Sound Source Localization by Phase Signature

David Lewis Graumann

A thesis presented to the faculty of the
OGI School of Science & Engineering
at Oregon Health & Science University
in partial fulfillment of the
requirements for the degree
Masters of Science
in
Electrical and Computer Engineering

June 2003

The Thesis "Sound Source Localization by Phase Signature" by David Lewis Graumann has been examined and approved by the following Examination Committee:

Dr. Eric Wan
Associate Professor
Thesis Research Advisor

Dr. Hynek Hermansky
Professor and Director of CIT

Dr. Xubo Song
Assistant Professor

ACKNOWLEDGEMENTS

Thank you Dr. Eric Wan for teaching me the fundamentals of digital signal processing in the classroom and encouraging their application to this Thesis. I am grateful. Your recommendation that a Thesis will be more rewarding than completing arbitrary engineering courses has certainly rung true. Thank you also for tolerating a few false starts and not giving up on my intent to graduate.

Thank you Professor Hynek Hermansky for your support of this work and in particular for two comments that have resonated in my thinking well beyond their RT60. I paraphrased them as follows: "The microphone transducer, designed for telecommunications, is one of the most limiting components of a speech recognition systems and the fundamentals of audio capture should be reexamined." and "It is probable that the human ear has evolved as an optimal receiver for speech and sound". These two comments have sparked my pondering, "Why are localization arrays designed to avoid the natural phenomenon of reverberation?"

DEDICATION

To My Love Heidi,

Who cared for three very young children and overlooked a disheveled garage during my studies.

TABLE OF CONTENTS

LIST OF TABLES & FIGURES

ABSTRACT

Sound Source Localization by Phase Signature

David Lewis Graumann, B.S. EE

M.S., Oregon Graduate Institute of Science and Technology
June 2003

Thesis Advisor: Dr. Eric Wan

Acoustic Source Localization involves methods that can detect the location of a particular audio signal within a two dimensional or three dimensional field. These methods require multiple microphones for simultaneously capturing the audio signal. The primary strategies used today are: Filter & Sum Beamforming, High Resolution Spectral Analysis, and Time Difference of Arrival Estimation. When any of these approaches are limited to only two microphones they are able to determine the sound's direction of arrival but unable to determine the sound's range. This thesis explores a novel approach to enhancing the spatial selectivity of a two-microphone array by combining Time Difference of Arrival methods with acoustical reverberation principles. Range determination using unique acoustical phase signatures within a designated 'listening region' of the array is demonstrated. The ability to make such a distinction with only two microphones can be used when building inexpensive audio capture devices requiring capabilities such as: camera steering, speech recognition end-pointing, and noise mitigation.

1 BACKGROUND, MOTIVATION, & COVERAGE

1.1 BACKGROUND

Techniques to identify the remote location of an emitted sound have been in use to solve real-world challenges for over 50 years. These methods have evolved out of sonar localization techniques which themselves have stemmed from methods applied to narrowband radio signals. Narrowband applications primarily include those for radar emission direction-of-arrival and geological surveillance. In some situations, the localization technology has the luxury of controlling the signals being used to determine the location. The transmitter and the receiver are matched in some optimal way to best measure the signal's location. These are termed *active* technologies [27]. Pings, Chirps, pseudo-noise signal sequences are tailored to the specific conditions being confronted. For example radar signals originally consisted of single sinusoids with precise modulation schemes. The other class of localization systems do not have the luxury of modifying the emitted signal and must resolve the location by interpreting the sound emitted from the object of interest. This class is called *passive*. In this case, at best, a priori knowledge of the sound source's characteristics are generalized and considered when designing the receiver.

This paper focuses on *passive* acoustic sound source localization within the frequency region of human speech. This situation is commonly encountered in telecommunications and man-machine interfaces when microphones are used to capture and analyze the human voice. This particular area of research has been underway for over 20 years and presents a unique set of challenges over narrowband signals used in sonar and radar communications. Applications that use these technologies include hands-free speech dictation, distance learning lecture broadcasting, large room video conferencing, and automatic speech end-pointing for telematics. In all of these applications,

sound source localization has been successfully applied using systems with multiple microphones. Telecommunications companies like Polycom™ and VTEL™ have deployed speech localization technologies to point video cameras at active talkers. The devices will estimate the location of talkers within a conference room and provide a steerable camera with direction-of-arrival information to assist in automatic transmission of well-framed video images. The camera with then use other means to adjust the focal range to the speaker [18][23][32]. These systems are costly and require room calibration sequences to operate. Less expensive systems have been introduced by companies such as Andrea Electronics™, Emkay Innovations™, and Acoustic Magic™ for use with desktop computers and kiosks. These devices often use localization to place acoustic nulls on arrival angles of disinterest. By doing so, they can establish better signal quality of the location of interest. These systems use 4-8 microphones and crude signal energy for range discrimination. This simplistic design principle can be easily shown to break down by speaking loudly at a distance beyond their recommended range.

The fundamental objective of these applications and devices is to identify the sound source in a 2D or 3D space. The signal processing strategies used to achieve this fall into three basic categories. The oldest and most rudimentary method is the Filter and Sum Beamformer. Originally promoted by Frost and Griffith & Jim [19][39], these methods search for a source location that maximizes the signal power output of several microphones. This is done by delaying and filtering each microphone input and then summing together. The delays that create the maximum output power are used in conjunction with knowledge of the microphone's physical arrangement to determine the angle of arrival of arrival of the dominant sound [21]. These methods have been shown to hold up well in reverberant conditions but do not offer a sharp peak at the sound source's angle of arrival [20]. The second class of sound localization methods examine Spectral Estimations stemming for the correlation matrix of several microphones. These methods look for spectral coherency between

microphone signals to predict the source location [5][21]. The primary focus is on eigenvalue decomposition techniques using several microphones and does not scale down to the two microphone configurations in a meaningful way. These methods have also been reported to have stability problems that can cause them to cancel out the sound of interest under reverberant conditions [20][21]. The third group of source localization methods, and one that is popular in small array configurations, are those that first parameterize the data into Time Delay Estimations (TDE) between pairs of microphones, translate the delays into angle of arrivals, then use geometric triangulation of the angles to pinpoint the source. This method provides an improvement over the other two in the ability to locate the angle of arrival of the sound source [6][14] and also easily scales down to two microphones.

All of these signal processing strategies utilize configurations with more than two microphones for acquiring the signal and even then often only calculate the sound source's direction of arrival and not the range to the microphones. Keeping this common limitation in mind, from these three technology options, the most suitable strategy for two microphones is the one based on Time Delay Estimation between microphone pairs. This was selected for its scalability and proven track record with 4-8 microphone arrays. It will be shown how using only two microphones with this method can be modified to resolve the range ambiguity.

Looking closer at Time Delay Estimation shows it was unified under a single formulation by Knapp & Carter [25]. They proposed a general description of Time Delay Estimation for narrowband signals termed Generalized Cross-Correlation (GCC). Their contribution detailed the various analytical options available for applying cross-correlation to the time delay problem. They investigate the various methods for calculating time delays between two signals that have propagated through two unknown channels. Their methods, however, were not specific to broadband sound source localization.

Specific to microphone arrays, TDE between two signals is most commonly tackled with methods that measure Time Difference of Arrival (TDOA) between to microphones. The difference of arrival time of a known or unknown acoustical source is measured between a pair of microphone signals [6][35]. Two microphones are used in free space to establish a hyperboloid that defines the set of possible source locations (this will be discussed further). Additional microphones are used to establish additional hyperboloids with their common intersect being the final location resolve [9][12]. These methods add cost with additional microphones and circuitry, they add latency in the pair-wise parameterization step, and the are shown to loose location resolution over more direct approaches [14].

An improvement on the TDOA parameterization was introduced by DeBiase [14]. In this work, he moved directly from the phase representation of the TDOA into the location's Cartesian coordinate system. The phase representation is created from pre-calculated TDOA values determined by the microphone arrangement. Doing show sharpened the peaks of the location measurement over the existing TDOA method. Results were shown using 15-512 microphones against an acoustically treated backdrop to avoid reflections. In this case, just like the other TDOA methods, very simple linear phase relationships are used, several microphones were use, and reverberation was avoided. Even under these conditions, this method is studied in a large conference room for determining angle of arrival measurements only and not for determining range.

1.2 MOTIVATION

This thesis work stems directly from observations of state-of-the-art audio capture techniques used on small form factor devices. The arrays being deployed today require many microphones and large apertures to perform well. Video conferencing systems require 30-100 cm microphone separation along 3 axes [16]. Desktop arrays on the market today use four to eight microphones with

30 cm end-to-end element placement. The microphone elements are either suspended in free space or dampened with acoustical foam. The most limiting aspect is that they only resolve the sound source to its direction of arrival and not its range. It would be preferred to resolve the location of the sound source in a three-dimensional space with a minimum hardware and manufacturing cost.

With the emergence of consumer electronic devices such as mobile laptop computers, camera phones, wireless video conferencing, and mobile speech recognition on low-power low-cost consumer devices, it becomes an interesting challenge to determining when a speech source is appropriate for audio analysis and transmission on a small form factor device when positioned at arm's length from a talker in an uncontrolled acoustical setting. Being able to spatially filter the direction and range of a sound could provide the ability to remove or select sounds of interest for the device.

Extending this concept a bit further, a futuristic motivation for inexpensive sound localization is the possibility of all home appliances and electronic equipment to contain built-in audio capture for speech activated command and control. Directing speech to recognition-enabled devices within the home without adjacent devices misinterpreting the commands would benefit from fine-grain spatial resolution of the sound source with minimum hardware requirements. Assuming all audio is routed to a centralized home computer, then it is conceivable that only the simple circuitry of the audio capture front-end needs to be embedded in devices such as microwave ovens, home stereo systems, and light switches.

One method to help distinguish when the audio is of interest is to consider its location with respect to the audio capturedevice. If it is being gene rated from within a prescribed location, then it is more likely to be produced by a single source. The intent is to better distinguish between background babble and those of the user of the device. When the talker of interest is active, their

voice can often mask these background signals. Just how to distinguish between the two is still of great importance. The need to provide small, spatially selective speech capture is of primary interest to manufactures with products within the cellular to sub notebook class of wireless devices. This could be used to center the camera on the user's face, provide robust end-pointing of speech utterances for front-end processing to a speech recognition engine, identify silent segments where noise characteristic can be evaluated, and assist in adaptive speech enhancement methods by tracking signals of both interest and disinterest.

In light of the scenarios described above, one final motivation for this thesis was to look at two microphone configurations. This is the minimal cost reduction available for Time Difference of Arrival techniques. Extending this work back to multiple microphones is always an option at a later point, but to achieve range and direction of arrival with two microphones with any sound source with in the speech frequency region is the motivational first step. So, although the work here extends easily to additional microphones, two elements were used in experimentations because these techniques lend themselves to implementation with inexpensive stereo analog to digital converters.

As previously stated, is was observed that all existing methods for determining sound source localization restrict reflections because they cause distortions in the easily calculated and predictable relationships between microphone topologies, signal propagation, and signal location. This seems counter intuitive when contemplating the acoustical and physical characteristics of the human pinna. The human ear provides a remarkable ability to estimate source location [40]. Characteristics include inter-aural time differences, amplitude alterations, head and shoulder reflections, as well as the ability to alter head position within the acoustical field [28]. However, the human ear did not evolve to remove reverberation. In fact it has reverberant characteristics [4][15]. Although the microphone is by no means a human ear, this thesis was intrigued by the idea that maybe sound source localization methods should look to exploit reverberation rather than avoid it. Though many of the previously

mentioned techniques analyze their methods in the presents of reverberation [3][8][10][11], only a few bodies of work attempted to capitalize on reverberation for sound enhancement [22][29] and none were identified that addressed its use for sound source localization.

From the technology review, motivation, and intrigue a very simple question is formulated to driver this work – Can the deliberate introduction of a reverberant surface in close proximity to two microphones be used to resolve both a sound source's direction of arrival and range from those microphones.

1.3 THESIS COVERAGE

First the theoretical background of Time Delay Estimation is presented with the primary focus on Time Difference Of Arrival. Next the concept of translating the TDOA measurements to a sound location will be described. The principle methods for solving the source localization problem will be provided as a foundation to the extensions set forth by this work. In the method proposed, we move directly from a measurement of microphone phase difference to the location that best matches the phase signature of the reverberant enclosure. This method can introduce mathematically complicated mappings between source location and phase angle. By doing so, an opportunity is created to resolve the sound source range using as few as two microphone elements.

Background information on acoustic wave propagation is established for justifying both the TDOA principles as well as the reverberation chamber analysis. A model of source location in a reverberant chamber is presented. Theoretical results from the model are shown, suggesting the possibility of real-world benefits to this approach. Next a real-time implementation is developed and instrumented into a complete test and logging harness. This facilitated the exploration of different reflective surfaces and enclosure arrangements. A reflective surface was chosen and measured for a set of 15 locations. Measurements of the microphone and reflective surface's geometrical

arrangement are made with a conventional tape measure and acoustically. The model is improved and compared to the theoretical phase signatures. After selecting final phase signatures for the enclosure, 15 locations are tested with a reflective surface. The results are compared to the non-reflective free space microphones. Conclusions are drawn on the over all performance of this new method and the improved contour of the location search space along the most difficult regions of range discrimination.

2 THEORETICAL

2.1 FUNDAMENTALS OF TDOA

The objective of Time Difference of Arrival for sound source localization is to accurately determine the difference in time it takes for a single sound to propagate to a pair of microphones. By obtaining this delay it is possible to calculate a set of positions for a point source. Consider the scenario where a sound source $s(n)$ emanating from single location is being recorded simultaneously by two microphones $m_1(n)$ and $m_2(n)$. See Figure 2.1. The same signal propagates along two different paths to each microphone. Given this configuration we have the relationship

$$m_i(n) = g_i(n) \otimes s(n) + v_i(n) \tag{1.1}$$

Where (n) represents discrete time sampling, $m_i(n)$ represents the signal received at the microphone, $g_i(n)$ represents the acoustical transfer function between the source $s(n)$ and the microphone's analog to digital converter (ADC), $v_i(n)$ represents noise that is uncorrelated with $s(n)$, and $\otimes$ denotes convolution. It is common practice to express $g_i(n)$ as a linear causal impulse response. In the typical case $g_i(n)$ is unknown and could be time varying. Fortunately we can learn something about the location of $s(n)$ by observing the characteristics between pairs of microphone signals without full knowledge of $g_i(n)$.

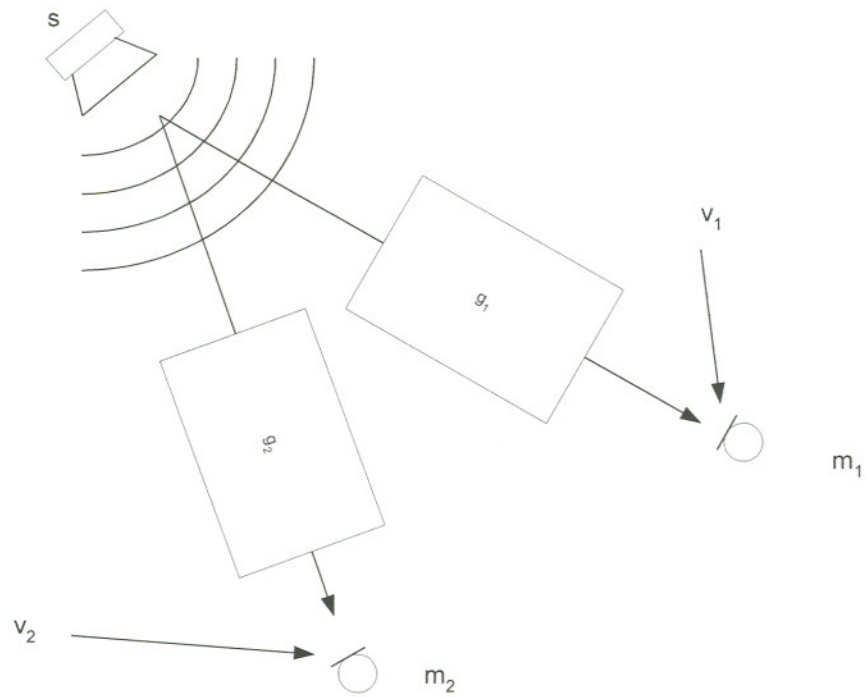


Figure 2.1: Acoustical Localization System Configuration.

As a first step towards understanding the signal characteristics of a microphone pair, we turn to the Generalized Cross-Correlation (GCC) put forth by Knapp and Carter [25] to described the unified theory behind several methods in Time Delay Estimation.. The GCC provides us with an optimal expression for the cross-spectrum approach to delay estimation. See Figure 2.2

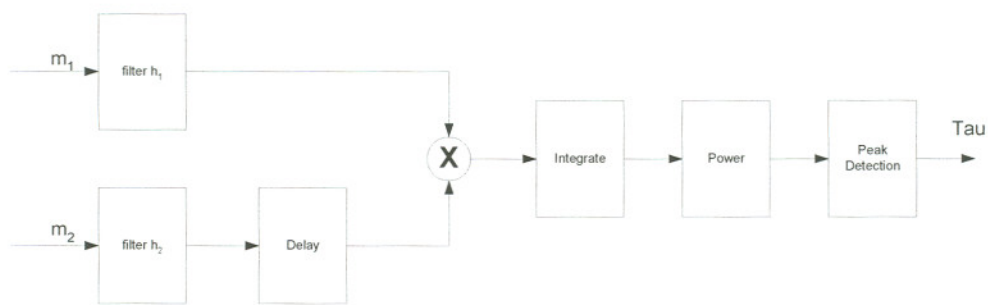


Figure 2.2: Generalized Cross Correlation Block Diagram by Knapp & Carter [25].

Here we consider the case where an unknown signal has propagated along two different acoustical paths to two microphones. As seen by the block diagram, the GCC states that a time delay can be estimated by first applying prescribed filters $h_i(n)$ to $m_i(n)$, then delaying one signal with respect to the other, multiplying the two signals together, integrating over time, and finding a peak in the signals output power. This works best for stationary non-periodic signals, but with some adjustment and short-term windowing this can be made suitable for time varying periodic signals as well. GCC can be applying in either the frequency or the time domain. We will develop the frequency domain representation because this is the domain where the final source location is determined. For the noiseless case, GCC in the frequency domain is expressed as:

$$R_{m_1 m_2}(\tau) \equiv \int_{-\infty}^{\infty} \left(H_1(\omega) G_1(\omega) S(\omega) H_2^*(\omega) G_2^*(\omega) S^*(\omega) \right) e^{j\omega\tau} d\omega \tag{1.2}$$

where $H_i(\omega)$, $G_i(\omega)$, & $S(\omega)$ are the frequency domain representations of $h_i(n)$, $g_i(n)$, & $s(n)$ respectively. This is simplified by the following assignments:

Let the GCC pre-filters be stated as

$$\Psi(\omega) \equiv H_1(\omega) H_2^*(\omega) \tag{1.3}$$

Let an observable microphone signal be stated as

$$X_{m_i}(\omega) \equiv G_i(\omega) S(\omega) \tag{1.4}$$

Further more, since only differences of microphone signals are used let

$$X_{m1m2}(\omega) \equiv X_{m1}(\omega) X_{m2}^*(\omega) \tag{1.5}$$

This form is often called the cross-power spectrum. We now have a manageable representation of the GCC for our situation.

$$\arg\max_{\tau}\left[R_{m1m2}(\tau)=\int_{-\infty}^{\infty}\Psi(\omega)X_{m1m2}(\omega)e^{j\omega\tau}d\omega\right] \tag{1.6}$$

The choice of $\Psi(\omega)$ is required for realization of the GCC. In the noiseless case, the optimal filter is the one that inverts the room acoustics along each path from the source. Specifically $\Psi(\omega)=G_1^{-1}(\omega)G_2^{-1}(\omega)$. Since there is no access to even an approximation of these functions, a suitable replacement is needed. Inspection of the simplest case where $G_i(\omega)$ is a single tapped delay line of fixed attenuation yields.

$$X_{m_i}(\omega)=G_i(\omega)S(\omega)=\alpha_i e^{j\omega\tau_i}S(\omega) \tag{1.7}$$

Substituting this into the cross-spectrum expectation leaves

$$X_{m_1m_2}(\omega)=\alpha_1\alpha_2 e^{j\omega(\tau_1-\tau_2)}\left|S(\omega)\right|^2 \tag{1.8}$$

This will be the simplest case used to give insight into the nature of the TDOA problem.

For our target application we do know that the ultimate signal of interest is speech. Though we build up the theories and experiments with a broadband test signals, the design choice at this stage is to use one that is eventually suitable for speech. This suggests that there will be harmonics, narrow voiced segments, as well as broader band aspirates and fricatives [26]. We also know that these characteristics will be time varying over 20-100 msec segments assuming the spoken English word. It has been shown in [30][31][33] that the selection of $\Psi(\omega)=\left|X_{m_1}(\omega)\right|^{-1}\left|X_{m_2}(\omega)\right|^{-1}$ is a very good

choice in practice primarily because it completely whitens the spectrum and does not overly emphasis strong transient tones. This selection is often referred to as the GCC Phase Transform (GCC-PHAT). Additionally, it has been shown that speech specific filters that interpret harmonic structures provide superior TDOA peaks under low signal-to-noise conditions[7][10]. For our investigation, we will use GCC-PHAT because it allows us to illustrate this research without unnecessary complications or simplifications. That said, GCC-PHAT should not be applied blindly, because in low signal-to-noise conditions or when the source is not present at a given frequency within a valid segment of audio, erroneous phase results will be obtained. Steps to alleviate these problems are described and implemented in later sections.

Substituting this choice of $\Psi(\omega)$ into (1.6) gives the final GCC-PHAT as

$$\arg\max_{\tau}\left[\hat{R}_{m1m2}(\tau)=\int_{-\infty}^{\infty}\frac{X_{m1m2}(\omega)}{\left|X_{m1}(\omega)\right|\left|X_{m2}(\omega)\right|}e^{j\omega\tau}d\omega\right] \tag{1.9}$$

where $\Psi(\omega)$ has simply normalized the magnitude of the microphone difference to unity, leaving only the phase angles as the TDOA determining factor.

In the absence of competing sound sources the time lag τ that created the largest spike in the cross-correlation function provides an estimate of the time difference of arrival for the two signals. The conventional solution to the localization problem at this point is to translate the estimates $\hat{\tau}=\arg\max_{\tau}\left[\hat{R}_{m1m2}(\tau)\right]$ into location coordinates. [5][6] [11][31][33][34][38]. Often this calculation is performed iteratively using expectations of the GCC function $E\left[\hat{R}_{m1m2}(\tau)\right]$ with signal windowing appropriate for speech and time constants appropriate for talker movement. The estimate of the TDOA between two microphones can then be geometrically related to a set of

possible source locations. This mapping is not unique and results in a locus of points along a hyperboloid as described in the following section.

2.2 SOUND PROPOGATION

The process of determining an unknown sound source's location in space from a TDOA measurement is grounded in the properties of acoustical wave propagation between the sound source and the microphones. To describe the source localization problem in terms of physical locations, we restate equation (1.1) for the simple acoustic scenario. Doing so will clearly illustrate the mapping from $\hat{\tau}$ to a physical location. Let $s(n)$ represent a signal at time n and at unknown location $\vec{s} \equiv (s_x, s_y, s_z)$. Let $m_i(n)$ represent the microphone signals at time n and known locations with respect to each other as $\vec{m}_i \equiv (m_{ix}, m_{iy}, m_{iz})$. The origin is arbitrary and will be chosen strictly for convenience.

