. Woodbury, N.Y.: Acoustical Society of America, 1989.

[171] M. Raffy and C. Gregoire, "Semi-empirical models and scaling: a least square method for remote sensing experiments," *International Journal of Remote Sensing,* vol. 19, pp. 2527-2541, Sep 10 1998.

[172] K. Hidajat and S. M. Chong, "Quality characterisation of crude oils by partial least square calibration of NIR spectral profiles," *Journal of near Infrared Spectroscopy,* vol. 8, pp. 53-59, 2000.

[173] Y. Chiu, C. W. Tsang, B. Nikolic, and P. R. Gray, "Least mean square adaptive digital background calibration of pipelined analog-to-digital converters," *Ieee Transactions on Circuits and Systems I-Regular Papers,* vol. 51, pp. 38-46, Jan 2004.

[174] D. R. Zhang, K. Beven, and A. Mermoud, "A comparison of non-linear least square and GLUE for model calibration and uncertainty estimation for pesticide transport in soils," *Advances in Water Resources,* vol. 29, pp. 1924-1933, Dec 2006.

[175] S. Bolognani, S. Del Favero, L. Schenato, and D. Varagnolo, "Consensus-based distributed sensor calibration and least-square parameter identification in WSNs," *International Journal of Robust and Nonlinear Control,* vol. 20, pp. 176-193, Jan 25 2010.

[176] H. S. Ege, Y. Ender, and D. Erdal, "Spectral Classical Least Square Calibration Approach for the Simultaneous Determination and Stability Test of Sulphadiazine and Trimethoprim in Bolus," *Journal of Animal and Veterinary Advances,* vol. 9, pp. 857-861, 2010.

[177] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems Man and Cybernetics,* vol. 15, pp. 116-132, 1985.

[178] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets and Systems,* vol. 28, pp. 15-33, Oct 1988.

[179] G. C. D. Sousa and B. K. Bose, "A fuzzy set-theory based control of a phase-controlled converter Dc machine drive," *IEEE Transactions on Industry Applications,* vol. 30, pp. 34-44, Jan-Feb 1994.

[180] L. Zhen and L. Y. Xu, "Fuzzy learning enhanced speed control of an indirect field-oriented induction machine drive," *IEEE Transactions on Control Systems Technology,* vol. 8, pp. 270-278, Mar 2000.

[181] S. N. Engin, J. Kuvulmaz, and V. E. Omurlu, "Fuzzy control of an ANFIS model representing a nonlinear liquid-level system," *Neural Computing & Applications,* vol. 13, pp. 202-210, Sep 2004.

[182] M. Kakar, H. Nystrom, L. R. Aarup, T. J. Nottrup, and D. R. Olsen, "Adaptive neuro fuzzy inference system (ANFIS) for prediction of respiratory motion," *Radiotherapy and Oncology,* vol. 76, pp. S91-S92, Sep 2005.

[183] M. Kakar, H. Nystrom, L. R. Aarup, T. J. Nottrup, and D. R. Olsen, "Respiratory motion prediction by using the adaptive neuro fuzzy inference system (ANFIS)," *Physics in Medicine and Biology,* vol. 50, pp. 4721-4728, Oct 7 2005.

[184] D. M. Gao and J. Chen, "ANFIS for high-pressure waterjet cleaning prediction," *Surface & Coatings Technology,* vol. 201, pp. 1629-1634, Oct 5 2006.

[185] F. Betin, A. Sivert, A. Yazidi, and G. A. Capolino, "Determination of scaling factors for fuzzy logic control using the sliding-mode approach: Application to control of a DC machine drive," *IEEE Transactions on Industrial Electronics,* vol. 54, pp. 296-309, Feb 2007.

[186] A. Depari, A. Flammini, D. Marioli, and A. Taroni, "Application of an ANFIS algorithm to sensor data processing," *IEEE Transactions on Instrumentation and Measurement,* vol. 56, pp. 75-79, Feb 2007.

[187] J. E. R. Dhas and S. Kumanan, "ANFIS for prediction of weld bead width in a submerged arc welding process," *Journal of Scientific & Industrial Research,* vol. 66, pp. 335-338, Apr 2007.

[188] S. Alavandar and M. J. Nigam, "Inverse kinematics solution of 3DOF planar robot using ANFIS," *International Journal of Computers Communications & Control,* vol. 3, pp. 150-155, 2008.

[189] C. Ozkan, "Adaptive Neuro-Fuzzy Inference System (Anfis) Based Local Ordinary Kriging Algorithm for Scattered Data Interpolation," *Survey Review,* vol. 41, pp. 395-407, Oct 2009.