Restating (1.1) with simple time delays gives us

$$m_i(n) = \alpha_i s(n - \tau_i) + v_i(n) \tag{1.10}$$

Again $v_i(n)$ is isotropic noise uncorrelated with $s(n)$. α_i is the attenuation factor of $s(n)$ as it propagates to each $m_i(n)$.

As stated above, τ_i and α_i are not known without additional knowledge of $s(n)$. The TDOA measurement has provided us with only $\tau_1 - \tau_2$. We will use this after setting up the basic problem statement.

Let d_1 and d_2 be the distance between $\vec{s}$ to $\vec{m}_1$ and $\vec{m}_2$ respectively. Let the difference in the distance the wave must travel between any two microphones be written as $D_{12} = | d_1 - d_2 |$. Regardless of whether we assume spherical or planar wave propagation principles [1][36], there is a measurable relationship between D_{12} and a set of locations where the sound source must reside. Though it will be shown that this will not matter in the final implementation we must assume something so a planar front is used. We now can state the relationship

$$D_{12} = \hat{\tau} C \tag{1.11}$$

where $C = 1087\, ft/\sec + (1.1\, ft/\sec) x (RoomTempFahrenheit - 32)$.

Expanding D_{12} gives us:

$$D_{12} = |\vec{m}_1 - \vec{s}| - |\vec{m}_2 - \vec{s}| \tag{1.12}$$

$$D_{12} = \sqrt{(s_x - m_{1x})^2 + (s_y - m_{1y})^2 + (s_z - m_{1z})^2} - \sqrt{(s_x - m_{2x})^2 + (s_y - m_{2y})^2 + (s_z - m_{2z})^2} \tag{1.13}$$

We can simplify these equations by selecting a convenient coordinate system. Let the x axis be the line defined by the points $(m_{1x}, 0, 0), (m_{2x}, 0, 0)$ with the origin in the middle of the two microphones. To illustrate the set of points defined by a single $\hat{\tau}$ measurement we will set all z to zero. If we let $c = |m_{1x} - m_{2x}|/2$ then we can restate (1.13) as

$$D_{12} = \sqrt{(s_x + c)^2 + s_y^2} - \sqrt{(s_x - c)^2 + s_y^2} \tag{1.14}$$

This equation can be reduced to a set of points on a hyperbola defined by

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1 \tag{1.15}$$

Where $a=\frac{|s-m_1|-|s-m_2|}{2}$ and $b=\sqrt{\left(\frac{|s-m_1|-|s-m_2|}{2}\right)^2+\left(\frac{|m1-m2|}{2}\right)^2}$. The foci are $\pm c$, the vertex is a, and the set of points is asymptotic to the line defined by $s_y=\frac{b}{a}s_x$.

This two dimensional equation defines a set of locations for source s on a hyperbola who's axis is along the line connecting the two microphones. The three dimensional locus is the hyperboloid around the same axis. In this way, by using two microphones, TDOA we can only determine the direction of arrival defined by the hyperboloid. We cannot determine the orientation along the space perpendicular to the $|m_1-m_2|$ axis. Referring to Figure 2.3 we can state this simply as - the difference between the time of flight of d_{a1} and d_{a2} will be the same as the difference between the time of flight of d_{b1} and d_{b2} along a hyperbola. The most obvious set of ambiguous locations is when the sound source lies along a plane perpendicular to $\overline{m_2-m_1}$ at $\left(m_{1x}+|m_{1x}-m_{2x}|/2,0,0\right)$. In this case all TDOA's are zero. This is a major limitation to determining range using only two microphones.

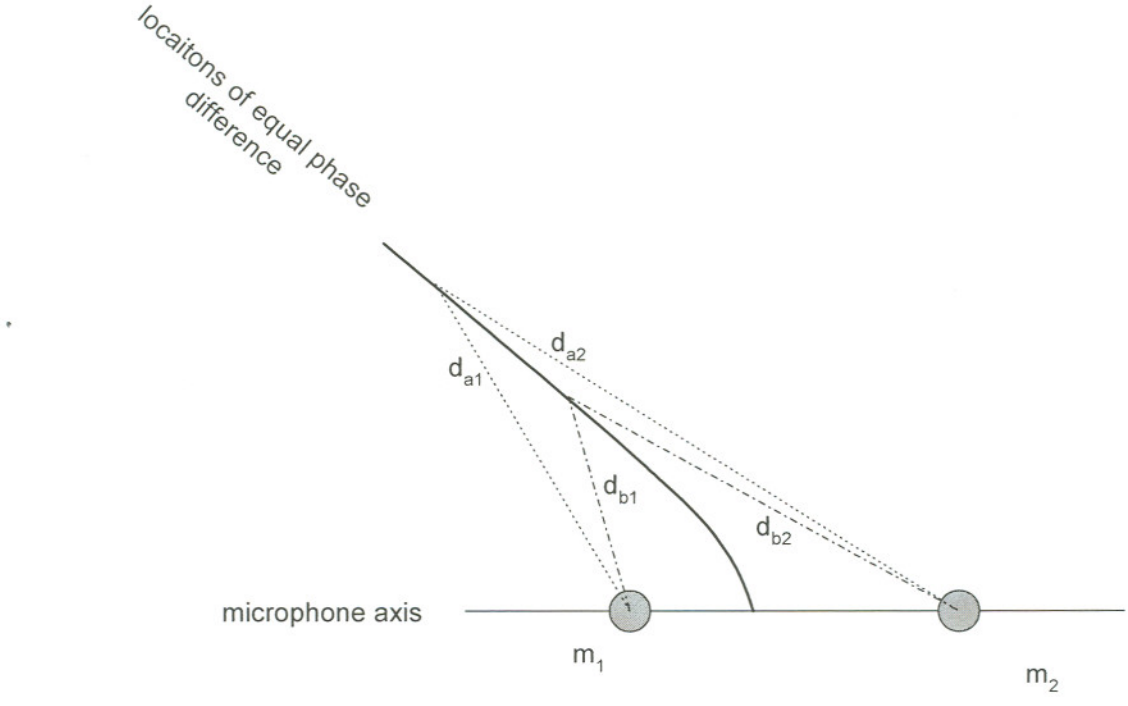


Figure 2.3: 2D Location of TDOA Ambiguity

Using methods based on this principle, which includes both Delay and Sum Beamforming and TDOA, there is simply no way to determine the range of the sound source without prior knowledge of the signal strength or transmission time. To overcome this ambiguity additional microphones are used with deliberate and often orthogonal positioning. [5][6][14][20][33] Doing so, can create a set of arrival angles or hyperboloids that can be solved simultaneously to a single point in space. Orthogonal positioning include placement on two orthogonal walls [33] as well as using 4 microphones arranged in a square [6]. These methods have been shown to work under low reverberation and low noise conditions. However, solving for the intersection of a set of hyperboloids has been shown to be sensitive to reverberation and microphone placement. It has been shown that closed form solutions do not exist under noisy conditions and maximum likelihood methods must be used to approximate the true location. [20]. Another important drawback to

adding additional microphones is the increased structure size and installation challenges to the localization device.

For some applications the Direction of Arrival (DOA) is sufficient and range is not necessary. For example, when pointing a single auto-focus camera at a talker. In this case, errors in triangulation are reduced by excluding the range and only resolving the source location to azimuth and elevation from the array origins. The camera then uses other means to determine the range and focus settings. But for cost sensitive applications, such as inexpensive consumer electronic equipment the size, cost, and installation overhead are fundamental engineering constraints to be tackled.

2.3 PHASE SIGNATURE APPROACH

The above signal relationships resolve to a manageable and realizable set of equations. They are used in both research and commercial products and seek to avoid rather than include the degradations due to reverberation. This can be observed by disassembling commercial units from Labtec™, Andrea Electronics™, Emkey Innovations™, Acoustic Magic™, Polycom™, and others. The enclosures used suspend the microphones in space and pad them with absorbent material to dampen reverberation effects. Here we will depart from this thinking and look to use the reverberations to resolve range ambiguities. The first step in creating a manageable set of equations is to establish a simple framework when considering the input signal's phase and amplitude that does not need to operate on a set of intersecting hyperboloids. To avoid awkward mathematic solutions encountered by triangulating on a source location from a set of noisy hyperboloid surfaces, DeBiase[14] proposed a method that combines the classic Delay & Sum beamformer with Time Delay Estimation. In his approach, he observed that the phase angles established when applying the

frequency domain TDOA can be compared to the expected phase angles for a TDOA for a single pair of microphones. The maximum of the sum of the phase differences for each location is a Maximum Likelihood function for TDOA. DiBiase labeled this the Steered Power Response-Phase Transform after noting that the GCC-PHAT sharpens the beam around a signal sound source while the Delay & Sum beamformer does a better job of handling reverberation and harsh environments. SRP-PHAT alleviates the need to transform the time difference of arrival into a single time delay τ, by simply finding the location where the pre-calculated phase differences maximize the sum of the phase difference between the location and the GCC-PHAT observation. Stating this in mathematical terms for the two-microphone case we have:

$$P_{12}(\vec{s}) = \int_{-\infty}^{\infty} \Psi_p(\omega) X_{m1}(\omega) X^*_{m2}(\omega) e^{-j\omega(\tau_{12}(\vec{s}))} d\omega \qquad (1.16)$$

where $P_{12}(\vec{s})$ is the sum of phase differences for a given location $\vec{s}$. $\Psi(\omega)$ is the PHAT whitener as before. $\tau_{12}(\vec{s})$ represents a priori calculations for the phase angle of the TDOA between m_1 and m_2 at location $\vec{s}$. The source location is then chosen to be the signal that maximized $P_{12}(\vec{s})$

$$\hat{s} = \underset{s}{\operatorname{argmax}}(P(\vec{s})) \qquad (1.17)$$

The main difference in this method verses the triangulation method is that there is no explicit conversion to a single time delay measurement for every pair of microphones. Instead each GCC-PHAT-whitened phase angle is considered separately before the final value of $P_{12}(\underset{\sim}{s})$ is calculated. For more than two microphones $P_{12}(\vec{s})$ is derived by simply added up all unique combinations of phase difference for all microphone pairs at each location $\vec{s}$. It can be shown that this is equivalent

to a filter and sum beamformer after applying GCC-PHAT less a scalar multiplier and a bias term[14].

For the two-microphone case, this method does not resolve the source location ambiguity found along any hyperboloids defined above. DiBiase solved this problem by adding more microphones. His experiments used anywhere from 15-512 microphones placed against an acoustically treated panel to lessen the effects of reverberation. In the work being presented here, we will use the SRP-PHAT method but take DiBiase's work in a different direction. Rather than add more microphones to sharpen the Direction of Arrival and determine the range, we will add a reverberant microphone enclosure to establish a range distinction previously unavailable with two microphones.

2.4 THE IMAGE MODEL

The above approach uses a simple geometric description in the absence of reverberation and noise to calculate the expected phase values for each microphone pair at each location. In general both the design and analysis of these methods avoid the inclusion of reflections because it greatly distorts $\tau_{12}(\vec{s})$. So, although it does not seek to estimate τ_{12} from the observed input signal, it does rely on the sound propagation model and microphone topology of the source and microphones to determine the proper phase models $e^{-j\omega\tau_{12}(\vec{s})}$ at each $P_{12}(\vec{s})$. To pursue the original premise that reverberation can be exploited we introduce a reflection that complicates the above simple model in order to disambiguate range. We will do this in two stages. First a simple model will be used for analytical purposes. This will show the effects of reflections on the overall math. Secondly real-world measurements and calculations will be made to confirm the predicted effects of reverberation on the GCC-PHAT transform.

To analyze the impact of reverberation we will use the popular sound source Image Model put forth by J. Allen and D. Berkley [2] and later by G. Kendall, W. Martens, and M. Wilde, [24] for simulating acoustical fields. As the name of this method implies, this approach treats reflective surfaces as mirrors for the acoustical wave front. When a point source's acoustical wave front strikes a wall, a mirror image of the signal is created as an 'imaged' point source. The image itself then becomes a point source with in a mirror image of the enclosure and the image model is again applied to the mirror image of the room. This occurs in 3-D and continues until the signal strength is calculated to have attenuated below a predetermined noise floor.

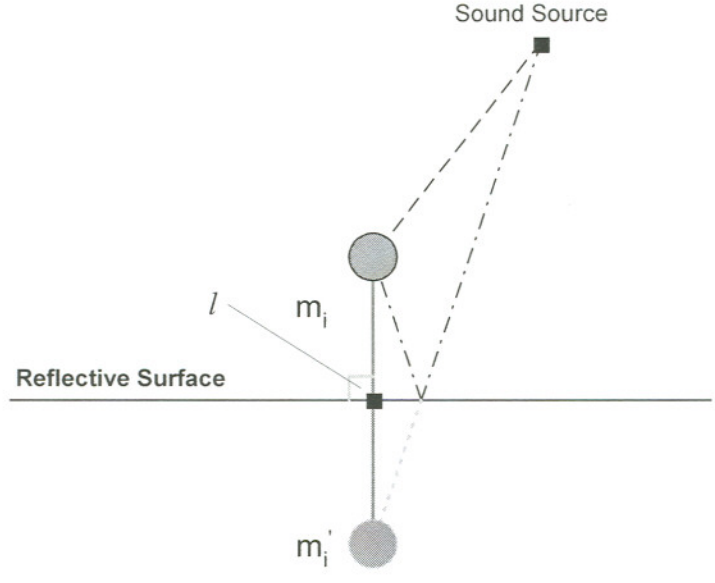


Figure 2.4: Acoustical Image Over A Pure Reflective Surface.

These methods are often used in 3D virtual room simulations. [4]. For now we are only interested in understanding the behaviors of a single reflective surface and can simplify the image model into a simple reflection of the microphone across the plane established by the reflective surface. In practice additional reflections will occur within the enclosure where the sound source

resides. In this geometrical configurations, $G(\omega)$, will still be unknown and possibly time varying. We are only modeling the known reflective surface being intentionally introduced.

For a single reflective surface a simple projection of the microphone's location onto the reflective plane can be determined by the Gram-Schmitt orthogonal projection. This represents the location on the plane closest to $\vec{m}_i$. The reflected image of the microphone can then be created by establishing the perpendicular location on the other side of the surface as twice the error vector between these two vectors. Figure 2.4. The general reflection equation can be stated as

$$\vec{m}_i' = \vec{m}_i + 2\left(\left(\frac{\langle \vec{l}, \vec{m}_i \rangle}{\|\vec{l}\|_2^2}\right)\vec{l} - \vec{m}_i\right) \tag{1.18}$$

Where l is the projection of the physical microphone onto the reflective surface and $\vec{m}_i'$ is the microphone image on the other side of the reflective surface. It can be easily shown that the distance an acoustic wave must travel between the source and the image microphone is identical to the distance between the source to the reflective surface and then to the physical microphone. This does not complete the simulation because the acoustic wave also refracts and attenuates as it travels[15]. The rules for refraction are a function of the reflective surface and become extremely cumbersome to analyze in practice. Fortunately we can select the surface as a design choice. For our purposes we use a hard surfaced floor tile to approximate a refraction of ~0[15] and attenuation factor of ~1. Attenuation will therefore be uniform between the source and both microphones. We only need to apply the inverse square law attenuation rule [17] that states that the intensity of the signal drops off as $\frac{1}{4\pi r^2}$ [4]. Where r is the distance between the source and either the microphone or microphone image. We will make another simplification that the distance between the source and the

microphone is >> the distance between the two microphones. In this way we can maintain a plane wave propagation assumption and avoid the spherical calculations at this stage. We can easily introduce them latter.

The signal at m_i can now be stated as the combined signals for the direct path and the reflected path:

$$m_i(n) = \frac{1}{4\pi|\vec{s}-\vec{m}_i|} s\left(n - \frac{|\vec{s}-\vec{m}_i|}{C}\right) + \frac{1}{4\pi|\vec{s}-\vec{m}_i'|} s\left(n - \frac{|\vec{s}-\vec{m}_i'|}{C}\right) \tag{1.19}$$

Where again C is the speed of sound and ' means microphone reflected image. This model can now be applied to our situation to produce the signal at each microphone.

2.5 ANALITICAL INTERPRETATION OF REFLECTION INFLUENCES

Combining the reflective surface into the SRP-PHAT can be interpreted as establishing a unique phase signature for each source location. By doing so, our intent to establish range resolution along the hyperboloid surface of the microphone pair can then be studied. This analysis will maintain the following simplifications: the transfer function between the source and microphone is simple attenuation of a single impulse function, the reflective surface does not refract the signal, there is no additive noise, and the signal is a plane wave. These simplifications are justified by showing that the analysis predicts real world measurements described in later sections.

Using the configuration from Figure 2.1 we have that a sound source $s(n)$ propagates to microphones m_1 and m_2. Assigning $A_s e^{-jwn}$ to $s(n)$ we can define the phase signature in terms of the microphone and reflective surface geometries.

Each microphone signal then becomes

$$m_i(n) = A_i e^{j\omega(n+\tau_i)} + A_i' e^{j\omega(n+\tau_i')} \tag{1.20}$$

where $A_i = \frac{A_s}{4\pi|\vec{s} - \vec{m}_i|}$, and τ_i, τ_i' are the time of propagation between the sound source and the microphone and the microphone image respectively.

The phase signature $\Phi(\omega)$ can be stated by taking the difference in the phase between two microphones.

$$m_i(t) = A_i e^{j\omega t} e^{j\omega\tau_i} + A_i' e^{j\omega t} e^{j\omega\tau_i'} \tag{1.21}$$

$$m_i(t) = e^{j\omega t}\left(A_i e^{j\omega\tau_i} + A_i' e^{j\omega\tau_i'}\right) \tag{1.22}$$

$$= e^{j\omega t}\left[A_i\left(\cos(\omega\tau_i) + j\sin(\omega\tau_i)\right) + A_i'\left(\cos(\omega\tau_i') + j\sin(\omega\tau_i')\right)\right] \tag{1.23}$$

Collecting the real and imaginary parts and solving for the microphone phase yields

$$\phi_i(\omega) = \arctan\left[\frac{A_i\sin(\omega\tau_i) + A_i'\sin(\omega\tau_i')}{A_i\cos(\omega\tau_i) + A_i'\cos(\omega\tau_i')}\right] \tag{1.24}$$

Stating in terms of our known topology yields

$$\phi_i(\omega) = \arctan\left[\frac{\frac{A_s}{4\pi}\left[\left(\left|\vec{s}-\vec{m}_i\right|^{-1}\right)\sin\left(\omega\frac{\left|\vec{s}-\vec{m}_i\right|}{C}\right)+\left(\left|\vec{s}-\vec{m}_i'\right|^{-1}\right)\sin\left(\omega\frac{\left|\vec{s}-\vec{m}_i'\right|}{C}\right)\right]}{\frac{A_s}{4\pi}\left[\left(\left|\vec{s}-\vec{m}_i\right|^{-1}\right)\cos\left(\omega\frac{\left|\vec{s}-\vec{m}_i\right|}{C}\right)+\left(\left|\vec{s}-\vec{m}_i'\right|^{-1}\right)\cos\left(\omega\frac{\left|\vec{s}-\vec{m}_i'\right|}{C}\right)\right]}\right] \quad (1.25)$$

Where again C is the speed of sound. The $\frac{A_s}{4\pi}$ cancels showing that the phase is not a function of the original source's amplitude. This is necessary for this method to function. Had the introduction of a reflective surface caused the phase difference to be a function of the original signal characteristics then this method would not work because we have no knowledge of the sound source's original amplitude. If the numerator/denominator < 0 then $\phi_i(\omega) = \phi_i(\omega) + \pi$. The phase signature is then created for any two microphone pairs as

$$\Phi_{ij}(\omega) = \phi_i(\omega) - \phi_j(\omega), \quad i \neq j \quad (1.26)$$

In our case we are only interested in two microphones so i=1 and j=2 which leaves us with a single phase signature.

As a check, for the non-reflective case we can drop the image microphone in (1.23) which reduces equation (1.25) to

$$\phi_i(\omega) = \arctan\left[\frac{\left(\left|\vec{s}-\vec{m}_i\right|^{-2}\right)\sin\left(\omega\frac{\left|\vec{s}-\vec{m}_i\right|}{C}\right)}{\left(\left|\vec{s}-\vec{m}_i\right|^{-2}\right)\cos\left(\omega\frac{\left|\vec{s}-\vec{m}_i\right|}{C}\right)}\right] \quad (1.27)$$

Which further reduces to

$$\Phi(\omega) = \frac{\omega\left(\left|\vec{s} - \vec{m}_1\right| - \left|\vec{s} - \vec{m}_2\right|\right)}{C} \tag{1.28}$$

Upon further inspection we see equation (1.28) defines the phases of the hyperbola

$$\frac{\Phi(\omega)C}{\omega} = \sqrt{\left(s_x - m_{1x}\right)^2 + \left(s_y - m_{1y}\right)^2 + \left(s_z - m_{1z}\right)^2} - \sqrt{\left(s_x - m_{2x}\right)^2 + \left(s_y - m_{2y}\right)^2 + \left(s_z - m_{2z}\right)^2} \tag{1.29}$$

Which, as expected, is identical to equation (1.13) derived from the geometries. However, in the case of the reflective surface this does not define a hyperbola. With careful placement of the reflective surface we can establish an arrangement that will cause a change in the phase difference even as the sound source travels along the surface of any direct path hyperbola.

3 PRACTICAL

Four steps were taken to determine the viability of this method. The first step established a plausible physical arrangement for the sound source, microphone, and reflective surface after imposing practical geometrical constraints on the overall setup. The second step established a real-time signal processing framework for on-line performance analysis and real-time inspection of internal system parameters. The third step measured the performance of this method using the image model's estimated phase signatures. A few discrepancies are shown between the image model and real-world signals that are resolved in the forth step by using acoustically measured phase signatures. The phase signature method is then used to determine the influences of a reflective surface to discriminate range along the lines of ambiguity.

3.1 MODEL ESTIMATION UNDER REAL-WORLD CONSTRAINTS

Before moving immediately to live audio signals a reasonable set of real-world physical constraints are imposed on the above analysis to verify that phase signatures are still discernable. For this the image model is used to establish the phase signatures. In both the reflective and non-reflected case equation (1.26) can be used with the microphone, reflective surface, and source locations known. The variables are:

- The location of the sound source with respect to the microphones.
- The signal characteristics of the sound source.
- The microphone spacing.

- The microphone sensitivity characteristics.
- The reflective surface characteristics.

As a first order of magnitude constraint the following choices were made: The sound source will reside within 5 feet of the microphones because we are looking for a distinction between sound directed at a device within a reasonable usage distance and those beyond that range. The sound source will be placed along broadside, end-fire, and diagonal surfaces to measure range along regions of ambiguity. The signal will cover the entire frequency spectrum of speech so we can see the entire signature at once. The reflective surface can be modeled by a simple reflection to match the Image Model assumption. The reflective surface and microphone will be placed in a way that suggests a credible enclosure for a home appliance or computer screen that may need to capture sound. The microphones will be omni directional to match the image model assumption.

These parameters were used in a MATLAB™ simulation. The reflective surface was assumed to be an infinite plane and placed in various locations with respect to the microphones. Microphone spacing was varied between 8 and 4 inches. A complete search of this space was not performed. Visual inspection and simple analysis of the phase signatures after imposing these limitations show that all produce similar results. The need to establish a method to maximize the spreading of the phase signatures was noted, however, the choice was to use a reasonable geometrical configuration to confirm the original premise rather than establish the optimal reflective surface and arrangement. The MATLAB simulation suggested that a microphone spacing of approximately 6 inches and a reflective surface that does not present symmetries with the microphone axis produces phase differences along regions of highest range ambiguity.

A set of 15 sound source locations on a single horizontal plane was established. The locations were arranged in a 48"x48" square with locations every 16 inches satisfies the above requirements. Figure 3.1 show location (32, 00) and (16', 48') respectively. X is the horizontal axis and Z is the vertical axis. The sloped green line is the reflective surface. The two magenta circles represent the microphones. The dotted lines show paths from the red-star true sound source and from the green-star 'imaged' sound source. (Equivalently, we could have reflected the microphones over the reflective plane as shown in the analytical section.)

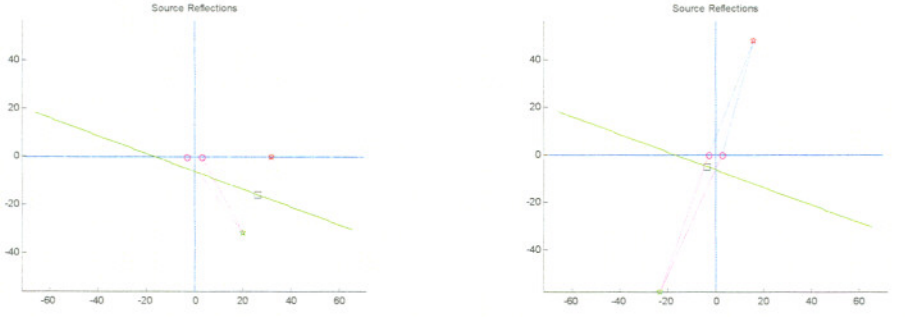
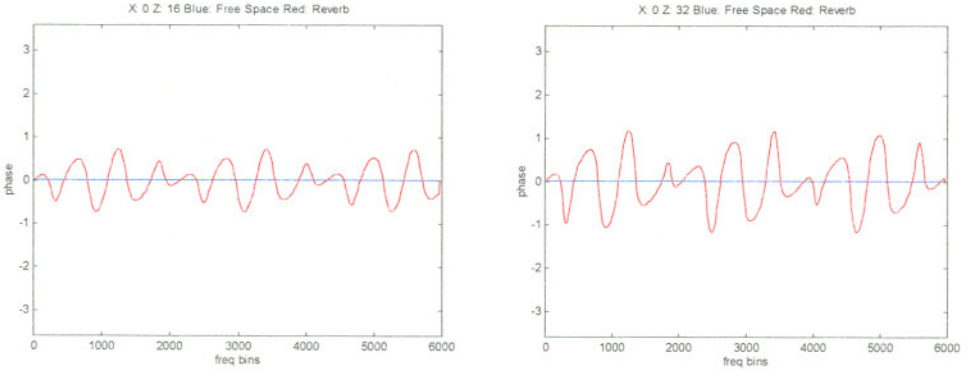
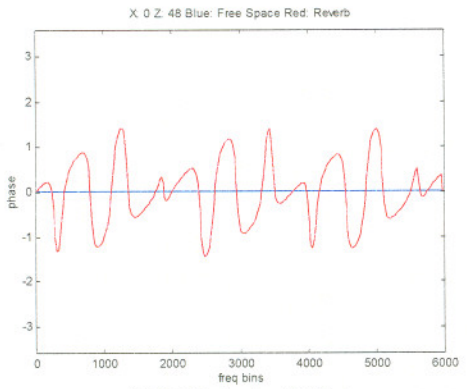
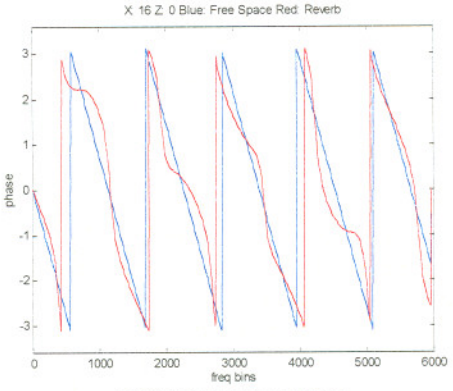


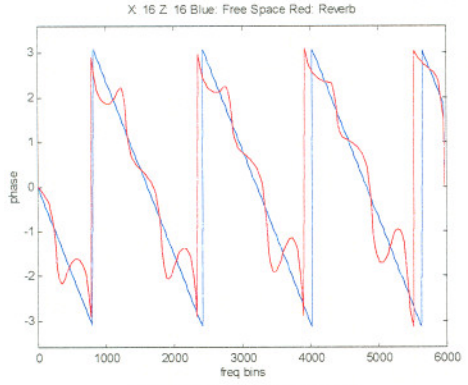
Figure 3.1: Simulation of Sound Source Imaging over Reflective Surface. Red Star is original sound source, green star is sound image over green line reflective surface. Magenta circles are microphones. All units are in inches. Black box is reflection point for source.

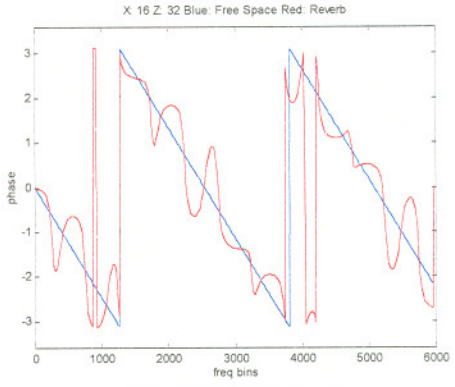
The phase signatures produced by this configuration for the set of 15 source locations is displayed in the following pictures.

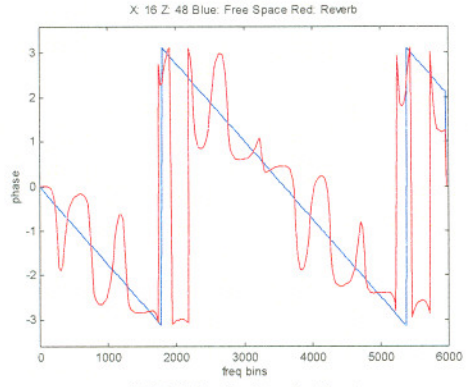


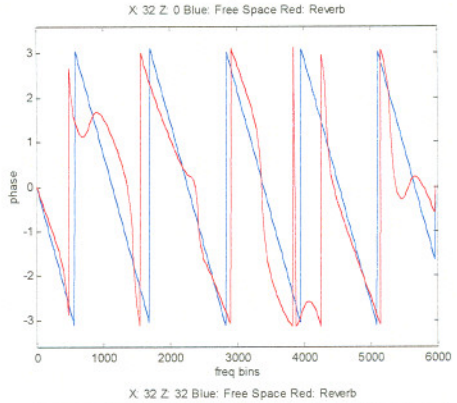

X: 0 Z: 48 Blue: Free Space Red: Reverb
phase
freq bins

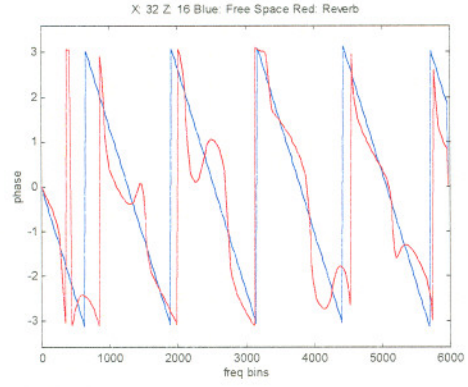

X: 16 Z: 0 Blue: Free Space Red: Reverb
phase
freq bins

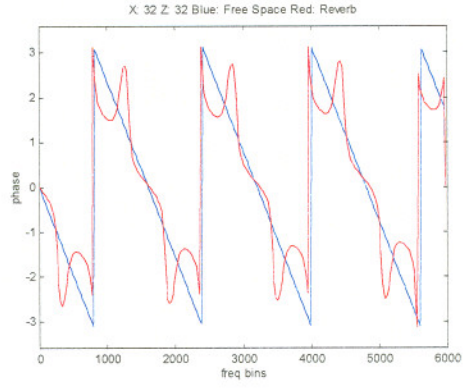

X: 16 Z: 16 Blue: Free Space Red: Reverb
phase
freq bins


X: 16 Z: 32 Blue: Free Space Red: Reverb
phase
freq bins


X: 16 Z: 48 Blue: Free Space Red: Reverb
phase
freq bins


X: 32 Z: 0 Blue: Free Space Red: Reverb
phase
freq bins


X: 32 Z: 16 Blue: Free Space Red: Reverb
phase
freq bins


X: 32 Z: 32 Blue: Free Space Red: Reverb
phase
freq bins

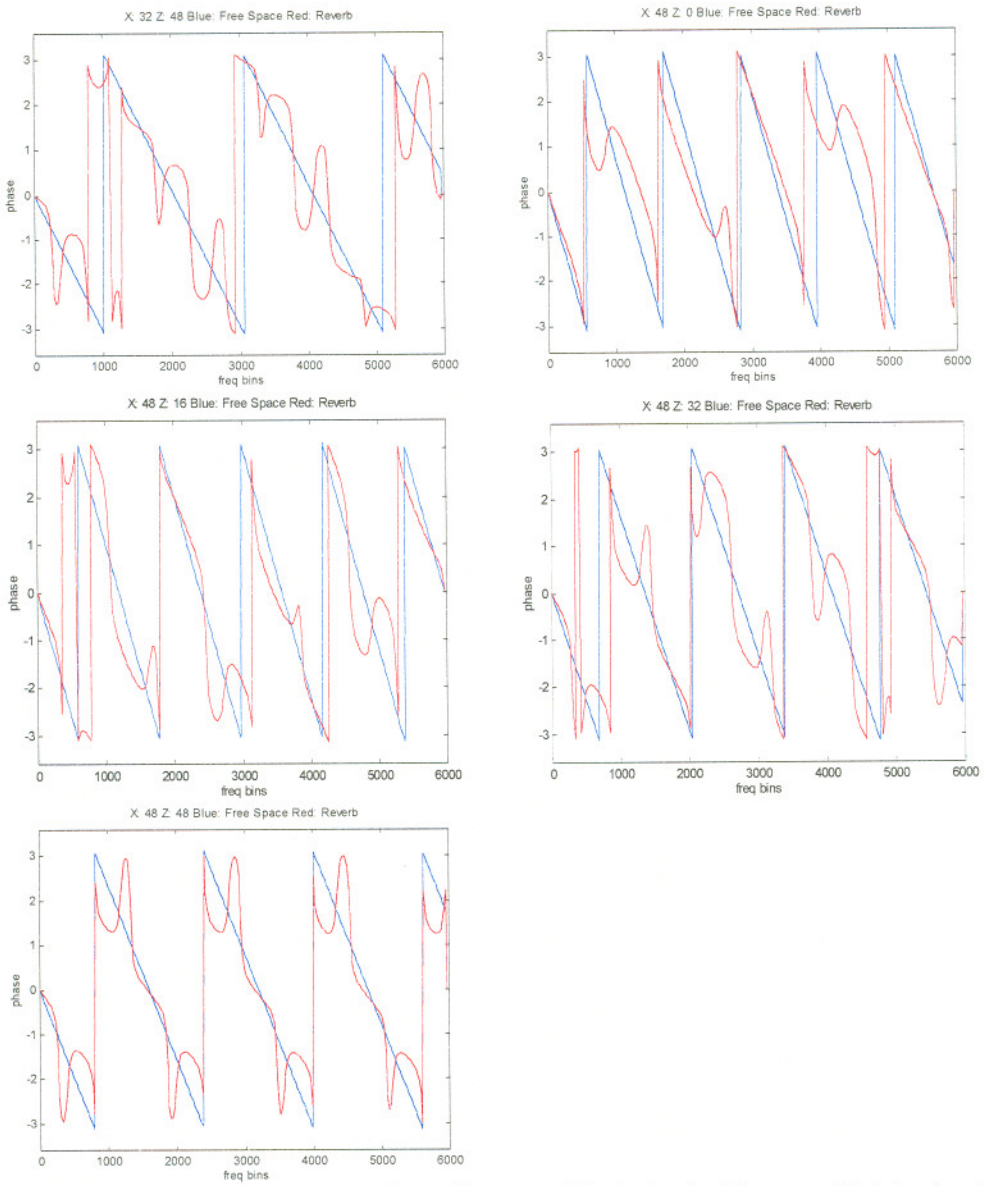


Figure 3.2: Phase Signature for Selected Testing Locations. Blue is simulated free-space. Red simulated Reflective Surface.

In these pictures the horizontal axis is frequency and vertical is wrapped phase. Blue is derived from the non-reverberant model while red is the reverberant model. We can see that the phase signature is disturbed by the reflective surface. This demonstrates the potential to modifying the

phase difference within the desired frequency range even with this small arrangement. This model is initially used for analyzing performance in the real-system.

3.2 SET UP & ACOUSTICAL ENVIRONMENT

Experiments were carried out in a 15'x16'x9' carpeted lab. Three sides of the lab contained linoleum desktops with mild acoustical absorbent back-panels reaching 5.5' in height. Starting above the cubicle panels and reaching the ceiling were additional acoustical panels. Under these conditions, the room has a RT60 rating of less than 94msec. The panels do not reduce the isotropic noise, which registers approximately 65dBC as measured with a handheld dB meter. This is still modestly high because of the HVAC ducts in the ceiling.

The phase signatures and the algorithm performance were measured at 15 preset locations covering a 48"x48" horizontally configured square located approximately 34" above the floor of the lab. (See Figure 3.3). The locations were placed at the intersections of each 16" grid. This set of locations provided three sets of three locations that move along a single hyperbola as well as six locations that do not. The microphones were placed 6 inches apart and suspended 5.5 inches above the counter with wire mesh. The mesh was ½" square fencing grid which will not impact the acoustical wave significantly because is dimensions are << that the wavelength of the highest sound. The entire setup was placed on top of an acoustic panel to avoid measuring reflections from the countertop. (See Figure 3.7). The reflective surface was created by a 13"x13" ceramic floor tile. Looking broadside to the microphone axis, the ceramic tile was placed behind the microphones with approximately equal span above and below the microphone pair. The specific dimensions are shown in detail in Figure 3.4. The reflective tile was held up from behind and its exact location was outlined with magic marker on the supporting acoustic panel for easy removal and replacement. Two

additional acoustical panels were placed behind the setup in the –x,-z quadrant to lessen the impact of room reverberation in both non-reflective and reflective configurations.

Taking the model observations as general configuration rules, and considering the limited size of the hard-walled lab being used in these experiments, the dimensions in Figure 3.3 and 3.4 were selected for the array and the 15 locations.

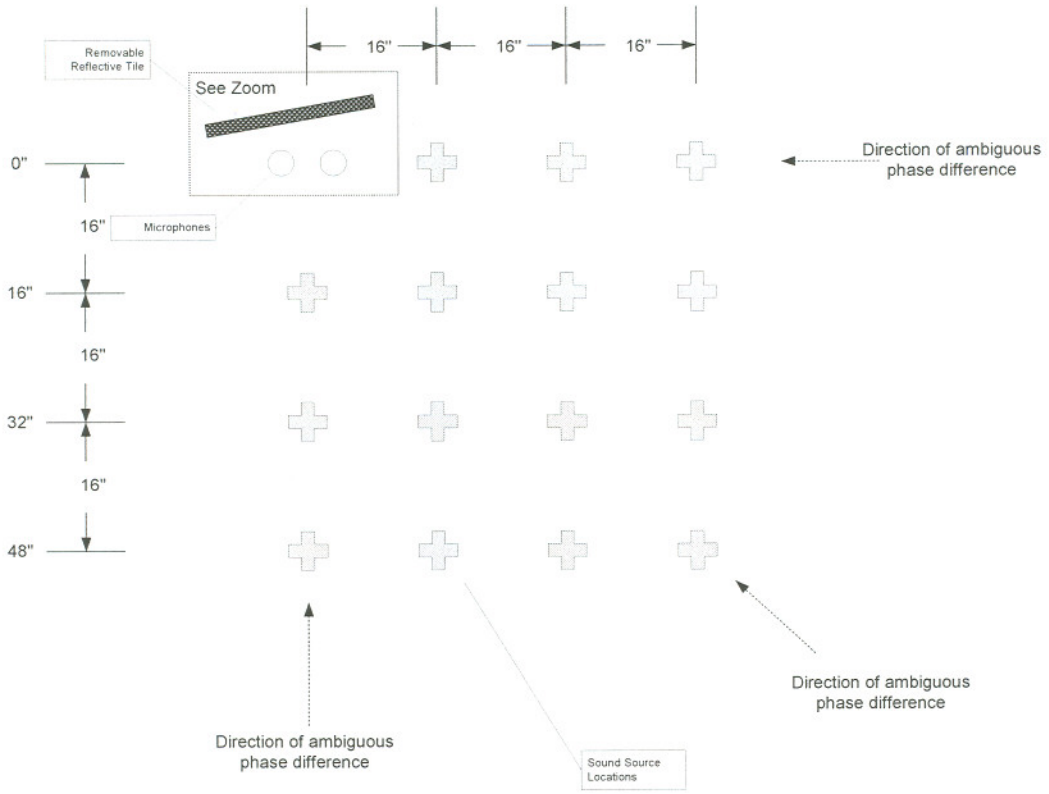


Figure 3.3: Test Configuration. The crosses represent the locations for testing. Lines of ambiguity are noted by perimeter arrows pointing directly at the microphone pair.

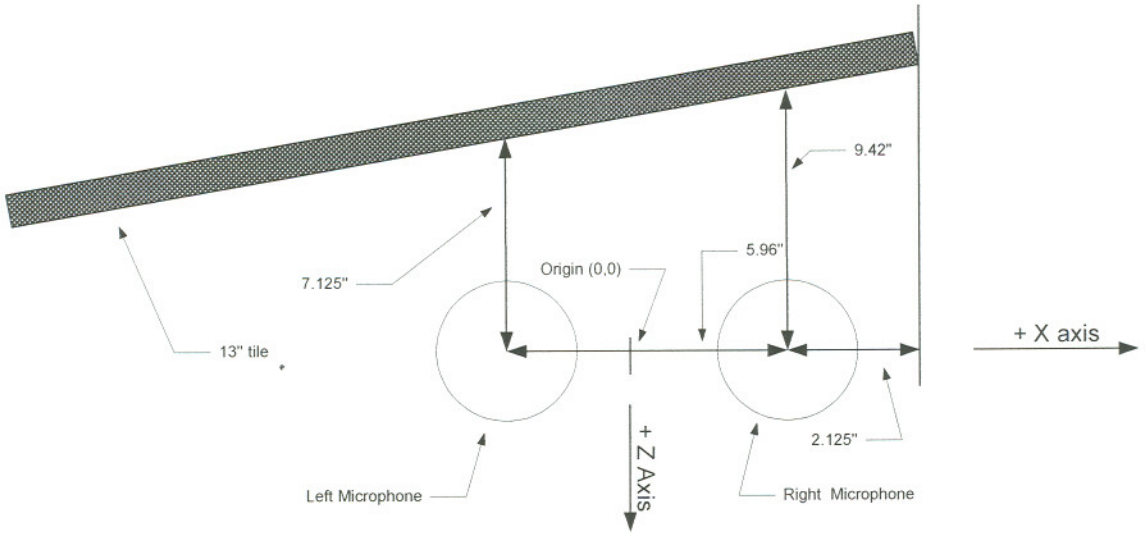


Figure 3.4: Close up of microphone and reflective surface configuration

To increase the ease of exploring a variety of enclosure shapes and surfaces a real-time test and measurement system was built. The real-time system was capable of capturing live audio, processing the phase signature algorithm, displaying the calculated phase differences, logging internal algorithm data, and importing simulated phase models for testing. The system was driven by an 866Mhz Pentium 4™ processor running National Instrument LabVIEW™ Virtual Instrument software under Windows 2000 (Figure 3.5 and 3.6). Audio I/O was established by an ASIO 2.0 full duplex driver feeding an RME-Audio Digi9652™ optical ADAT PCI add-in card. Two channels were used for input and one channel used for output. All signal processing was performed in real-time using Intel's optimized signal processing libraries. The optical I/O was fed into a Creamware A16™ 16 channel ADC/DAC. The microphones used were generic Panasonic condenser microphones attached to a custom DC Bias circuit. The microphone level was amplified with a Mackie™ adjustable pre-amp. Output test and measurement signals were generated by a high-current amplifier and B&W bookshelf loudspeakers.

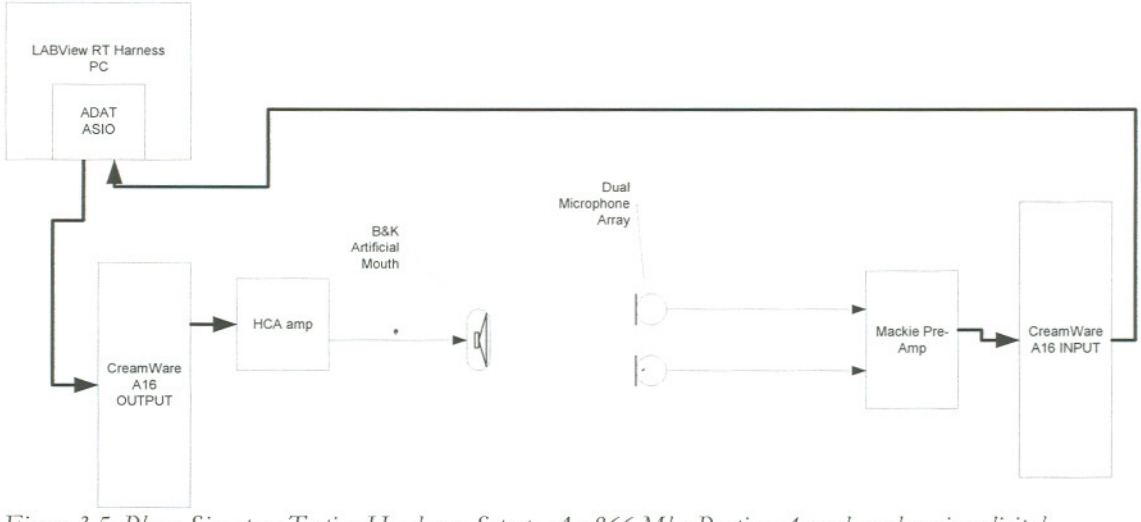


Figure 3.5: Phase Signature Testing Hardware Set up. An 866 Mhz Pentium 4 sends and receives digital audio attached to an amplifier and two microphones.