[190] J. A. Perez, M. Gonzalez, and D. Dopico, "Adaptive neurofuzzy ANFIS modeling of laser surface treatments," *Neural Computing & Applications,* vol. 19, pp. 85-90, Feb 2010.

[191] Z. L. Yang, Y. Q. Liu, and C. R. Li, "Interpolation of missing wind data based on ANFIS," *Renewable Energy,* vol. 36, pp. 993-998, Mar 2011.

[192] Mathworks, "Fuzzy logic toolbox 2 user's guide," pp. 2-107-142 2011.

[193] H. Gavert, J. Hurri, J. Sarela, and A. Hyvarinen, "FastICA for Matlab 7.x and 6.x Version 2.5," *Helsinki University of Technology, Finland,* 2005.

ABSTRACT

LOCATING AND EXTRACTING ACOUSTIC AND NEURAL SIGNALS

by

NA ZHU

May 2012

Advisor: Dr. Sean F. Wu

Major: Mechanical Engineering

Degree: Doctor of Philosophy

This dissertation presents innovate methodologies for locating, extracting, and separating multiple incoherent sound sources in three-dimensional (3D) space; and applications of the time reversal (TR) algorithm to pinpoint the hyper active neural activities inside the brain auditory structure that are correlated to the tinnitus pathology. Specifically, an acoustic modeling based method is developed for locating arbitrary and incoherent sound sources in 3D space in real time by using a minimal number of microphones, and the Point Source Separation (PSS) method is developed for extracting target signals from directly measured mixed signals. Combining these two approaches leads to a novel technology known as Blind Sources Localization and Separation (BSLS) that enables one to locate multiple incoherent sound signals in 3D space and separate original individual sources simultaneously, based on the directly measured mixed signals. These technologies have been validated through numerical simulations and experiments conducted in various non-ideal environments where there are non-negligible, unspecified sound reflections and reverberation as well as interferences from random background noise. Another innovation presented in this dissertation is concerned with applications of the TR algorithm to pinpoint the

exact locations of hyper-active neurons in the brain auditory structure that are directly correlated to the tinnitus perception. Benchmark tests conducted on normal rats have confirmed the localization results provided by the TR algorithm. Results demonstrate that the spatial resolution of this source localization can be as high as the micrometer level. This high precision localization may lead to a paradigm shift in tinnitus diagnosis, which may in turn produce a more cost-effective treatment for tinnitus than any of the existing ones.

AUTOBIOGRAPHICAL STATEMENT

NA ZHU

EDUCATION

Ph.D. in Mechanical Engineering, Wayne State University, 2012

M.S. in Mechanical Engineering, Wayne State University, 2009

B.S. in Automobile Engineering, Tongji University, 2007

PUBLICATIONS

S. F. Wu and N. Zhu, "Locating arbitrarily time-dependent sound sources in 3D space in real time," Journal of the Acoustical Society of America, Vol. 128, 728 – 739 (2010).

N. Zhu and S. F. Wu, "Sound source localization in three-dimensional space in real time with redundancy checks," Journal of Computational Acoustics (accepted, 2011).

N. Zhu and S. F. Wu, "Error analysis of triangulation based source localizations," IEEE Transactions on Instrumentation and Measurement (submitted, 2011).

S. F. Wu and N. Zhu, "Sonic detection and ranging," Journal of the Acoustical Society of America (submitted, 2011).

S. F. Wu and N. Zhu, "Blind extraction of sound signals," Journal of the Acoustical Society of America (submitted, 2011).

RESEARCH GRANTS

N. Zhu, "Development of an innovative, 3D computer aided diagnostic system for tinnitus," American Tinnitus Association (ATA), $10,000, 05/01/11 – 04/40/12. Principal Investigator with Dr. S. F. Wu and Dr. J. Zhang as Co-Principal Investigators.

N. Zhu, Supplemental Research Grant from the Office of the Vice President for Research, WSU, $1,000, 05/01/11 – 04/40/12.

RESEARCH AWARDS

Winner of the Best Paper in Signal Processing by a Young Presenter, the 159th Meeting of the Acoustical Society of America, Baltimore, Maryland, April 2010. (N. Zhu and S. F. Wu, "Error analysis of locating arbitrary sound sources in 3D space in real time").

Winner of the 2009 Best Student Paper Competition and Grand Price Cash Award, the 2009 ASME International Mechanical Engineering Congress and Exposition, Lake Buena Vista, Florida, November 2009. (N. Zhu and S. F. Wu, "Track and trace multiple incoherent sound sources in 3D space in real time").

The Second Place Award at the Best Students Papers Competition, the 157th Meeting of the Acoustical Society of America, Portland, Oregon, May 2009. (N. Zhu and S. F. Wu, "Locating arbitrary noise sources in 3D space in real time").