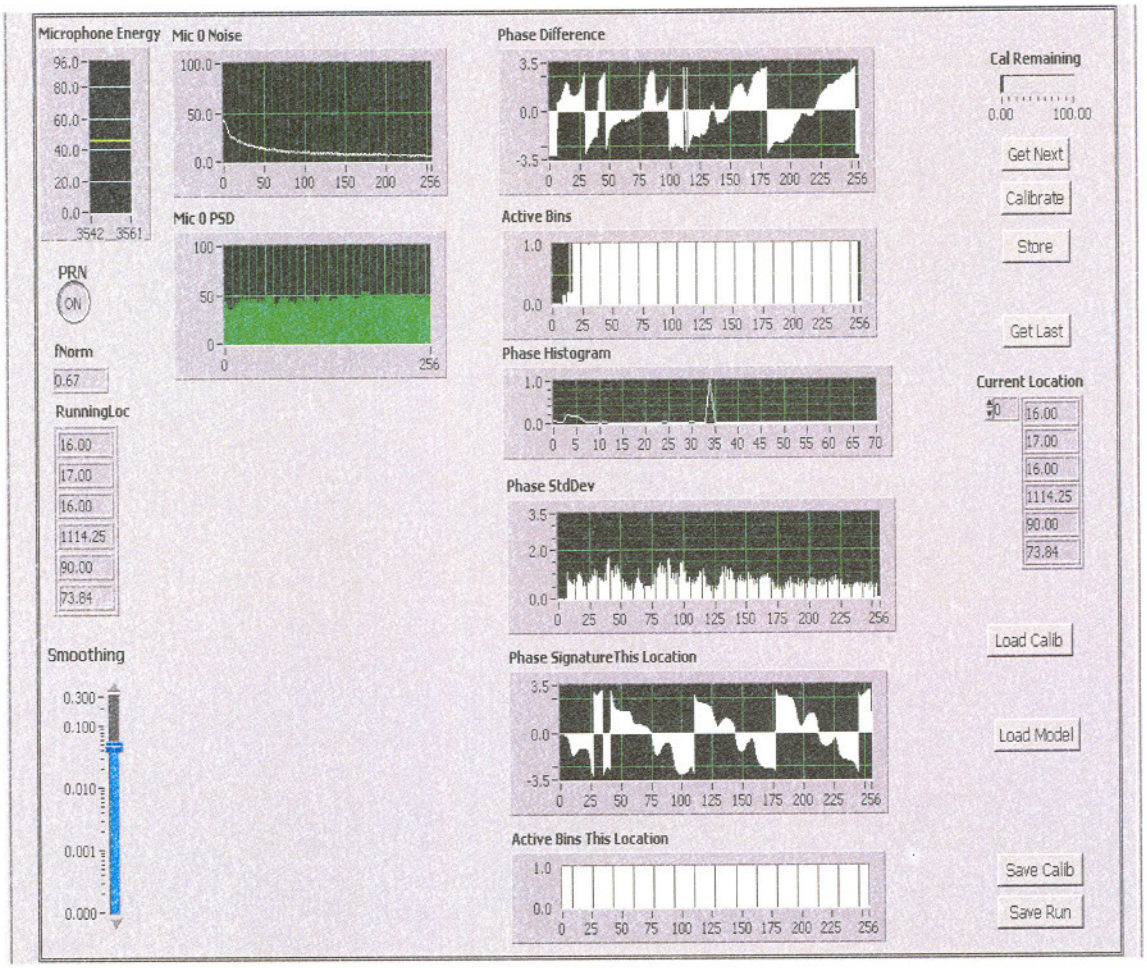


Figure 3.6. LabView™ real-time GUI. This application allows for real-time control and display of the Phase Signature Algorithm. From Left to Right, the column displays the mi microphone energy in dB, signal output on/off button, argmax(P(s)) and s. A real-time smoother for the noise floor. Second to Left column is the Noise Floor Energy and microphone cross-spectrum in dB. The Third from left column contains the live phase signature, the active bins above 20 dB, a single bin phase histogram for debugging, the Phase Signature StdDev, the modeled phase signature, and the model phase signature active bin mask. The Right most column contains four buttons to move, calibrate, and store phase signatures, the current signature in the Phase Signature This Location window, and a set of buttons to load and store Models and P(s) results.

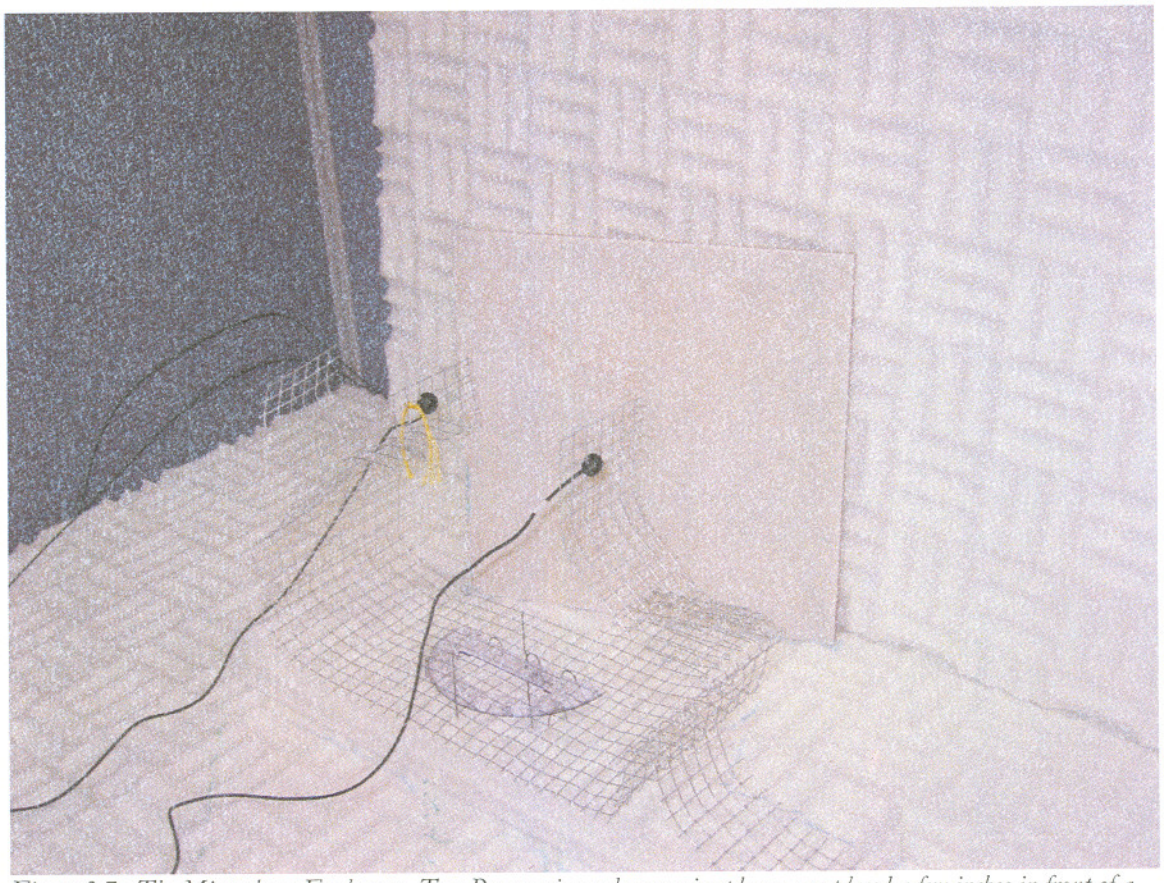


Figure 3.7: The Microphone Enclosure. Two Panasonic condenser microphones are placed a few inches in front of a reflective surface during reverberation testing. Wire mesh suspends the microphones several inches above the table. Acoustic panels are used to isolate the effects of the reflective surface when comparing to the free space configuration.

3.3 ALGORITHM REALIZATION & PRACTICAL ENHANCEMENTS

To process real-time audio input the standard practices were used. The sampling rate of the Creamware™ A16 was fixed at 48kilosamples per second. The sampling size was 16 bit twos complement. The DC-offset was removed by a high-pass filter with a cut off just above 60Hz. The signal was then low-pass-filtered and down-sampled from 48kps to 24kps. Both filters were implemented as FIRs to maintain linear phase. Since the group delay through all filters was the same

for both microphones, no additional time adjustments were needed. A 512-point FFT was used with a 50% overlapping window providing 21.3 msec buffers to the algorithm. Each frequency bin is 46.875Hz wide. The standard 512 point Hanning window was used to lessen the edge effects of the rectangular buffering window during the Fourier Transform. The noise floor was calculated after this filtering to match the gain stages in all signals being compared. The following block diagram (Figure 3.8) shows the stages from the theoretical derivation along with additional components needed to handle real-world signals.

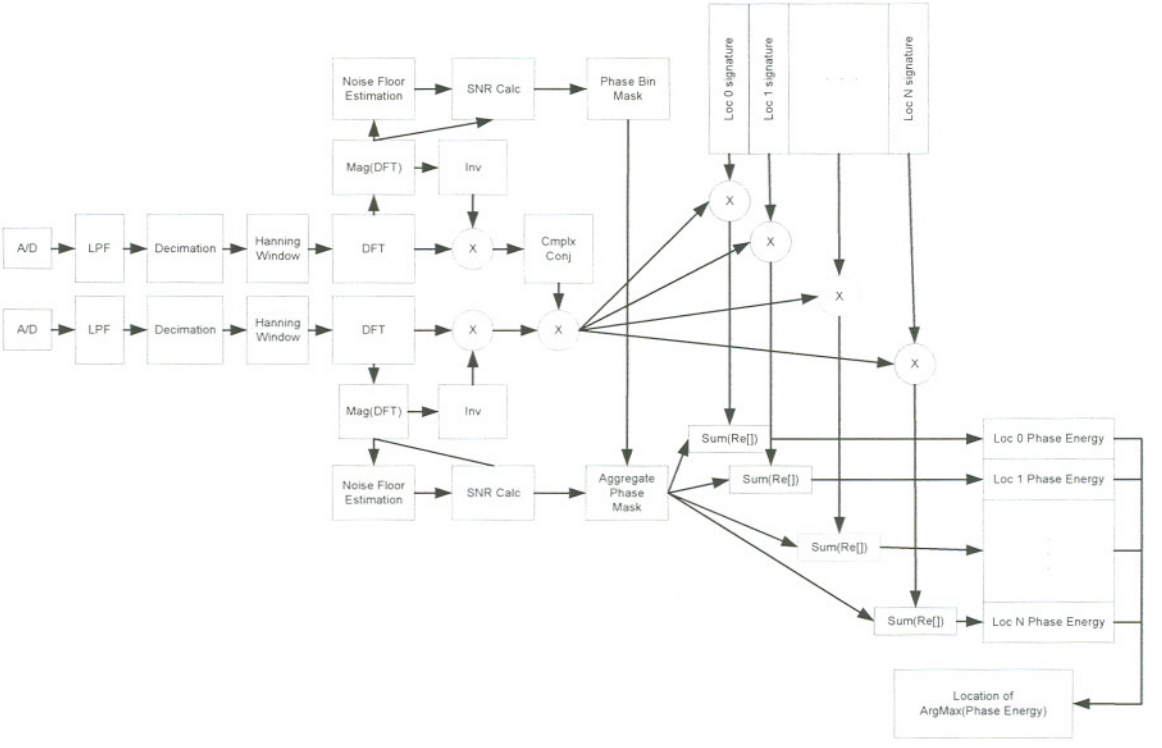


Figure 3.8: Phase Signature Block Diagram.

To make measurements on real acoustic data several practical enhancements were introduced to handle the effects of stationary noise, low signal strength, real-time streaming, and phase variations. First the noise floor power spectral density was estimated after the initial filtering stages. The estimation as placed after the initial filters so they match the gain stages in all signals being compared.

This was achieved by applying long term averaging to the power spectral density during the first 20 seconds of a trial. This noise floor was then frozen for the duration of the data collection. The power in each frequency bin of the phase difference was then compared to this noise floor. All bins registering 20dB above the noise floor were considered to contain valid phase information. All frequency bins below this threshold were not used in the calculation of that frame. It was possible that the room's HVAC system could turn on or off at any time and change the noise floor. This was monitored and fortunately never occurred during data collection.

To avoid erroneous conclusions being made from only a small number of phase readings 20dB above the noise floor, an additional threshold was placed on the number of bins that must be above the noise for a single reading. This level was set imperially at 25% of all FFT bins and needs to be reconsidered when discriminating for human speech. The phase bins that passed this criterion are then used in $X_i(\omega)$ calculations.

For each test a uniformly distributed pseudo-random noise (PRN) source was used. The output level at 1 foot from the loudspeaker was set to 72dBC averaged over 2 seconds. This signal was selected to cover nearly the full frequency spectrum of the sampled signal. The level was measured at 1 foot and not at the microphones. Thus the signal at the microphone varies depending on the location and is predicted not to matter. Informal testing at various levels suggested this is true so it was removed as a variable.

The real-time system was used to load, calibrate, and save the phase and location information. For each location the PRN was played and the phase signature was calculated. This was done for either creating a new signature for this location or comparing against the currently loaded model. In addition to the phase signature, a Boolean flag was stored to indicate which frequency bins had valid phases and which did not receive enough signal strength. If either the phase signature model

frequency bin or the input phase difference frequency bin did not have a valid flag, then that bin was discarded. This turns out to be necessary because several of the low frequency bins (<280Hz) could not achieve a valid SNR during calibration at the farther distances.

A forgetting factor with a time constant of 4 seconds was used to allow tracking of a moving sound source. Lastly, the mean and variance of $P(\vec{s})$ over all locations were recorded into a log file. These enhancements made the testing of the phase signatures consistently operate in real-time on any signal. As a quick informal check of these enhancements a variety of sound source signals were tested for proper functioning. Talking, clapping, shaking keys, and two versions of PRN signals were used to confirm that the masking and filtering functions operated in a reasonable fashion.. Formal testing then proceeded using a PRN signal.

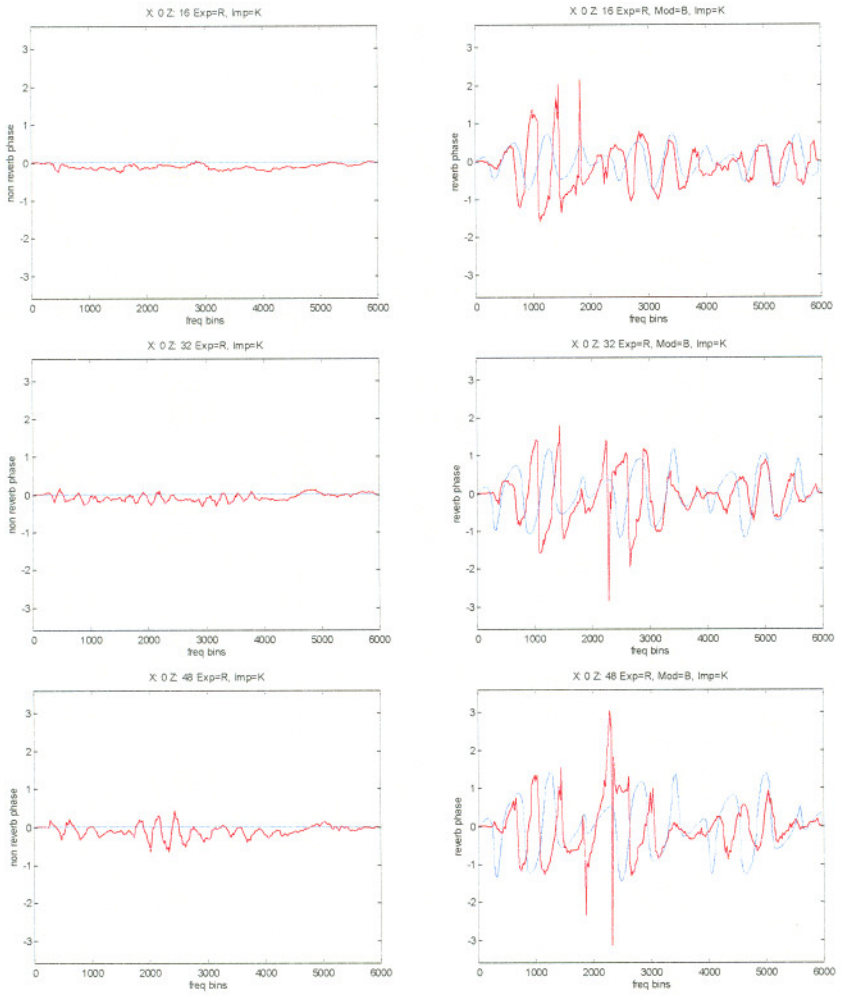
3.4 EXPERIMENTS

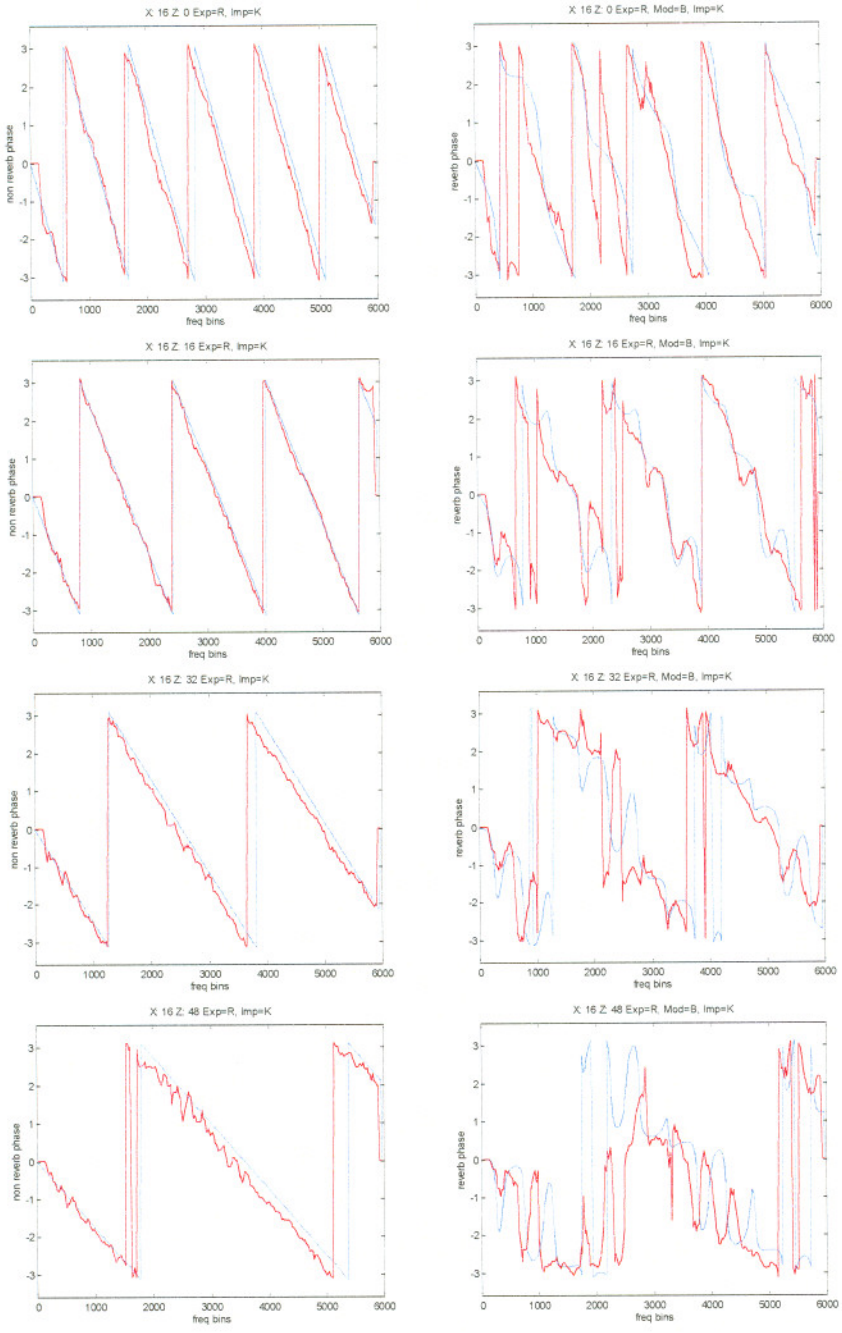
Two main experiments were performed. One used the phase signatures from the Image Model and the other from acoustical measurements. The first, using the Image Model outputs, determined rather quickly that though the simple image model demonstrates the basic behavior of the phase signature it does not represent a few characteristics of the true real-world signature accurately enough for this algorithm to function properly in all locations. Next the true phase signature is measured using audio signals and an accurate model is selected. This model is loaded and the testing results are logged and analyzed.

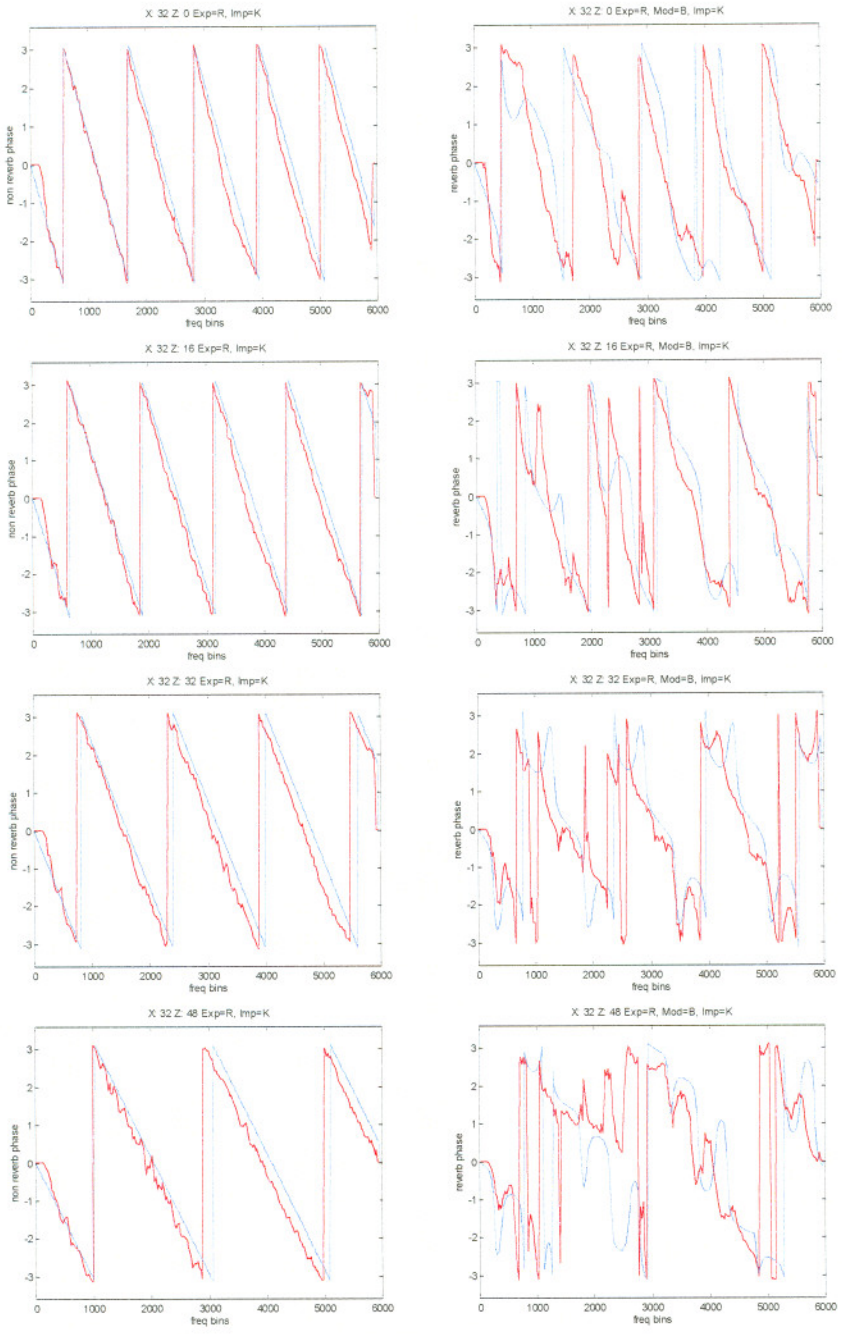
3.4.1 HAND MEASURED IMAGE MODEL

The first experiment uses the predicted phase signatures from the Image Model in the real-time demonstrator. The arrangement was carefully measured by hand and entered into the MATLAB

simulation. The phase angles and bin masks were used during live testing. The test was carried out as described in the set up section above with $P(\vec{s})$ and its variance recorded for each location. The phase models used verses the actual phase signature received are shown below. The left column shows the non-reflective case and the right column shows the reflective case. For all pictures the model is in dashed blue and the measured phase signature in solid red.




X 16 Z 0 Exp=R, Imp=K
non reverb phase
freq bins
X 16 Z 0 Exp=R, Mod=B, Imp=K
reverb phase
freq bins
X 16 Z 16 Exp=R, Imp=K
non reverb phase
freq bins
X 16 Z 16 Exp=R, Mod=B, Imp=K
reverb phase
freq bins
X 16 Z 32 Exp=R, Imp=K
non reverb phase
freq bins
X 16 Z 32 Exp=R, Mod=B, Imp=K
reverb phase
freq bins
X 16 Z 48 Exp=R, Imp=K
non reverb phase
freq bins
X 16 Z 48 Exp=R, Mod=B, Imp=K
reverb phase
freq bins


X 32 Z 0 Exp=R, Imp=K
X 32 Z 0 Exp=R, Mod=B, Imp=K
X 32 Z 16 Exp=R, Imp=K
X 32 Z 16 Exp=R, Mod=B, Imp=K
X 32 Z 32 Exp=R, Imp=K
X 32 Z 32 Exp=R, Mod=B, Imp=K
X 32 Z 48 Exp=R, Imp=K
X 32 Z 48 Exp=R, Mod=B, Imp=K
non reverb phase
reverb phase
freq bins

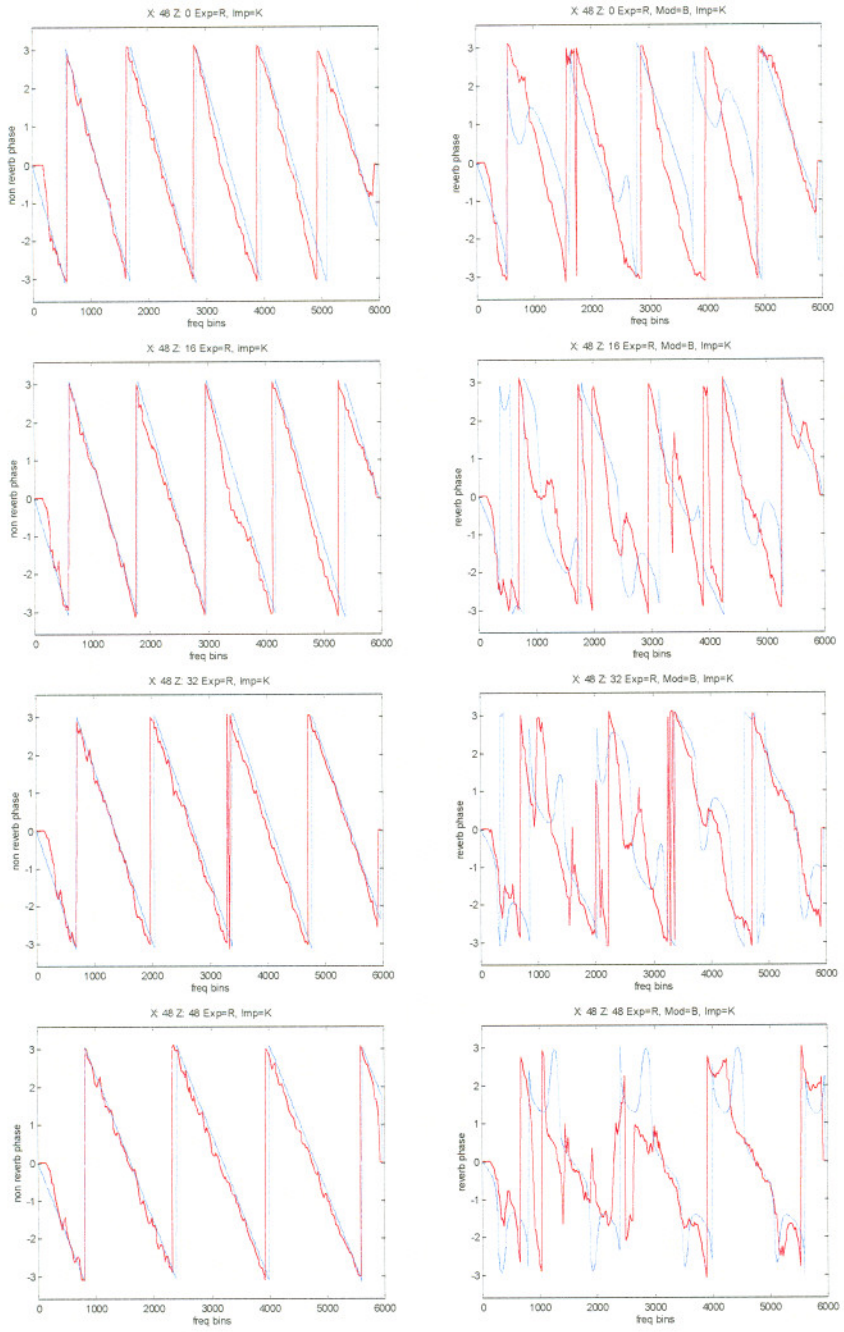


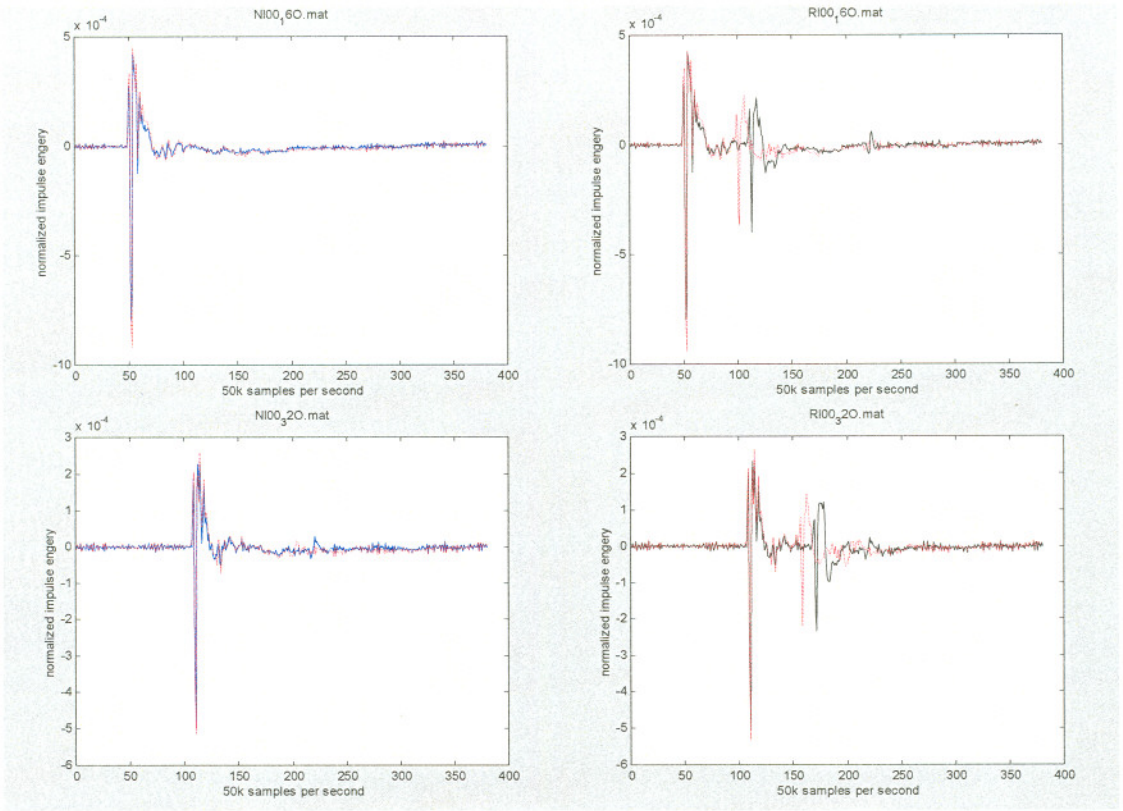
Figure 3.9: Hand Measured Phase Signatures vs Actual. Dashed Blue is model, Solid Red is actual.

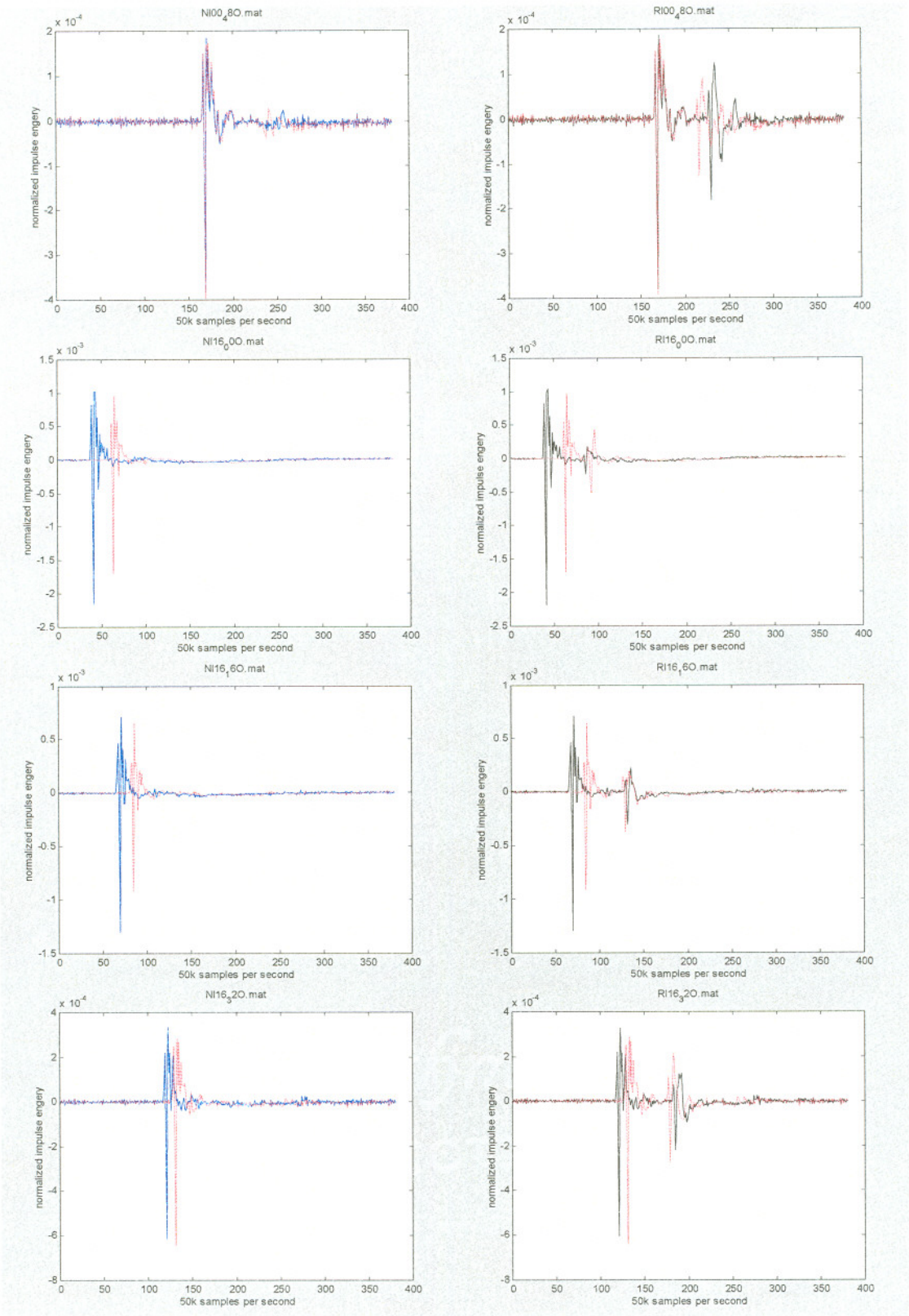
With these phase difference models a handful of locations were measured for $P(\vec{s})$. The maximum values achieved were under 0.6 for the reverberant case and the non-reverberant models scored higher during the reverberant configuration. It was obvious by viewing the real-time screen display of the phase signatures that the model was not accurately describing the input signal. We can see from inspection of the overlay graphs in Figure 3.9 that, although the reverberant model comes up with a similar representation, it does not track the measured phase angles closely in all positions. This can be seen clearly in the (0,16), (0,32), (16,48), (32,48) cases. Testing was stopped and it was decided to look for a better model because the intent of the premise was to focus on the ability to use reflections to disambiguate range and not to focus on building accurate acoustic models for small enclosures. A better understanding of the discrepancy between the model and real-world signal was sought.

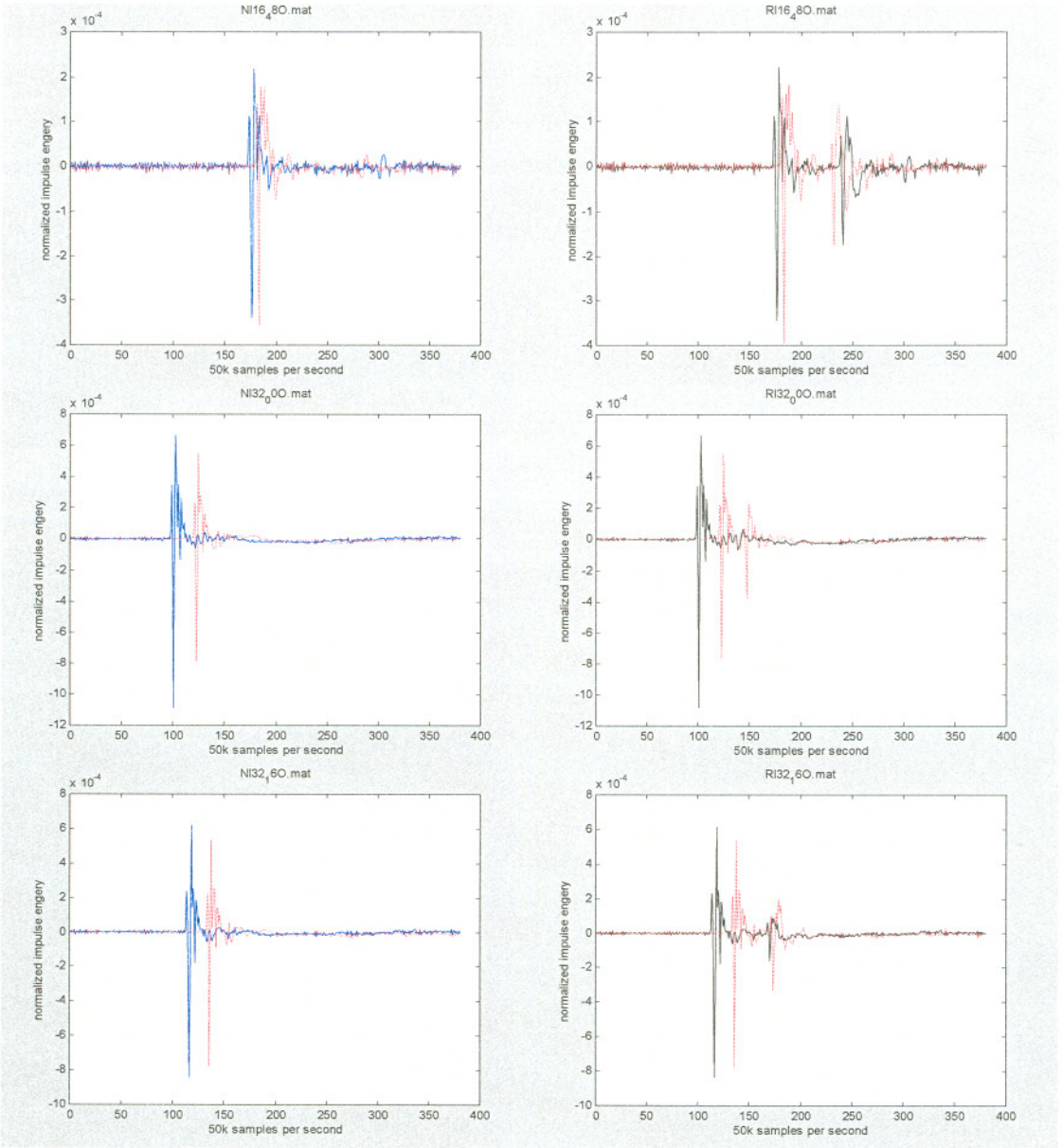
3.4.2 IMPULSE RESPONSE INSPECTION

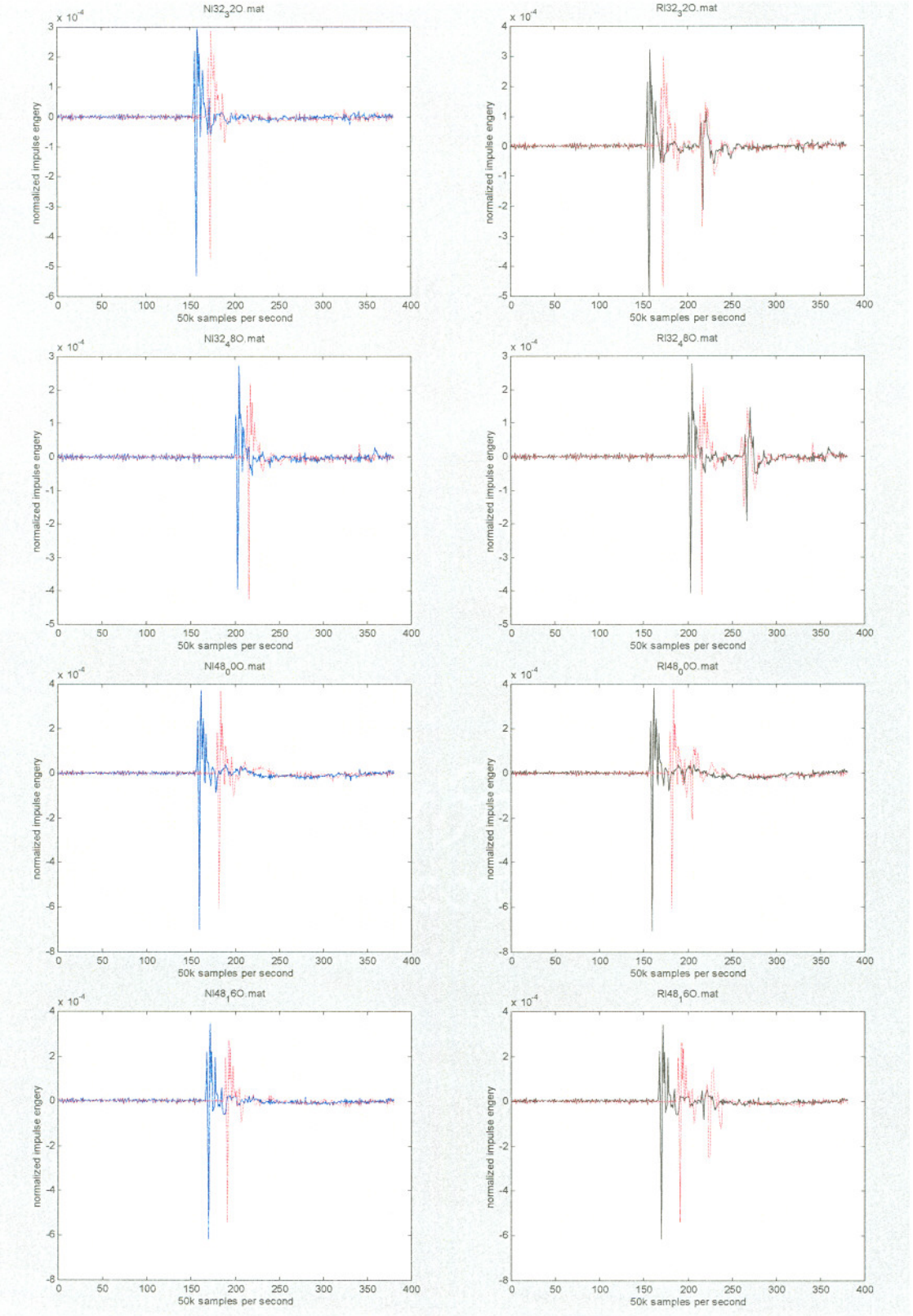
There are a number of sources for errors in the Image Model method. This includes: errors in hand measuring the configuration with a tape measure, assuming plane wave propagation rather than spherical, modeling the tile surface as a pure reflection, and assuming the tile has infinite size. To shed light on the source of the error the impulse response of the reverberant tile was first measured. The loudspeaker was placed at one of the 15 locations. One microphone was recorded at a time using the Maximum-Length Sequence(MLS) [13][37] to stimulate the room. The signal was sampled at 50Hz and correlated with the original sequence using the Least Means Squared method.[21][39] This produced the time series impulse for that microphone given a sound source at that location $(G_i(\vec{s}))$. The reflective tile was then introduced and the impulse was measured again. This procedure was repeated for the other microphone. Doing so for all locations created two sets of impulses, one set for the microphone pair without the reflector and one set for the microphone pair

with the reflector. Next, the impulse responses were scanned for the direct path and reflected path signal time of arrival. For the non-reflective case only the direct path impulse was used. The time delays between the signals arriving at the right microphone were subtracted from those arriving at the left microphone. The following graphs of the impulse responses show how the reflector affects the impulse response as predicted by the Image Model.




NI00_48O.mat
RI00_48O.mat
NI16_00O.mat
RI16_00O.mat
NI16_16O.mat
RI16_16O.mat
NI16_32O.mat
RI16_32O.mat
normalized impulse engery
50k samples per second


NI16_4O.mat
RI16_4O.mat
NI32_0O.mat
RI32_0O.mat
NI32_1O.mat
RI32_1O.mat
x 10^-4
normalized impulse engery
50k samples per second


NI32_32O.mat
RI32_32O.mat
NI32_48O.mat
RI32_48O.mat
NI48_00O.mat
RI48_00O.mat
NI48_16O.mat
RI48_16O.mat
x 10^-4
normalized impulse engery
50k samples per second

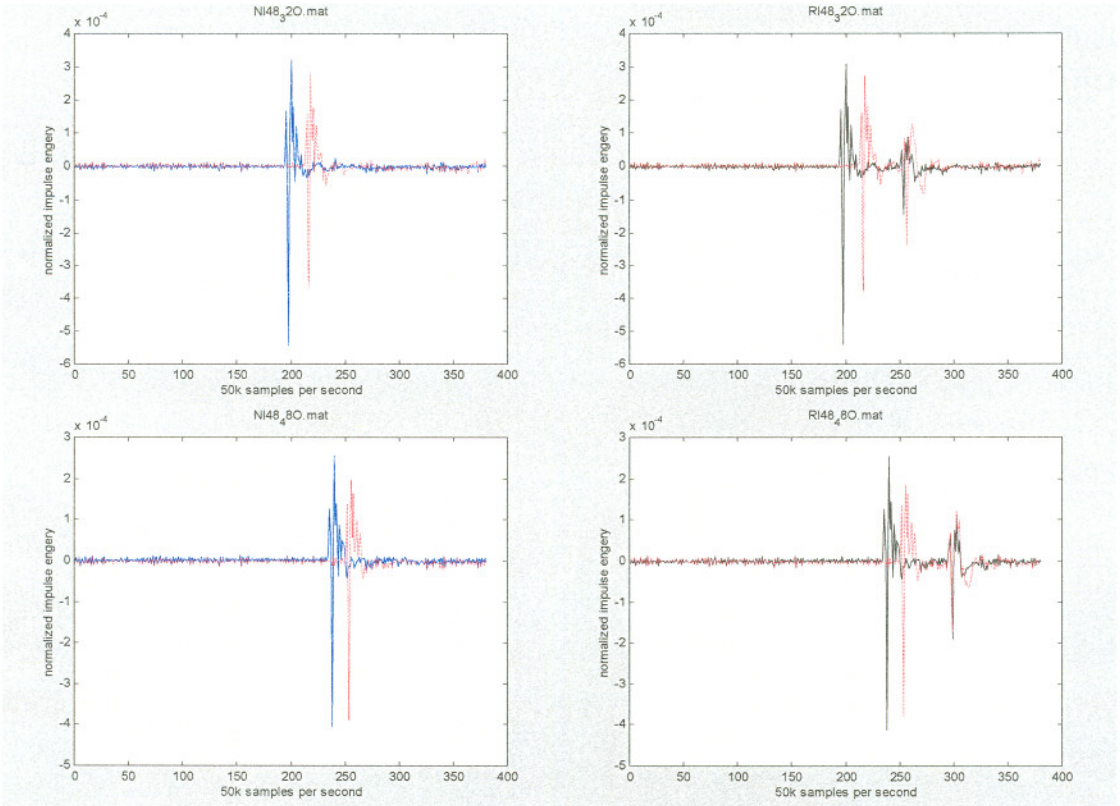


Figure 3.10: Impulse Responses for non-reverberant and reverberant configurations. Blue & Red are Left & Right microphones respectively in free space. Black & Red are Left & Right microphones respectively in reverberant conditions.

From these impulse responses we obtain a set of delay tables. (Table 3.1-3.3). First for the direct path measured separately and then for both the direct and reflected paths measured together. Since the sound source was not moved during a single location measurement (only the microphone jacks and the tile were altered) the values in the tables show remarkably persistent microphone and reflective surface measurements that are consistent beyond the forth decimal place.

Table 3.1 – Non-Reflective Direct Path Distance

Z\X	0"	16"	32"	48"
0"	N/A	-5.96376	-5.96376	-5.96376
16"	-0.27108	-4.0662	-5.4216	-5.69268
32"	0	-2.7108	-4.33728	-5.15052
48"	0	-2.16864	-3.52404	-4.33728

Table 3.2 - Reflective Direct Path Distances

Z\X	0"	16"	32"	48"
0"	N/A	-5.96376	-5.96376	-5.96376
16"	-0.27108	-4.0662	-5.4216	-5.69268
32"	0	-2.7108	-4.33728	-5.15052
48"	0	-2.16864	-3.52404	-4.33728

Table 3.3 - Reflective Image Path Distances

Z\X	0"	16"	32"	48"
0"	N/A	-1.62648	-2.43972	-7.31916
16"	2.98188	0.81324	-1.08432	-1.62648
32"	3.52404	1.62648	0.27108	-0.81324
48"	3.79512	2.16864	1.08432	0

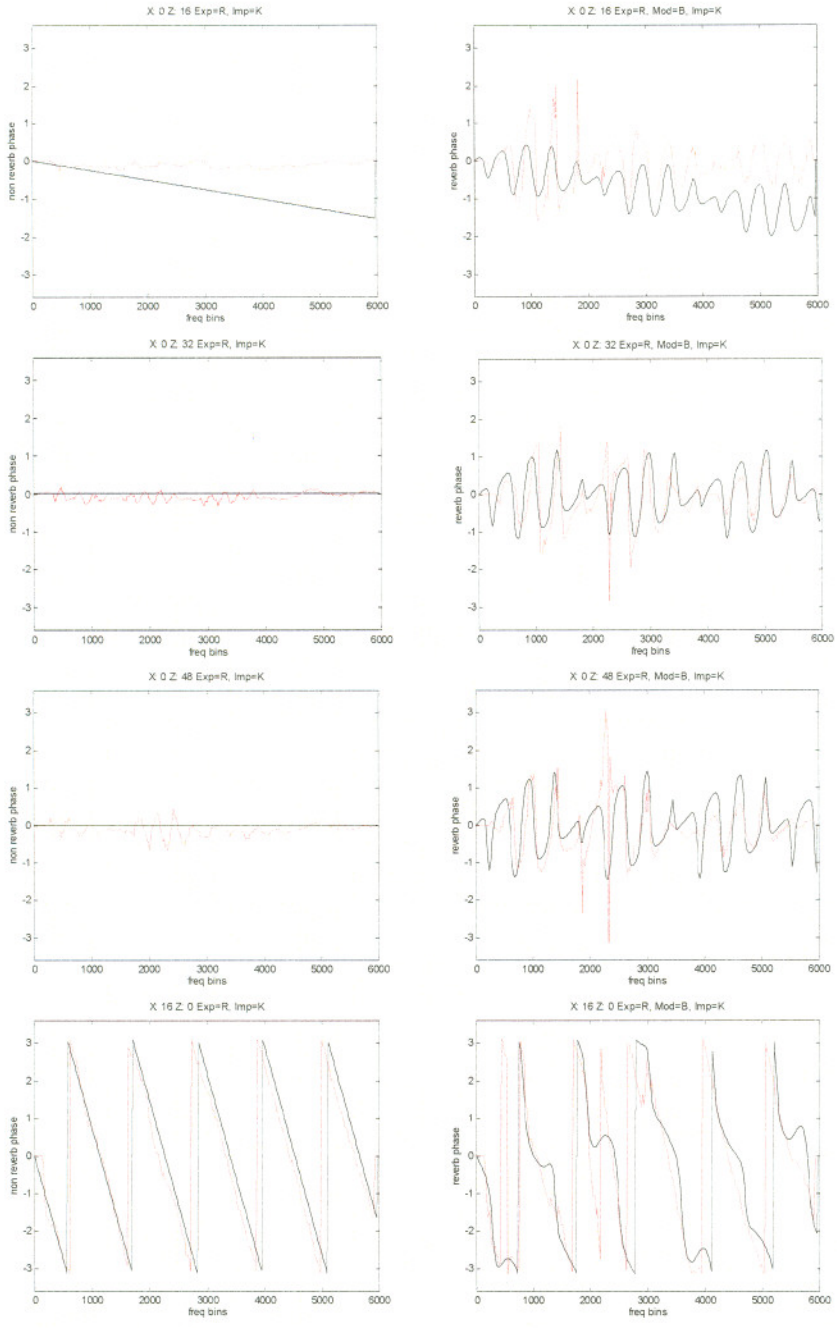
These measurements provided different values than the previous hand measured approach. The impulse measurements uncovered the following discrepancies with the Image Model.

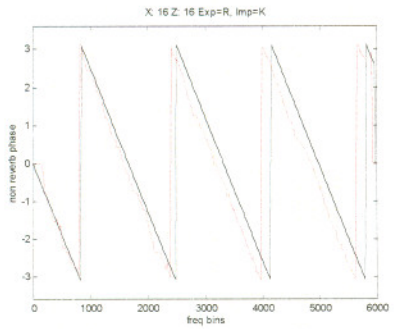
1. The physical placement of location (0,16) was incorrect. It should be zero for the direct path case and flat for the reverberant case. This was simply corrected.

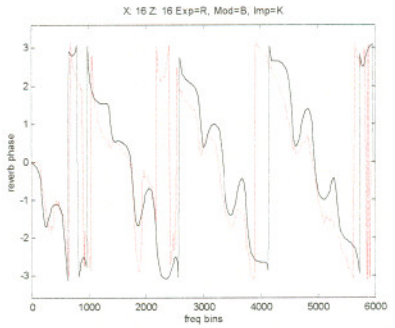
2. Locations (32,0), (48,0), and (48,16) have no strong reflection off the reflective surface. This was attributed to the physical placement of the tile. See Figure 3.4. Because the leading edge of the reflective surface along the X axis was placed almost even with the right microphone, there is no tile to reflect off. However, as the sound source nears the microphone along the X axis the surface become useful and we see (16,0) showing a strong image.

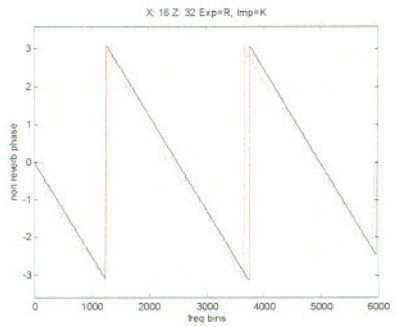
3. Locations (16,48) & (32,48) have a phase inversion during a portion of their overall signature. In this segment of the overall phase difference it appears that the phase is changing opposite to the model prediction.

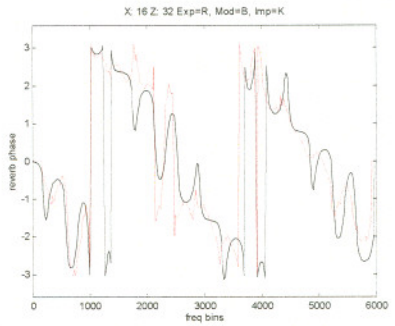
The direct and reflected delays were then used in the Image Model in place of the values measured by hand. The theoretical attenuations over the new distances were still used rather than differencing the amplitude values measured by the impulse response. This was done to isolate the effects of this improvement. Figure 3.11 show the phase signatures graphed as before. The left column shows the non-reflective case with black being the model and dashed red the measured values. The right column displays the phase signature case also with the modeled and measured in black and dashed red respectively.

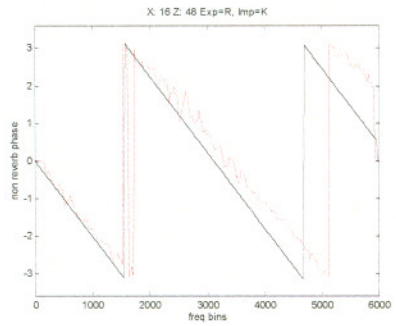

X: 0 Z: 16 Exp=R, Imp=K
non reverb phase
freq bins
X: 0 Z: 16 Exp=R, Mod=B, Imp=K
reverb phase
freq bins
X: 0 Z: 32 Exp=R, Imp=K
non reverb phase
freq bins
X: 0 Z: 32 Exp=R, Mod=B, Imp=K
reverb phase
freq bins
X: 0 Z: 48 Exp=R, Imp=K
non reverb phase
freq bins
X: 0 Z: 48 Exp=R, Mod=B, Imp=K
reverb phase
freq bins
X: 16 Z: 0 Exp=R, Imp=K
non reverb phase
freq bins
X: 16 Z: 0 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

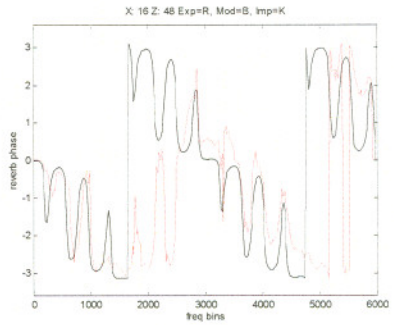

X: 16 Z: 16 Exp=R, Imp=K
non reverb phase
freq bins

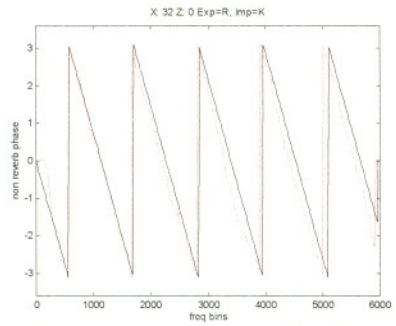

X: 16 Z: 16 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

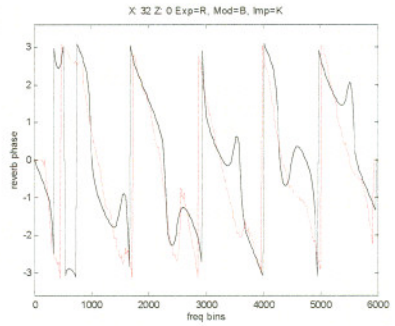

X: 16 Z: 32 Exp=R, Imp=K
non reverb phase
freq bins

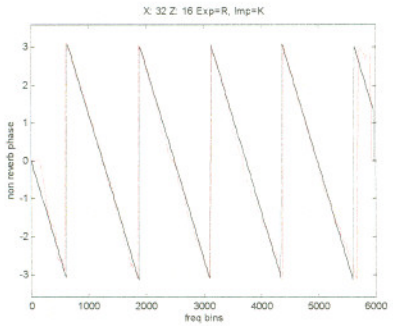

X: 16 Z: 32 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

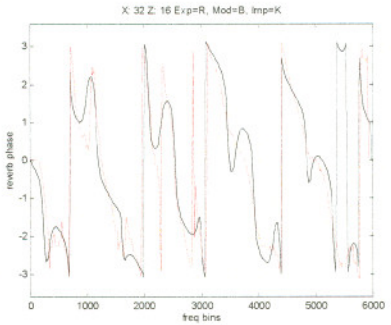

X: 16 Z: 48 Exp=R, Imp=K
non reverb phase
freq bins

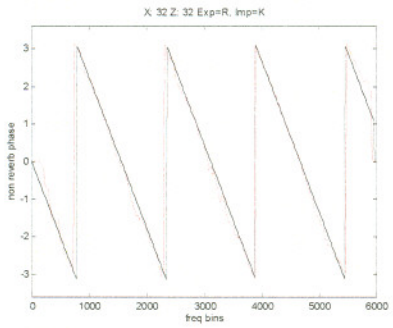

X: 16 Z: 48 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

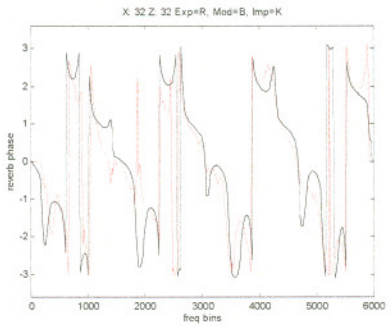

X: 32 Z: 0 Exp=R, Imp=K
non reverb phase
freq bins

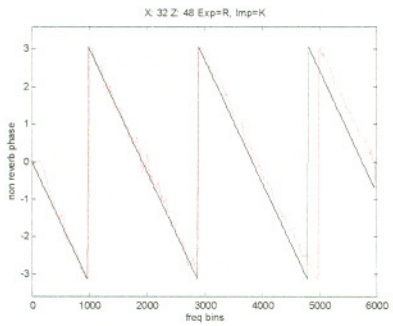

X: 32 Z: 0 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

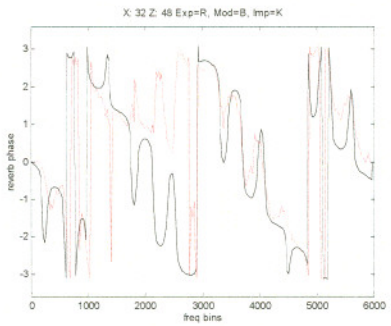

X: 32 Z: 16 Exp=R, Imp=K
non reverb phase
freq bins

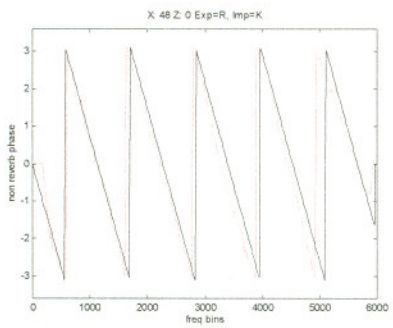

X: 32 Z: 16 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

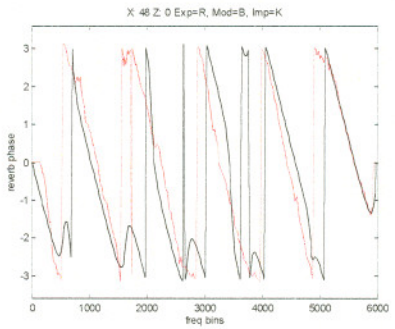

X: 32 Z: 32 Exp=R, Imp=K
non reverb phase
freq bins


X: 32 Z: 32 Exp=R, Mod=B, Imp=K
reverb phase
freq bins


X: 32 Z: 48 Exp=R, Imp=K
non reverb phase
freq bins


X: 32 Z: 48 Exp=R, Mod=B, Imp=K
reverb phase
freq bins


X: 48 Z: 0 Exp=R, Imp=K
non reverb phase
freq bins


X: 48 Z: 0 Exp=R, Mod=B, Imp=K
reverb phase
freq bins

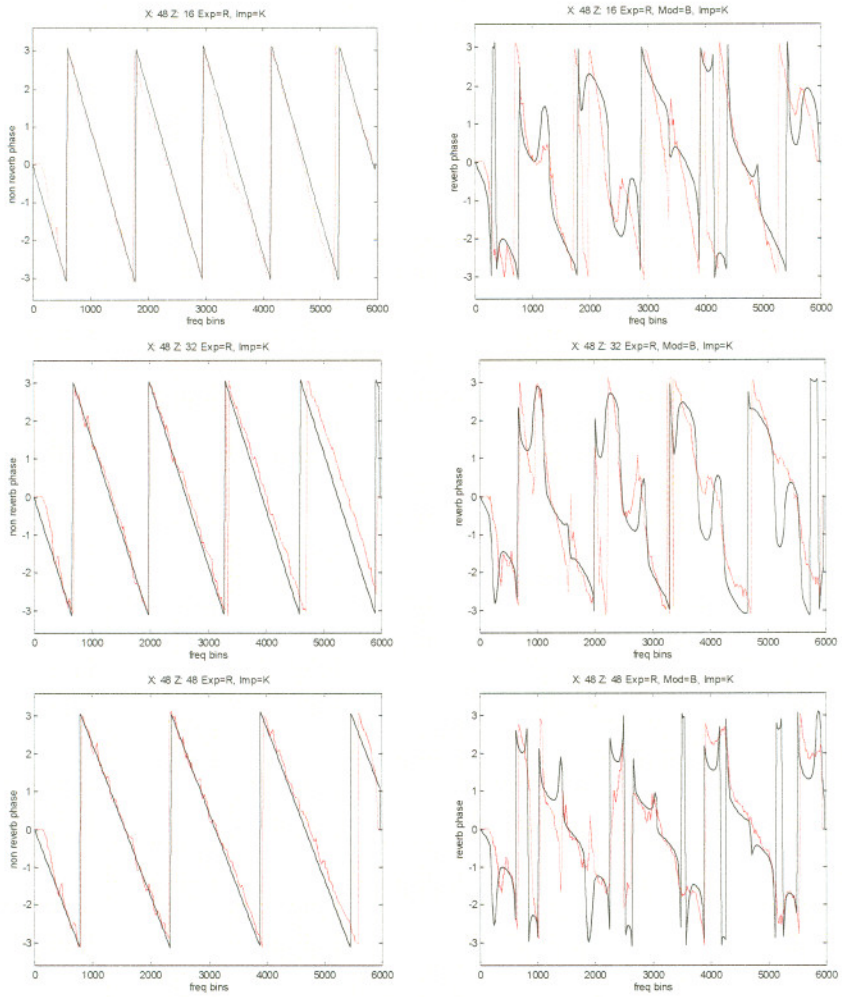


Figure 3.11: Phase Signature Impulse Model vs Actual Signature. Dashed Red is the actual signature. Solid Black is the new acoustically measured Image Model.

We can see visually from phase signatures in Figure 3.11 that the phase signature models are well aligned with most of the fluctuations encountered in the calibrated images. All of them are far better than those used in Figure 3.9. However, locations (16,48) & (32,48) still have the inverted phase over a small frequency range.

3.4.3 ACOUSTICALLY MEASURED METHOD

The above experiments show that the phase model is more complex than the one represented by the Image Model. Though it is very close, using this model will not produce phase signatures that represent the real-world condition when dealing with a reflective surface. This is not true for the non-reverberant case which does in fact closely follow the TDOA model as shown in the left column of Figure 3.11.

The final experiment seeks to eliminate the discrepancies between model complexities and measure only the benefits of the phase signature. To do this, focus is placed on measuring range discrimination along the hyperbolic paths of equal phase difference using an acoustically measured phase signature.

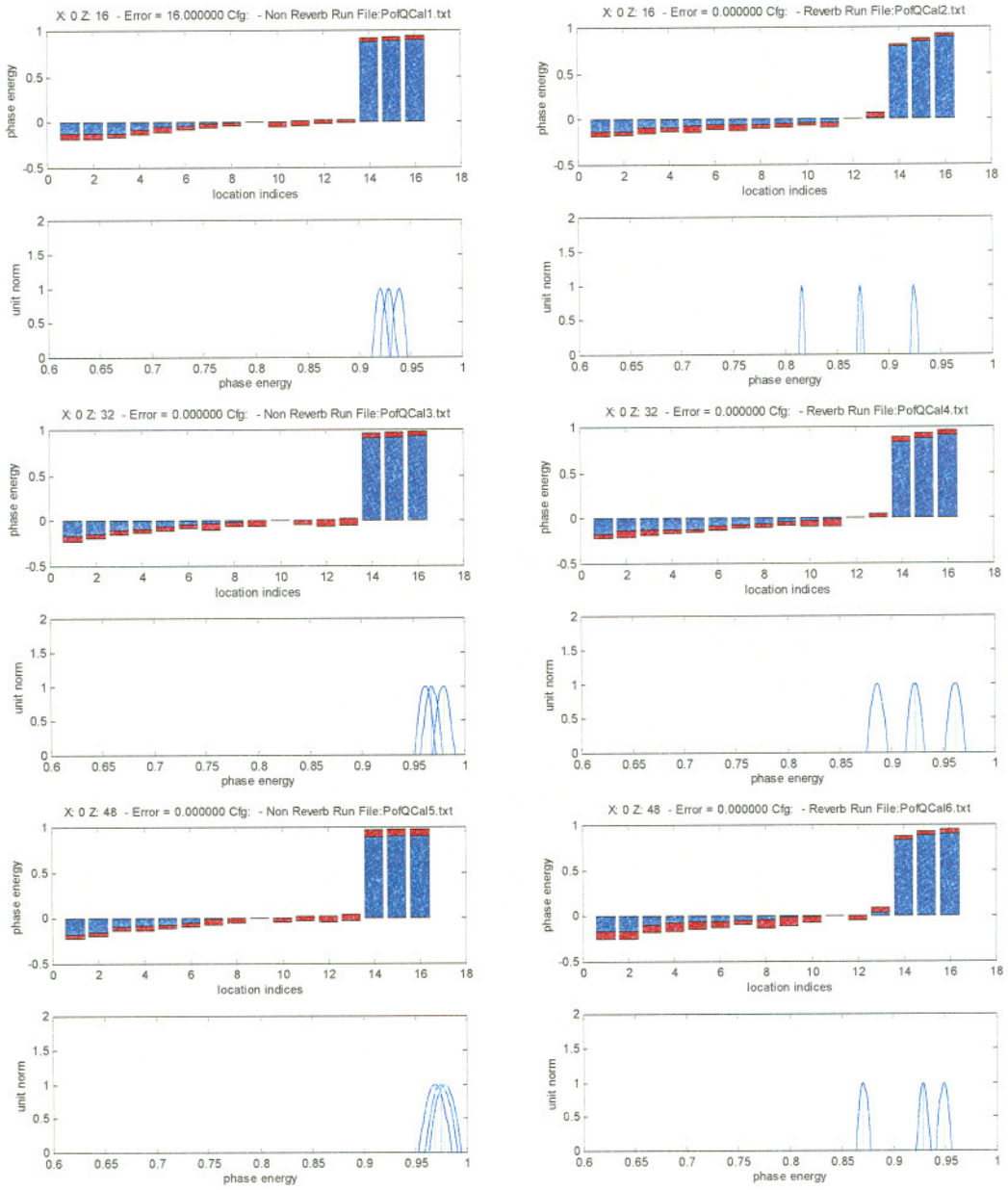
Using acoustically measured phase signatures for the reflective case will naturally include reflections from the surrounding room enclosure as well as the reflective surface. If we compare this to the simple non-reflective model of $e^{-j\omega(\tau_1-\tau_2)}$ we might draw poor conclusions because the calibration process itself could be the source of range discrimination. So the final experiment measures the performance of both reflective and non-reflective calibrated phase signatures. The original simple non-reflective model of $e^{-j\omega(\tau_1-\tau_2)}$ is not used at all because the phase signature for all signals along a single line of ambiguity (0,z) (x,0) or (x,x) consist of the same values. Since there is no difference any selection of range along these paths is arbitrary and we simply state that the distinction between ambiguity locations is zero.

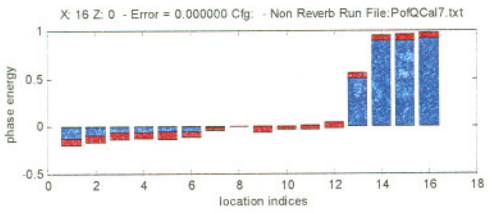
At each of the 15 locations, the PRN signal was played for 15 seconds and turned off by a timer. This was done once with out the reflective tile and once with the tile. The sets of phase signatures were then stored as the 'calibrated model' that included phase radians, phase variance, and 'SNR achieved' flag. This test was performed three times and the signatures were averaged. The

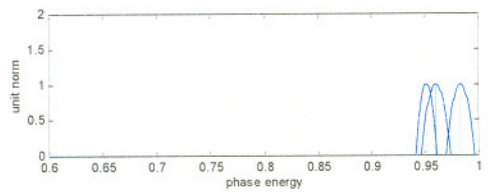
difference in the mean of the phase signatures across all frequencies from one pass to another was extremely small when the sound source was left untouched in the same location. However, it did vary when the sound source was moved and then 'replaced' on the testing location. This suggests that the method of placing the sound source into the identical location is more of a cause for error than averaging multiple calibrations. Therefore, during all calibrations for these experiments, the sound source was left untouched between non-reverberant and reverberant calibrations. This was done to reduce the problem of calibrating slightly different locations for the reverberant verses non-reverberant signatures. When we came back and placed the sound source at this location for testing, any error in the placement will be the same for both cases. Even doing this is insufficient, because we do not know the sensitivity of misplacement between the non-reverberant phase signatures and the reverberant phase signatures. So when we place the sound source back into position when running our performance test we will NOT compare the final power strength of $P(\vec{s})$ but only the differences between the method's ability to discriminate between locations. After clearing $P(\vec{s})$, the PRN sequence was played for 10 seconds in one of the 15 locations. The entire $P(\vec{s})$ for all locations and their corresponding variances were logged. This was done with and without the tile in place. The sound source was not touched during this time. This was then repeated for all 15 locations.

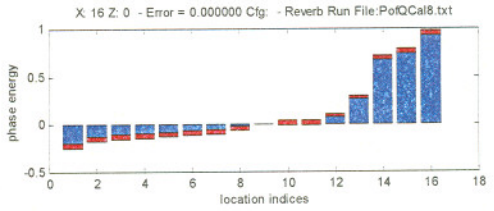
The following plots showcase results of this testing for each location. The 15 $P(\vec{s})$ values are plotted in ascending order in a bar graph for each location. On closer inspection it was determined that all locations registering similar $P(\vec{s})$ values resided along the regions of ambiguity. This makes these graphs easy to interpret. The three bars to the right with similar values are physically adjacent to one another and our objective has been to separate them. The negative slope of these three bars is an indication of the shaped of the $P(\vec{s})$ surface as that location. Below the bar

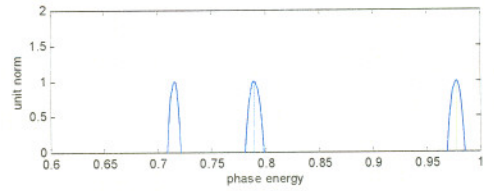
graph the same test results are used to plot all $P(\bar{s})$ locations that registered greater than 0.6 for that test run. These are plotted on the X axis along with their corresponding variances. Only the 1 sigma point is plotted to make the graphs easier to view.

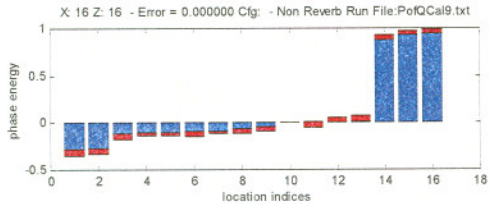


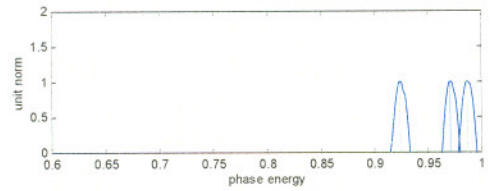

X: 16 Z: 0 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal7.txt
phase energy
location indices

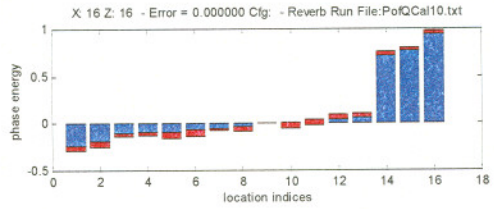

unit norm
phase energy

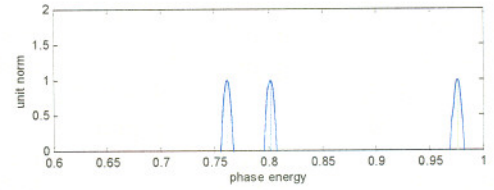

X: 16 Z: 0 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal8.txt
phase energy
location indices

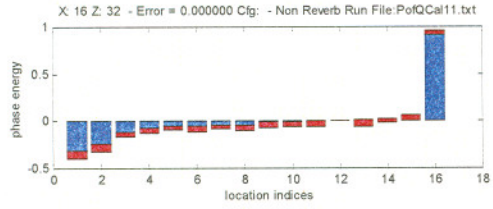

unit norm
phase energy

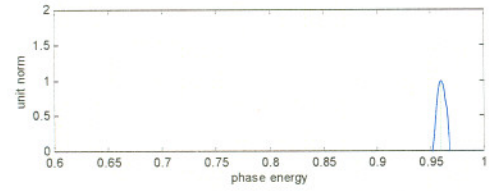

X: 16 Z: 16 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal9.txt
phase energy
location indices

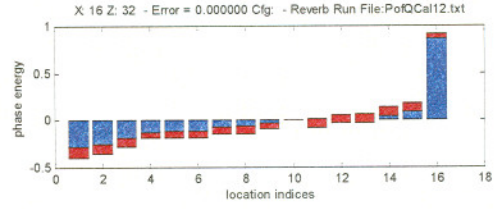

unit norm
phase energy

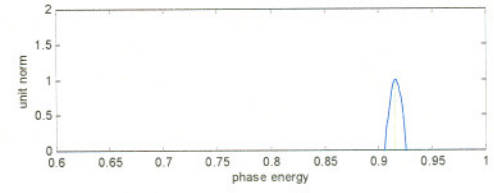

X: 16 Z: 16 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal10.txt
phase energy
location indices

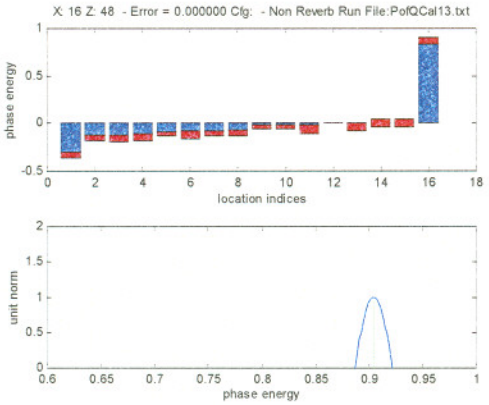

unit norm
phase energy

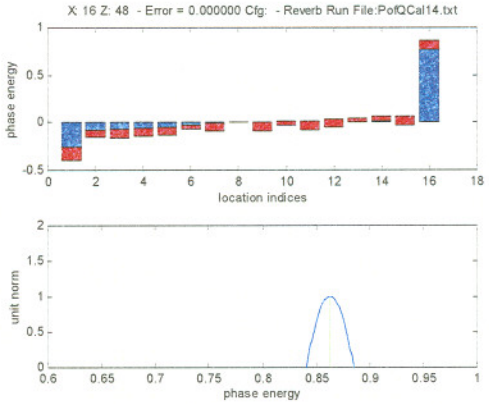

X: 16 Z: 32 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal11.txt
phase energy
location indices

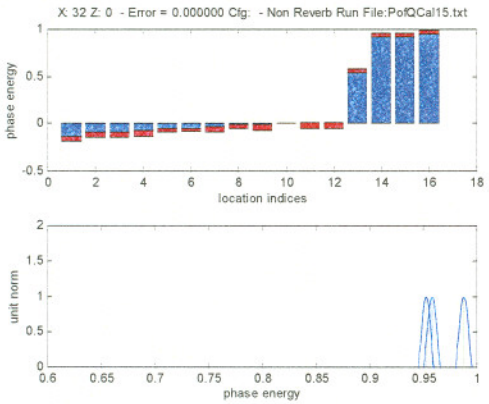

unit norm
phase energy

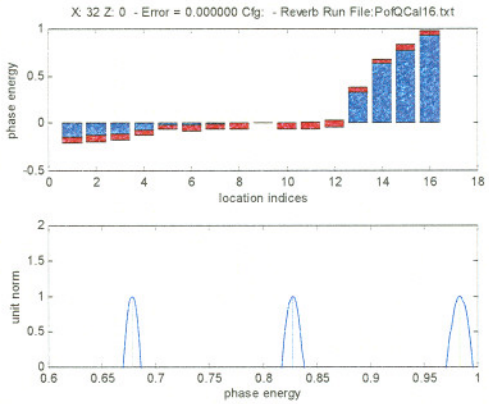

X: 16 Z: 32 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal12.txt
phase energy
location indices

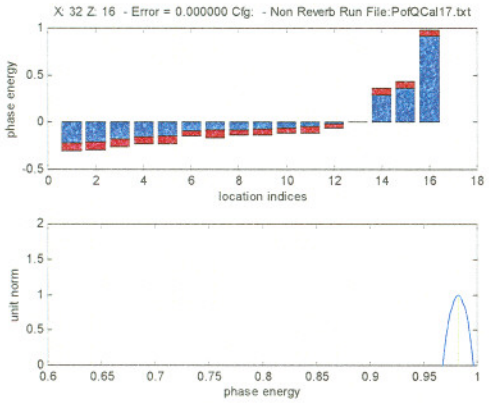

unit norm
phase energy

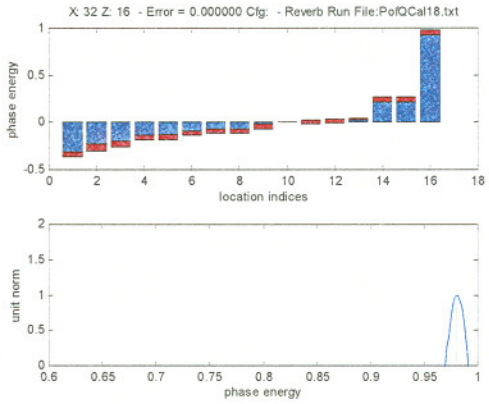

X: 16 Z: 48 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal13.txt
phase energy
location indices
unit norm
phase energy

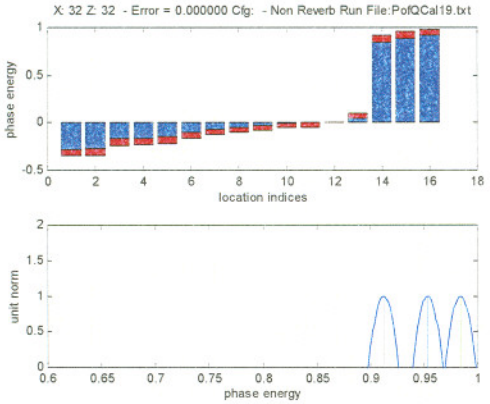

X: 16 Z: 48 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal14.txt
phase energy
location indices
unit norm
phase energy

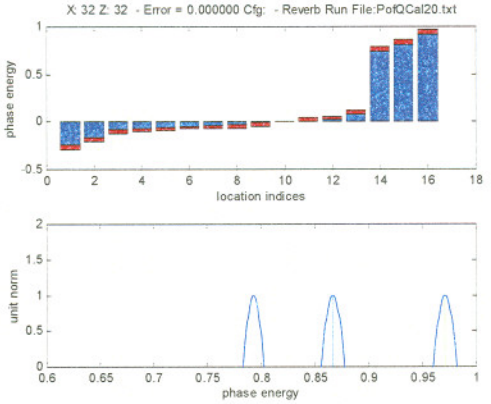

X: 32 Z: 0 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal15.txt
phase energy
location indices
unit norm
phase energy

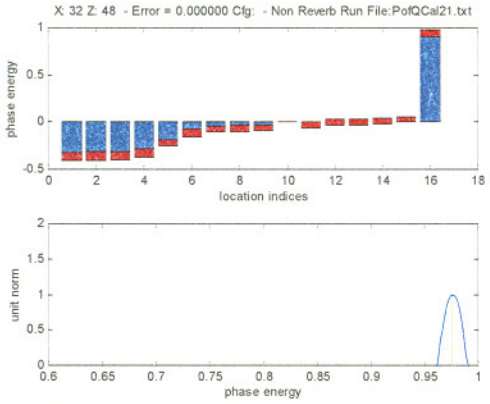

X: 32 Z: 0 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal16.txt
phase energy
location indices
unit norm
phase energy

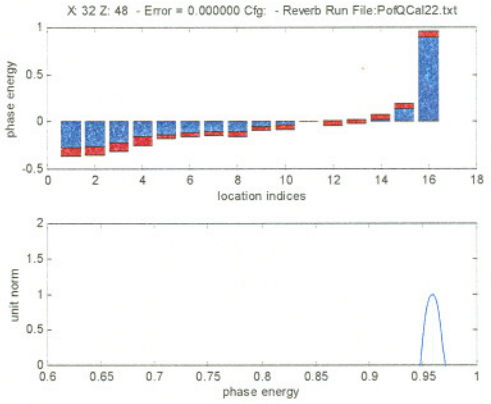

X: 32 Z: 16 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal17.txt
phase energy
location indices
unit norm
phase energy

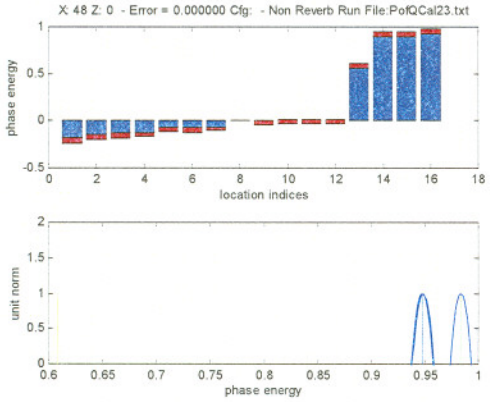

X: 32 Z: 16 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal18.txt
phase energy
location indices
unit norm
phase energy

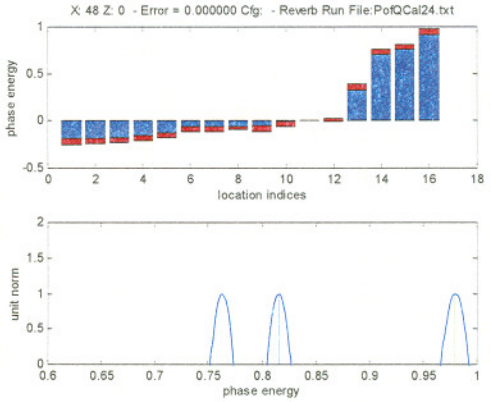

X: 32 Z: 32 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal19.txt
phase energy
location indices
unit norm
phase energy


X: 32 Z: 32 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal20.txt
phase energy
location indices
unit norm
phase energy


X: 32 Z: 48 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal21.txt
phase energy
location indices
unit norm
phase energy


X: 32 Z: 48 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal22.txt
phase energy
location indices
unit norm
phase energy


X: 48 Z: 0 - Error = 0.000000 Cfg: - Non Reverb Run File:PofQCal23.txt
phase energy
location indices
unit norm
phase energy


X: 48 Z: 0 - Error = 0.000000 Cfg: - Reverb Run File:PofQCal24.txt
phase energy
location indices
unit norm
phase energy

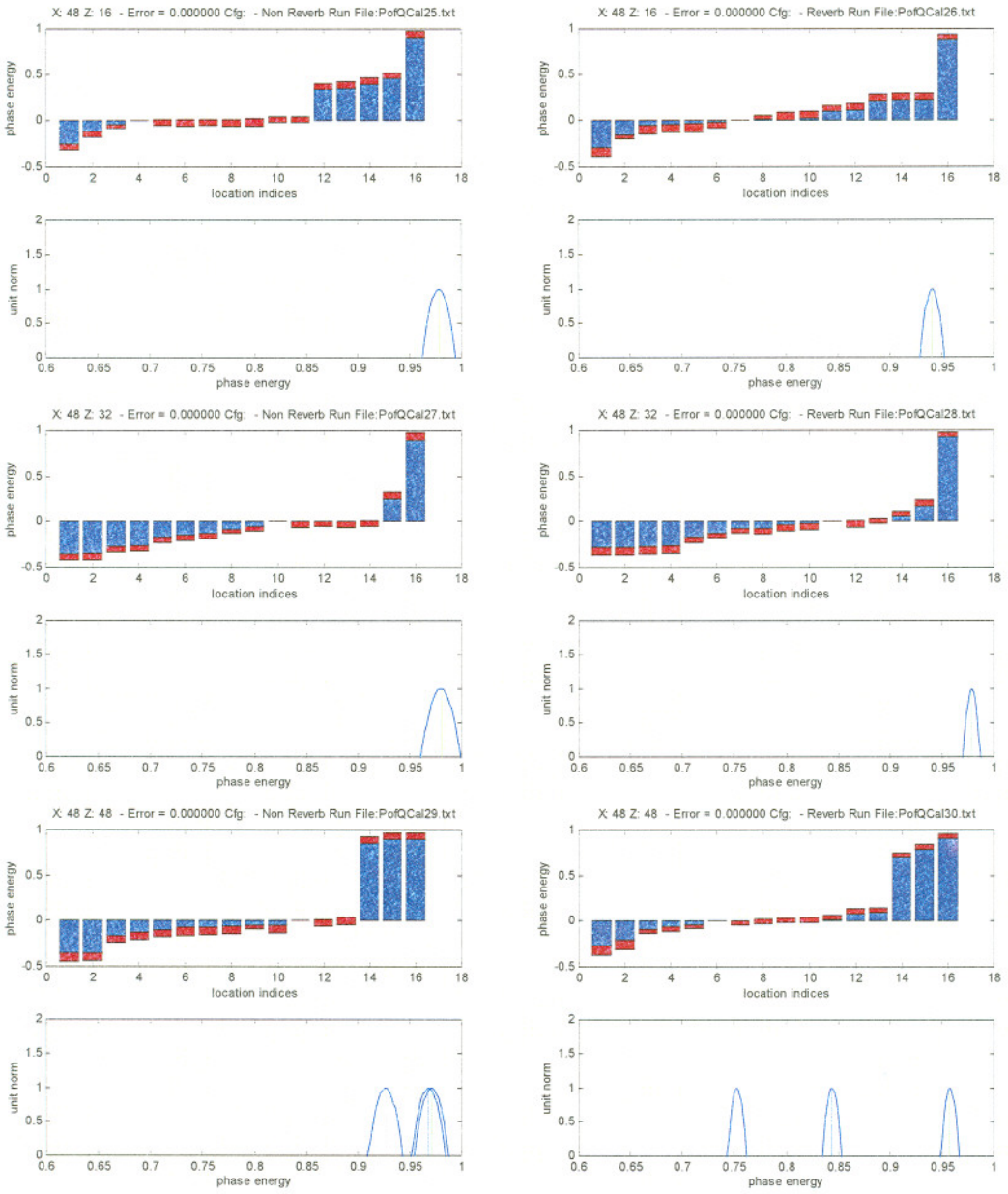


Figure 3.12: Average P(s) Bar Graph. Average performance of P(s) at each of the 15 locations with sound source emitting from a single location. The bars represent the average P(s) value. They are in ascending order left to right. The subplot below the Bar Graph represents an alternative view with P(s) along the x axis. Only the top 0.6 contenders are displayed along with their 1 sigma spread.

All test cases show both the non-reverberant and reverberant configurations improved ability to distinguish the exact location over all other locations. The ambiguity angles defines as broadside

(0,16)(0,32)(0,48), diagonal (16,16)(32,32)(48,48), and end-fire (16,0)(32,0)(48,0), each demonstrate an increase in the slope of the $P(\vec{s})$ surface with respect to their nearest neighbors. The non-reverberant calibration slightly improves the ability to distinguish previous areas of ambiguity showing that even simple reverberation can provide an indication of range. However, as seen in the non-reverberant cases (16,0) (0,16) (32,0) (0,32) & (48,48), this improvement was less than the 1 sigma of the measurement noise, suggesting that this improvement is of limited value.

The reverberant configuration outperforms the non-reverberant case along all angles of ambiguity. Though the reverberation configuration provided a superior signal characteristic for resolving range ambiguity, inspection of $P(\vec{s})$ at angles other than broadside, diagonal, and end-fire reveal that reverberation has flattened the overall height of the surface. First the tests at (16,32) and (32,48) depict a decrease in the difference in power between the true location and the nearest neighbor compared to the non-reflective case. Second, $P(\vec{s})$ registers an average of 2% smaller in the reverberation configuration as compared to the non-reverberant configuration. This may not be significant but it was consistent. The combination of these two effects creates a surface for the reverberant case that is shorter and flatter over all locations. Fortunately the $P(\vec{s})$ for these nearest neighbors is still well below the 4 sigma of the strongest $P(\vec{s})$ location.

Closer inspection of the reverberant cases shows that there is a marked different between the three ambiguity angles. The broadside angles did not separate the signature as well as well as the diagonal and end-fire cases. For example (0,48) did not establish a reliable distinction beyond the 1 sigma mark for $P(\vec{s})$. It did however succeed in improving the distinction between the true location and its second nearest signature beyond the 1 sigma point, which outperforms the non-reverberant case.

The following charts summarize the above performance plots of the Phase Signature approach compared to the calibrated version of the simple pair-wise TDOA model.

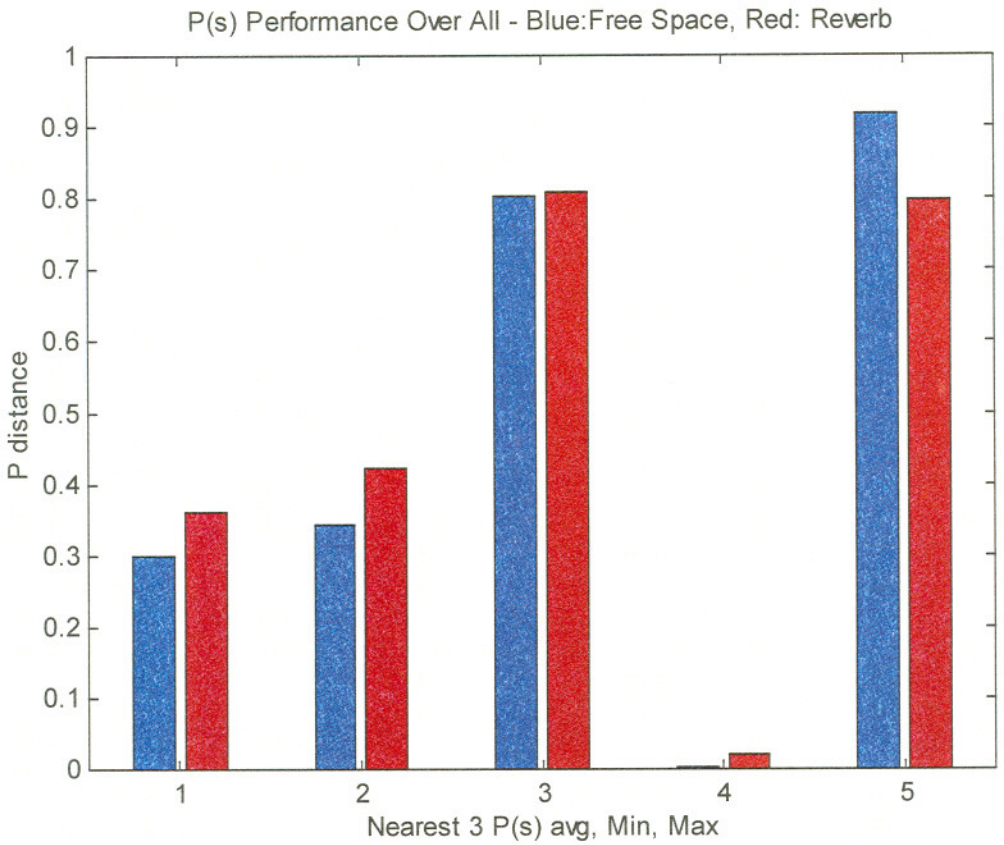


Figure 3.13: P(s) Overall Performance. Blue – non-reverberant, Red – reverberant. Column 1, 2, & 3 represent difference between peek and 1st, 2nd, and 3rd nearest neighbors respectively. Columns 4 and 5 represent the minimum and maximum separations from the peak respectively.

Figure 3.13 charts $P(\vec{s})$ performance of the system. The first thee columns present average performance measured by the difference between the maximum $P(\vec{s})$ and the first three nearest $P(\vec{s})$ values. Column 4 charts the minimum difference between the maximum $P(\vec{s})$ and its nearest value. Column 5 charts the maximum difference between the maximum $P(\vec{s})$ and its nearest thee values. In all of these tests it turned out that the nearest $P(\vec{s})$ values also represented the nearest physical locations so distance is not added to the chart. On average the reverberant phase signature

tests slightly outperformed the non-reverberant calibrated test. The reverberant phase signature shows only slight improvement over the calibrated case. As we have seen from inspecting the individual test cases this improvement is not enough to successfully resolve the two locations. Column 5 shows that the maximum different of the first three nearest $P(\vec{s})$ values is greater in the non-reverberant case than in the reverberant case. This suggests that the peak of $P(\vec{s})$ in the reverberant case is not as sharply defined along locations of non-ambiguity.

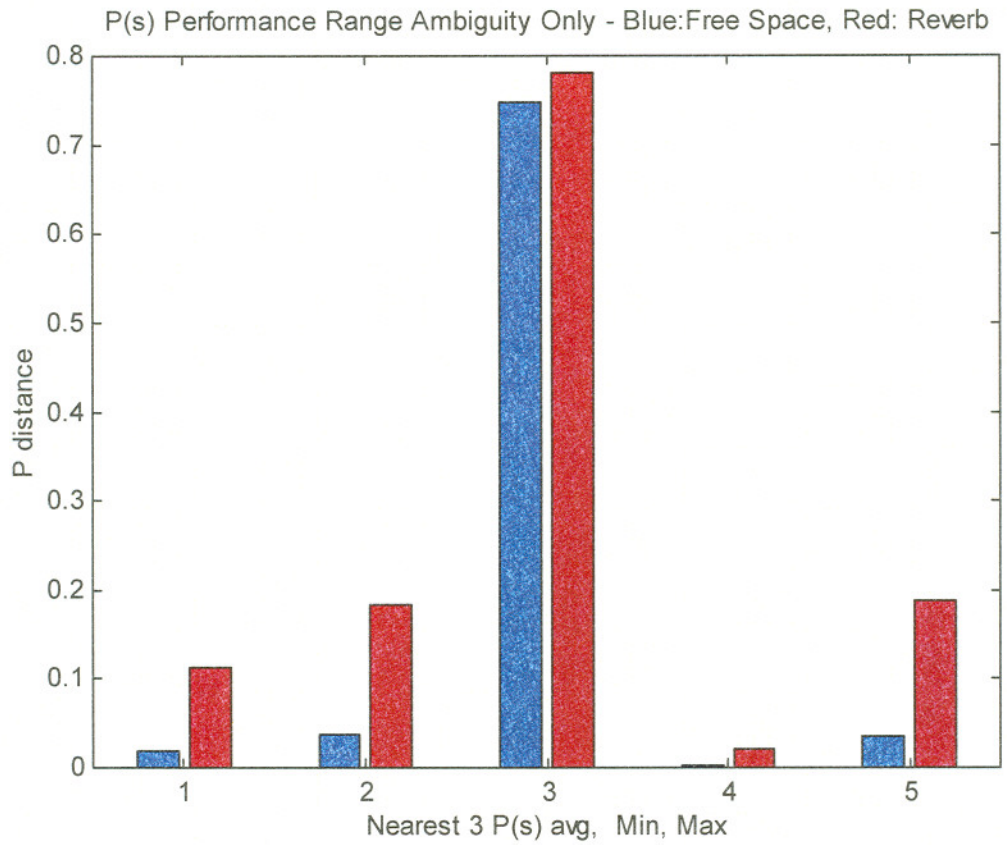


Figure 3.14: P(s) Performance Range Ambiguity Only. Blue – non-reverberant, Red – reverberant. Columns 1, 2, &3 represent the difference between peak and 1st, 2nd, and 3rd nearest neighbors of 9 locations with range ambiguity. Columns 4 & 5 represent the minimum and maximum separations respectively.

Figures 3.14 charts $P(s)$ performance only over locations with range ambiguity. This view creates additional insights into the benefits of the phase signature approach. The bars are defined as in Figure 3.13 except the averages only include the 9 locations used in broadside, diagonal, and end-fire testing. With this view the following observations are made.

1. Column one shows that the reverberant method is able to distinguish the location of the sound source over the calibrated non-reverberant model.

2. The reverberant enclosure was able to distinguish the locations of ambiguity 10% better than those calibrated to the room, which suggest that the reflector was effective.

3. Looking at the 1st and 2nd distances suggests a clear sharpening of the $P(s)$ contour beyond its nearest neighbor of 18%.

4. Column 4 indicates that there was at least one case where the reverberant enclosure did not improve the range ambiguity significantly. Further investigation shows that this was location (0,32). With location (0,48) also showing limited signs of improvement along the array broadside.

Further breakdown of the data as presented in Figures 3.15, 3.16 and 3.17 shows the individual performance along the three different angles of ambiguity. The average slope of $P(s)$ for the three locations of ambiguity is plotted over the 32" separation in points. (normalized by $\sqrt{2(32^2)}$ for the diagonal case). The yellow plot is always unity and is added as a reminder that without any calibration the phase signatures are identical for all locations so the $P(s)$ difference is always zero. The blue plot shows the slope of $P(s)$ for the non-reverberant case. The green plot shows the slope of $P(s)$ for the reverberant case. Plotting over 32" in some cases represents the shape of

$P(s)$ while in others represents the slope for 16" and then projects below the true surface for the remaining 16".

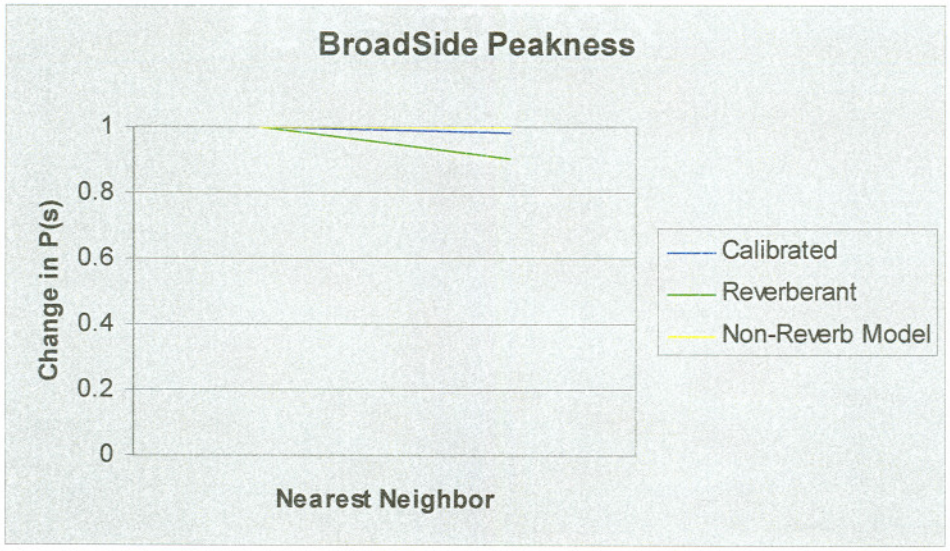


Figure 3.15: Broadside Ambiguity Improvement. P(s) slope over 32 inches.

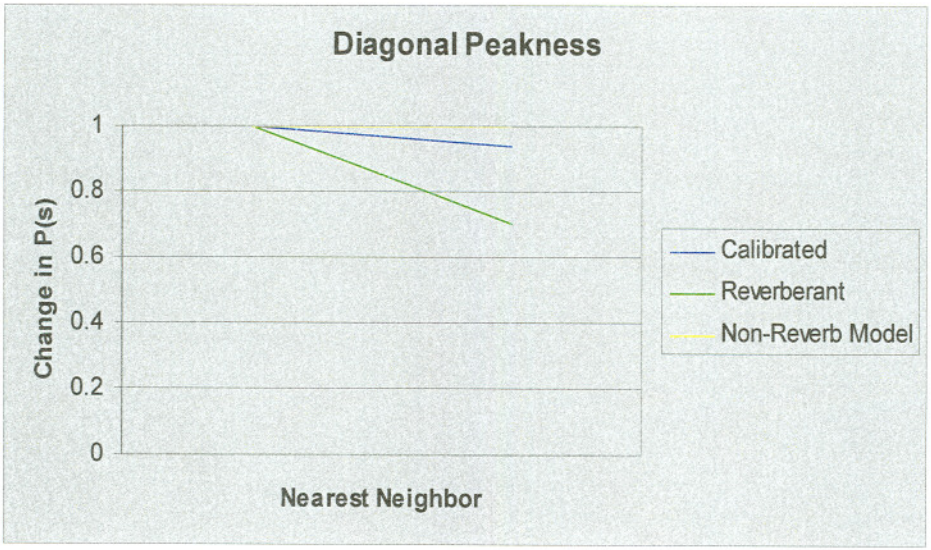


Figure 3.16: Diagonal Ambiguity Improvement. P(s) slope over 32 inches.

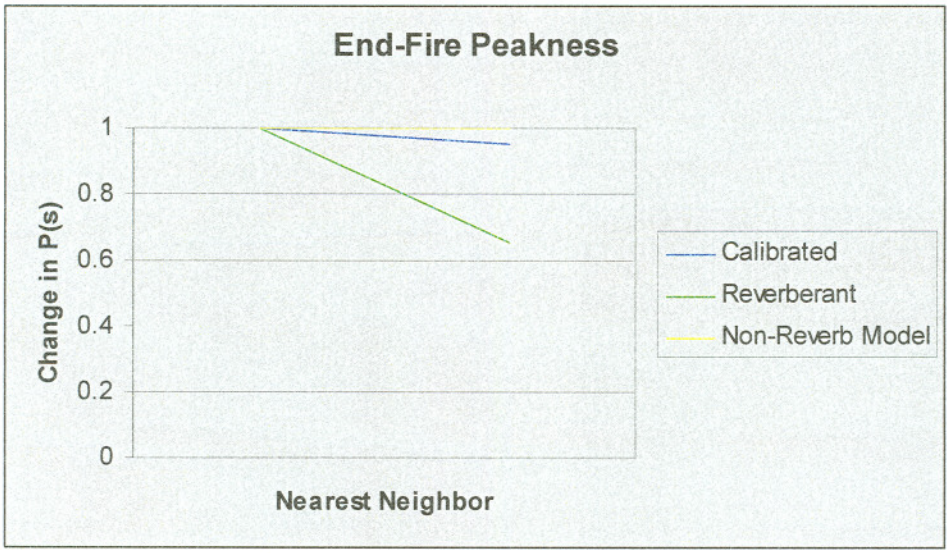


Figure 3.17: End-Fire Ambiguity Improvement. P(s) slope over 32 inches.

In all cases, as readily notable in the bar graphs, the phase signature method outperforms the calibrated non-reverberant case. Comparing the improvements between the three different ambiguity axes we see that the greatest improvement in slope was found along the X axis or (x,0). This is referred to as the end-fire direction. The least improvement was obtained along the Z axis or (0,z). This is referred to as the broadside direction. The performance improvement along the diagonal angle of ambiguity was comparable to the end-fire case. This may be attributed to the fact that the impulse response has an intentional anomaly along the end-fire direction especially at location (0,48) where the right microphone does not receive a sound source image while the left microphones does. In this way, the phase signature had greater changes along the end file direction than in the broadside case where the phase signature varied only slightly because the distance between the reflective paths varies only slightly. These differences indicate that a richly varied phase signature can be produced and used to distinguish range though non-trivial acoustic modeling may be required to establish the phase signatures in advance for all locations.

4 FUTURE WORK

Small Enclosure Modeling. The primary next step for this effort is to fully explore the modeling of the acoustical enclosure. Very little work, if any, has been done to deliberately place a reverberant chamber around microphones used in sound source localization. Doing so in the manner presented in this work creates a need for very complex models of the physics behind wave propagation. One starting place could be in the ongoing work by researchers to produce life-like 3-D virtual reality sound. Another would be to pursue efforts in building acoustics and scale them down to the type of enclosures of interest. During the investigation for optimal reflector positioning, many microphone and reflector arrangements, beyond the one selected, were considered. Examples include wrapping the microphones in a plastic PC Tablet enclosure and placing the reflector between the two microphones. Both of these created more phase distinction than the simple reflective arrangement used in testing, however, the model was impractical to predict and arguably overly complex for this introductory research. Elaboration on the enclosure modeling could possibly produce two benefits to this work:

1. It could be of great interest to produce a method to accurately model the phase signature of every location with a fixed area and eliminate the need for the impractical calibration process.

2. Using accurate acoustical modeling may provide a means to search for an enclosure shape that separates all phase signatures in some optimal way. This might mean that all signatures are separated by at least a specified distance while still maintaining that the closest signatures reside at adjacent physical locations.

Speech Specific Filtering. The secondary next step for this effort is to establish full operation on a speech signal. The real-time set up described in this work functions on broadband noise as well as speech. However the speech signals responds slowly because few bins register above 20dB in a given frame. Frequency bins less than 20dB above the noise floor are difficult to distinguish between speech and high variance noise and so, without a more sophisticated discriminator, are discarded. Though not formally reported here, preliminary investigation suggested that there could be a benefit to considering <20dB SNR frequency bins based on harmonics of stronger signals within an audio frame. By including more phase signatures the energy recorded in $P(\vec{s})$ will react more in accordance to natural movement and speech patterns.

5 CONCLUSIONS

The goal of this thesis was to explore the plausibility of resolving the range of a sound source using only two microphones. It has been shown that at a minimum, using state-of-the-art TDOA in conjunction with a reflective surface placed on the opposite side of two microphones from a broadband sound source it possible to determine the range of the sound source from those microphones. Based purely on mathematical derivation, traditional TDOA model assumptions cannot resolve this range. The reflective surface was shown to sharpen the peak of a phase difference distance contour beyond simple room calibration. By accurately measuring the impulse response of the enclosure we can determine the accurate physical geometries of the enclosure and use them to establish the basic structure of the phase signatures for the majority of locations calibrated. However, there are phase anomalies encountered in the real-world measurement not fully characterized in the common 3-D simulation model used, preventing a continuous phase signature surface from being accurately described. By using a calibration method for phase signatures, a rigid surface that does not completely reflect the entire near-field wave form is used to establish an identifiable phase signature along the end-fire and diagonal source directions which is shown to outperform the broadside angle. A loosely drawn conclusion of this behavior is that there are non-trivially shaped enclosures that could establish phase signatures capable of range discrimination beyond a simple reflective plane.

In conclusion, reverberation is a natural phenomenon and this study shows one way to capitalize on its characteristics to enhance the performance of microphones arrays used in sound source localization. Showing that the pair-wise comparison of two microphones within a prescribed enclosure can resolve the range along regions of ambiguity is one step towards designing inexpensive audio capture modules embedded in walls, appliances, and electronic equipment that spatially filter sound based on both its distance and angle from the microphones.

6 REFERENCES

[1] J. Abel, J. Smith, *The spherical interpolation method for closed-form passive source localization using range difference measurements*, in Proc. IEEE ICASSP, vol 12, April 1987, pp. 471-474.

[2] J. B. Allen, D. A. Berkley, *Image Model for efficiently modeling small-room acoustics*, Journal of the Acoustical Soc. of America, vol 65, 1979, pp. 943-950.

[3] S. Bedard, B. Champagne, A. Stephenne. *Effects of room reverberation on time-delay estimation performance*, in Proc. IEEE ICASSP, vol 2, April 1994, pp. 261-264.

[4] D. Begault, *3-D Sound for virtual reality and multimedia*, Academic Press, 1994.

[5] J. Benesty, *Adaptive eigenvalue decomposition algorithm for passive acoustic source localization*, Journal of the Acoustical Soc. of America, vol 107, 2000, pp. 384-391.

[6] M. S. Brandstein, *A Framework for Speech Source Localization using Sensor Arrays*, PhD Thesis, Brown University, Providence, RI, May 1995.

[7] M. S. Brandstein, *On the use of explicit speech modeling in microphone array applications*, in Proc. ICASSP, vol 6, May 1998, pp. 3613-3616.

[8] M. S. Brandstein, J. E. Adcock, H. F. Silverman, *A closed-form location estimator for use with room environment microphone arrays*, IEEE Trans. on Speech and Audio Processing, vol 5, 1997, pp. 45-50.

[9] M. S. Brandstein, J. E. Adcock, H. F. Silverman, *A closed-form method for finding source locations from microphone-array time-delay estimates*, in Proc. IEEE ICASSP, vol 17, April 1995, pp. 3019-3022.

[10] M. S. Brandstein, H. F. Silverman. *A robust method for speech signal time-delay estimation in reverberant rooms*, in Proc. IEEE ICASSP, vol 1, April 1997, pp.375-378.

[11] B. Champagne, S. Bedard, A. Stephenne, *Performance of time-delay estimation in the presence of room reverberation*, IEEE Trans. on Speech and Audio Processing, vol 4, 1996, pp. 148-152.

[12] Y. Chan, K. Ho, *A simple and efficient estimator for hyperbolic location*, IEEE Trans. on Signal Processing, vol 42(8), 1994, pp. 1905-1915.

[13] W. Chu, *Impulse-response and reverberation-decay measurements made by using a periodic pseudorandom sequence*, Applied Acoustics, vol 29, 1990, pp. 193-205.

[14] J. DiBiase, *A high-accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments*, PhD Thesis, Brown University, Providence, RI, May 2000.

[15] R. Duncan Luce, *Sound & Hearing: A Conceptual Introduction*, Lawrence Erlbaum Assoc., 1992.

[16] G. Elko, *Microphone array systems for hands-free telecommunication*, Speech Communication, vol 20, 1996, pp 229-240.

[17] A. Evans, *Making Sense of Sound: The basics of Audio Theory and Technology*, Prompt Publications, 1992.

[18] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West, M. M. Sondhi, *Autodirective microphone systems*, ACUSTICA, vol 73, 1991, pp. 58-71.

[19] O. L. Frost III, *An algorithm for linearly constrained adaptive array processing*, in Proc. of the IEEE, vol 60, 1972, pp. 926-935.

[20] S. Gay, J. Benesty, *Acoustic Signal Processing for Telecommunication*, Kluwer Academic Publishers, March 2000.

[21] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 3rd edition, 1996.

[22] E. Jan, P. Svaizer, J. Flanagan, *Matched-Filter processing of microphone array for spatial volume selectivity*, in Proc. of International Symposium on Circuits and Systems, 1995, pp. 1460-1463.

[23] W. Kellermann, *A self-steering digital microphone array*, in Proc. ICASSP, vol 5, April 1991, pp. 3581-3584.

[24] G. Kendall, W. Martens, M. Wilde, *A spatial sound processor for loudspeaker and headphone reproduction*, in Proc. of the AES 8th International Conference. New York 1990.

[25] Knapp, C.H. and Carter, G.C., *The generalized correlation method for estimation of time delay*, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol ASSP-24, 1976, pp 320-327.

[26] P. Ladefoged, *A Course In Phonetics*, Harcourt & Brace, 3rd edition, 1993.

[27] J. Minkoff, *Signals, Noise, and Active Sensors: Radar, Sonar, Laser Radar*, Wiley-Interscience, 1992.

[28] B. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 1997.

[29] T. Nishiura, S. Nakamura, K. Shikano, *Speech Enhancement by multiple beamforming with reflective signal equalization*, in Proc. IEEE ICASSP, vol 1, May 2001, pp. 189-192.

[30] M. Omologo, P. Svaizer, *Acoustic Event Localization using a Crosspower-Spectrum Phase based Technique*, in Proc. IEEE ICASSP, vol 2, April 1994, pp. 273-276.

[31] M. Omologo and P. Svaizer, *Acoustic source location in noisy and reverberant environment using CSP analysis*, in Proc. IEEE ICASSP, vol 2, May 1996, pp. 921-924.

[32] D. Rabinkin, *Digital Hardware and Control for a Beamforming Microphone Array*, M.S. Thesis, Rutgers University, New Brunswick, NJ, January 1994.

[33] D. Rabinkin, *Optimum Sensor Placement for Microphone Arrays*, PhD Thesis, Electrical and Computer Engineering, Rutgers University, May 1998.

[34] D. Rabinkin, R. Renomeron, J. Fench, J. Flanagan. *A DSP implementation of source location using microphone arrays*, in Proc. SPIE, vol 2846, 1996, pp. 88-99.

[35] S. Reddi, *An exact solution to range computation with time delay information for arbitrary array geometries*, IEEE Trans. on Signal Processing, vol 41(1), 1993, pp. 485-486.

[36] H. Schau, A. Robinson, *Passive source localization employing intersecting spherical surfaces from time-of-arrival differences*, IEEE Trans. on Acoustics, Speech, and Signal Processiong, vol ASSP-35(8), 1987, pp 1223-1225.

[37] M. R. Schroeder, *New method of measuring reverberation time,*, Journal of the Acoustical Soc. of America, vol 37(1), 1965, pp. 409-412.

[38] H. Wang, P. Chu, *Voice source localization for automatic camera pointing system in videoconferencing*, in Proc. IEEE ICASSP, vol 1, 1997, pp. 187-190.

[39] B. Widrow, S. D. Stearns, *Adaptive Signal Processing*, London Press, 2nd edition, 1985.

[40] D. Wright, J.H. Hebrank, B. Wilson, *Pinna reflections as cues for localization*, Journal of the Acoustical Soc. of America, vol 56, 1974, pp. 957-962.